INTRODUCTION TO GENETICS



Ying Liu
City College of San Francisco

LibreTexts™

Introduction to Genetics

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 05/03/2026

TABLE OF CONTENTS

Licensing

1: DNA, Genes, Chromosomes and Genome

- 1.1: Introduction
- 1.2: DNA is the Genetic Material
- 1.3: DNA Structure
- 1.4: Genes are the Basic Units of Inheritance
- 1.5: Genes and Genomes
- 1.6: Model Organisms Facilitate Genetic Advances
- 1.E: Overview, DNA, and Genes (Exercises)
- 1.S: Overview, DNA, and Genes (Summary)

2: Cells and Mitosis

- 2.1: Introduction
- 2.2: Studying Cells
- 2.3: DNA is Packaged into Chromatin
- 2.4: The Cell Cycle and Changes in DNA Content
- 2.5: Karyotypes Describe Chromosome Number and Structure
- 2.6: Key Words and Terms

3: Meiosis - Sexual Reproduction

- 3.1: Overview of Meiosis
- 3.2: Meiosis I
- 3.3: Meiosis II
- 3.4: Meiosis
- 3.5: Comparing Meiosis and Mitosis
- 3.6: Errors in Meiosis

4: Mendelian Genetics

- 4.1: Introduction
- 4.2: Introduction
- 4.3: Mendel's Experiments
- 4.4: Punnett Squares and Test Crosses
- 4.5: Laws of Inheritance
- 4.6: Mendel's First Law
- 4.7: Relationships Between Genes, Genotypes and Phenotypes
- 4.8: Pedigree Analysis
  - 4.8.1: Symbols used in Pedigree Charts
  - 4.8.2: Modes of Inheritance
  - 4.8.3: Sporadic and Non-Heritable Diseases
  - 4.8.4: Calculating Probabilities Using Pedigree Charts
  - 4.8.5: Chapter 4 Summary
  - 4.8.6: Key Terms in Chapter 4
  - 4.8.7: Chapter 4 Study Questions

- 4.8.8: Chapter 4 Answers
- 4.9: Pedigrees and Punnett Squares
- 4.10: Biochemical Basis of Dominance
- 4.11: Black fur color - a dominant trait
- 4.12: Yellow fur color - a recessive trait
- 4.13: Sex-Linkage- An Exception to Mendel's First Law
- 4.14: Phenotypes May Not Be As Expected from the Genotype
- 4.15: Phenotypic Ratios May Not Be As Expected
- 4.16: Extensions of the Laws of Inheritance
- 4.E: Genetic Analysis of Single Genes (Exercises)
- 4.S: Genetic Analysis of Single Genes (Summary)

5: Genetic Analysis of Multiple Genes

- 5.1: Dihybrid Crosses
- 5.2: Epistasis and Other Gene Interactions
- 5.3: Example of Multiple Genes Affecting One Character
- 5.4: Alleles at a Single Locus
  - 5.4.1: Introduction
  - 5.4.2: Somatic vs. Germline Mutations
  - 5.4.3: Alleles- Hetero-, Homo-, Hemizygosity
  - 5.4.4: Pleiotropy vs. Polygenic Inheritance
  - 5.4.5: Types of Dominance
  - 5.4.6: Biochemical Basis of Dominance
  - 5.4.7: Classification of Mutants
  - 5.4.8: Muller's Morphs
  - 5.4.9: Chapter 6 Summary
  - 5.4.10: Key Terms in Chapter 6
  - 5.4.11: Chapter 6 Study Questions
  - 5.4.12: Chapter 6 Answers
- 5.E: Genetic Analysis of Multiple Genes (Exercises)
- 5.S: Genetic Analysis of Multiple Genes (Summary)

6: Sex Chromosomes Sex Linkage

- 6.1: Introduction
- 6.2: Autosomes and Sex Chromosomes
- 6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
- 6.4: Sex Linkage- An Exception to Mendel's First Law
- 6.5: Y-Linked Genes
- 6.6: Chapter 10 Summary
- 6.7: Key Terms in Chapter 10
- 6.8: Chapter 10 Study Questions
- 6.9: Chapter 10 Answers

7: DNA Structure and Function

- 7.1: Introduction
- 7.2: Historical Basis of Modern Understanding
- 7.3: DNA Structure and Sequencing
- 7.4: Basics of DNA Replication

- 7.5: DNA Replication in Prokaryotes
- 7.6: DNA Replication in Eukaryotes
- 7.7: DNA Repair
- 7.8: Key Terms
- 7.9: Chapter Summary
- 7.10: Visual Connection Questions
- 7.11: Review Questions
- 7.12: Critical Thinking Questions

8: Genes and Proteins

- 8.1: Introduction
- 8.2: The Genetic Code
- 8.3: Prokaryotic Transcription
- 8.4: Eukaryotic Transcription
- 8.5: RNA Processing in Eukaryotes
- 8.6: Ribosomes and Protein Synthesis
- 8.7: Gene Expression
  - 8.7.1: Introduction
  - 8.7.2: Regulation of Gene Expression
  - 8.7.3: Prokaryotic Gene Regulation
  - 8.7.4: Eukaryotic Epigenetic Gene Regulation
  - 8.7.5: Eukaryotic Transcription Gene Regulation
  - 8.7.6: Eukaryotic Post-transcriptional Gene Regulation
  - 8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
  - 8.7.8: Cancer and Gene Regulation
  - 8.7.9: Key Terms
  - 8.7.10: Chapter Summary
  - 8.7.11: Visual Connection Questions
  - 8.7.12: Review Questions
  - 8.7.13: Critical Thinking Questions
- 8.8: Key Terms
- 8.9: Chapter Summary
- 8.10: Visual Connection Questions
- 8.11: Regulation of Gene Expression
  - 8.11.1: The lac Operon
  - 8.11.2: The Use of Mutants to Study the lac Operon
  - 8.11.3: Eukaryotic Gene Regulation
  - 8.11.4: Regulatory Elements in Evolution
  - 8.11.5: Additional Levels of Regulating Transcription
  - 8.11.6: Epigenetics
  - 8.11.7: Regulation of Gene Expression (Exercises)
  - 8.11.8: Gene Expression Essentials
  - 8.11.S: Regulation of Gene Expression (Summary)
- 8.12: Review Questions
- 8.13: Critical Thinking Questions

9: Mutation and Variation

- 9.1: Mutation and Polymorphism
- 9.2: Origins of Mutations
- 9.3: Genetic screening for mutations- Forward genetics
- 9.4: Types of Mutations
- 9.5: Some mutations may not have detectable phenotypes
- 9.6: Complementation tests and Allelism
- 9.7: Example of Human Mutations
- 9.8: Linkage and Mapping
  - 9.8.1: Linkage
  - 9.8.2: Recombination
  - 9.8.3: Linkage Reduces Recombination Frequency
  - 9.8.4: Crossovers Allow Recombination of Linked Loci
  - 9.8.5: Inferring Recombination From Genetic Data
  - 9.8.6: Genetic Mapping
  - 9.8.7: Mapping With Three-Point Crosses
  - 9.8.E: Linkage and Mapping (Exercises)
  - 9.8.S: Linkage and Mapping (Summary)
- 9.E: Mutation and Variation (Exercises)
- 9.S: Mutation and Variation (Summary)

10: Changes in Chromosome Number and Structure

- 10.1: Changes in Chromosome Number
- 10.2: Changes in Chromosome Structure
- 10.3: Chromosome Abnormalities in Humans
- 10.4: Diagnosing Human Chromosome Abnormalities
- 10.E: Changes in Chromosome Number and Structure (Exercises)
- 10.S: Changes in Chromosome Number and Structure (Summary)

11: Pedigrees and Populations

- 11.1: Prelude to Pedigrees and Populations
- 11.2: Pedigree Analysis
- 11.3: Inferring the Mode of Inheritance
- 11.4: Sporadic and Non-Heritable Diseases
- 11.5: Calculating Probabilities
- 11.6: Population Genetics
- 11.E: Pedigrees and Populations (Exercises)
- 11.S: Pedigrees and Populations (Summary)

12: Techniques of Molecular Genetics

- 12.1: Prelude to Molecular Genetics
- 12.2: Isolating Genomic DNA
- 12.3: Isolating or Detecting a Specific Sequence by PCR
- 12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
- 12.5: Cloning DNA - Plasmid Vectors
- 12.6: DNA Analysis - Gel Electrophoresis
- 12.7: DNA Analysis- Blotting and Hybridization
- 12.8: Transgenic organisms

- 12.9: Techniques of Molecular Genetics (Exercises)
- 12.S: Techniques of Molecular Genetics (Summary)

13: Cancer Genetics

- 13.1: Classification of Cancers
- 13.2: Cancer Cell Biology
- 13.3: Hallmarks of Cancer
- 13.4: Mutagens and Carcinogens
- 13.5: Oncogenes
- 13.6: Tumor Suppressor Genes
- 13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
- 13.8: Genes and COVID-19 Susceptibility in Humans
  - 13.8.1: Introduction
  - 13.8.2: How Does SARS-Cov-2 Infect Humans?
  - 13.8.3: Genes Implicated in Severe COVID-19 Infection in Humans
  - 13.8.4: Approaches for Vaccine Development Against SARS-Cov-2
  - 13.8.5: Chapter 13 Summary
  - 13.8.6: Key Terms in Chapter 13
- 13.E: Cancer Genetics (Exercises)
- 13.S: Cancer Genetics (Summary)

14: Genomics and Systems Biology

- 14.1: ‘Omics Technologies
- 14.2: DNA Sequencing
- 14.3: Whole Genome Sequencing
- 14.4: Functional Genomics – Determining Function(s)
- 14.E: Genomics and Systems Biology (Exercises)
- 14.S: Genomics and Systems Biology (Summary)

15: Molecular Markers and Quantitative Traits

- 15.1: Some Variations in the Genome Affect Complex Traits
- 15.2: Origins of Molecular Polymorphisms
- 15.3: Classification and Detection of Molecular Markers
- 15.4: Applications of Molecular Markers
- 15.5: Quantitative Trait Locus (QTL) Analysis
- 15.E: Molecular Markers and Quantitative Traits (Exercises)
- 15.S: Molecular Markers and Quantitative Traits (Summary)

16: Appendix 1 – Simple Nomenclature

- 16.1: Textbook Introduction
- 16.2: Why Use Nomenclature?
- 16.3: Biochemistry of Gene Behavior
- 16.4: Basic Nomenclature
- 16.5: Three-Letter Symbols
- 16.6: Linked Genes
- 16.7: Summary

Index

Glossary

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

1: DNA, Genes, Chromosomes and Genome

The molecular foundation of all living organisms lies in **DNA** — **deoxyribonucleic acid** — a remarkable molecule that encodes the instructions needed for growth, development, and reproduction. DNA is organized into **genes**, the functional units of heredity, which guide the production of proteins that determine the structure and function of cells. These genes are meticulously arranged on **chromosomes**, thread-like structures within the cell nucleus that ensure accurate distribution of genetic information during cell division.

Collectively, an organism's complete set of DNA, including all its genes, is known as its **genome**. The human genome, for instance, spans over 3 billion base pairs and contains about 20,000-25,000 genes distributed across 23 pairs of chromosomes. Understanding the interplay between DNA, genes, chromosomes, and the genome is crucial for exploring the complexity of genetic inheritance, disease mechanisms, and the diversity of life. This chapter delves into the intricate architecture and function of these genetic elements, providing a foundation for the study of molecular biology and genetics.

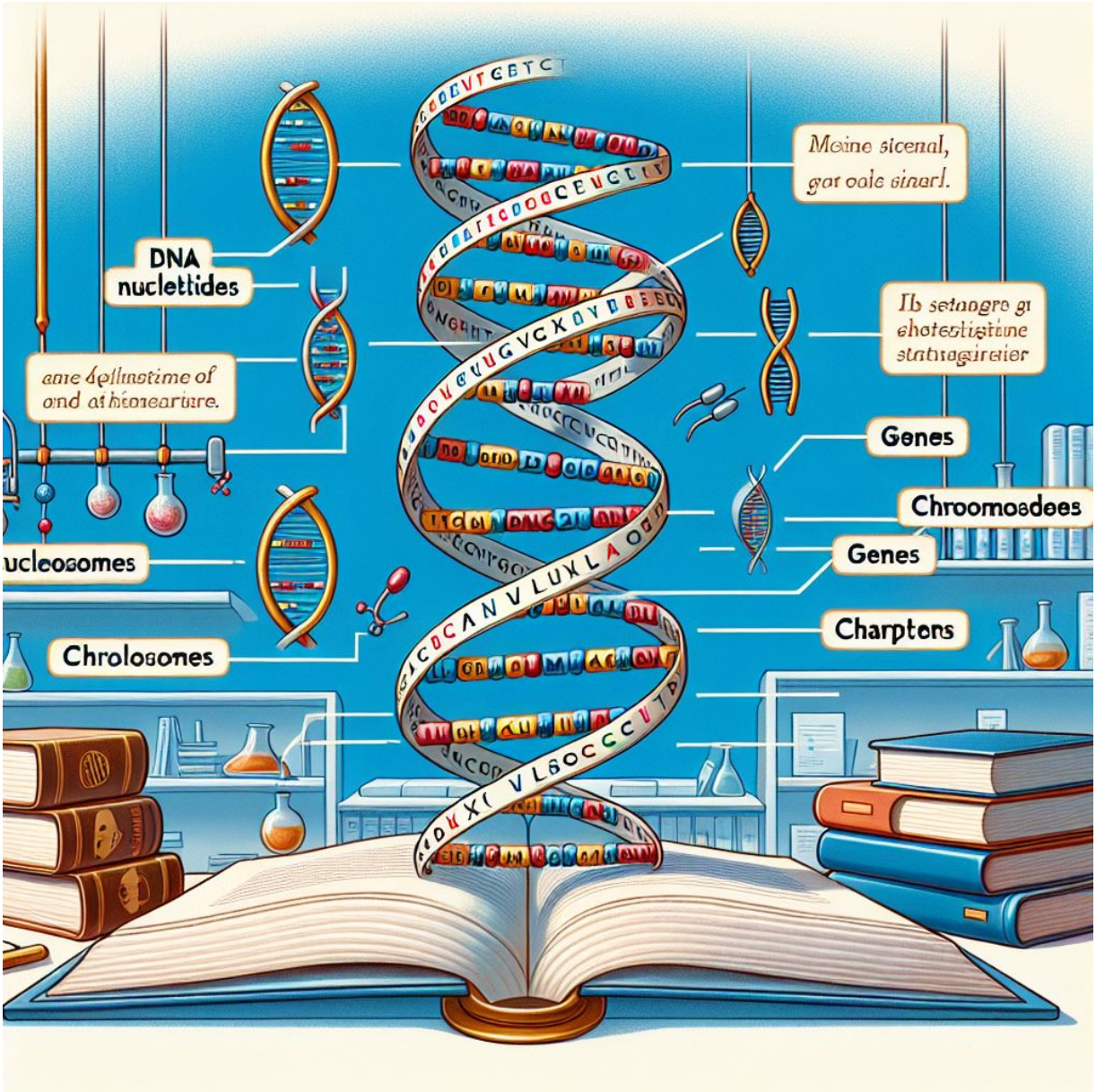


Figure 1.1: Genome, chromosome and genes, the book analogy. All the DNA in your cell makes up your genome, an equivalent of a book. The DNA in a human cell is organized in 46 chromosomes, or chapters. Each chromosome contains hundreds to thousands of genes, or sentences. And DNA is a long polymer made of 4 different nucleotides, or letter. (Copyright; author via source)

This page titled 1: DNA, Genes, Chromosomes and Genome is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

1.1: Introduction

Learning Objectives

- Define genes, genome, chromosomes and karyotype
- Explain how DNA carries and maintains genetic information
- Describe DNA sequencing and genomics
- Discuss how genomics can be useful for fields such as conservation biology, forensic science and epidemiology

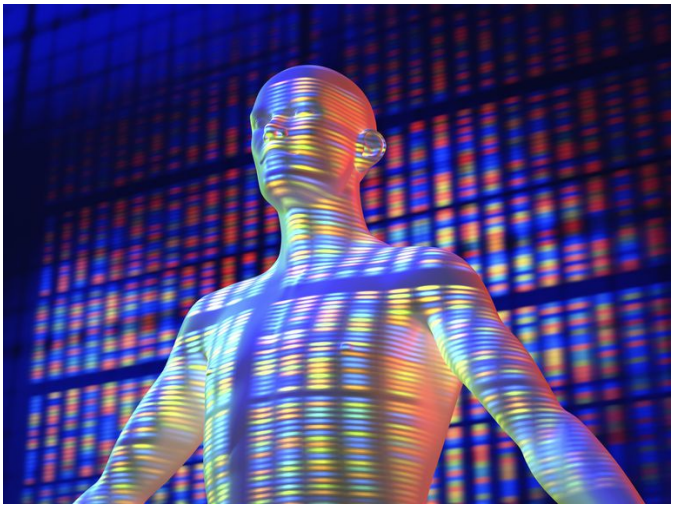


Figure 1.1.1: The Human Genome Project (HGP) made a map of the genes of a human being. (Copyright; author via source)

We are living in the “Post-genomic era” in which you can pay a company like “23 and Me” or “Ancestry” $100-200, spit into a tube, and in about two months, find out everything you need to know about your ancestry, genetic traits, risks for getting certain genetic diseases, etc. These advances would not be possible without the efforts of scientists across the world studying human genetics and genomics. The US National Institute of Health (NIH) and the US Department of Energy (DOE) launched the Human Genome Project (HGP) in 1990, an international consortium with participating scientists from the UK, France, Australia, China, Germany, etc. aimed at sequencing the entire human genome. Scientists started out mapping the human genome (Figure 1), by determining the locations of many genes on the 23 haploid chromosomes. They then determined the nucleotide sequences of the genes (called coding sequences) and other non-coding sequences in the genome. The project took 13 years (completed in 2003), and provided the first blueprint of the human genome.

Genetics – is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Heredity – Humans have always been aware that the characteristics of an individual plant or animal in a population could be passed down through the generations. Offspring look more like their parents.Humans also knew that some heritable characteristics (such as the size or color of fruit) varied between individuals, and that they could select or breed crops and animals for the most favorable traits. Knowledge of these hereditary properties has been of significant value in the history of human development. In the past, humans could only manipulate and select from naturally existing combinations of genes. More recently, with the discovery of the substance and nature of genetic material, DNA, we can now identify, clone, and create novel, better combinations of genes that will serve our goals. Understanding the mechanisms of genetics is fundamental to using it wisely and for the betterment of all.



Fig.1-1: *Parent and offspring. Wolf's Monkey. (Flickr-eclectic echos-CC:AND)*



This page titled 1.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **1.1: Overview** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

1.2: DNA is the Genetic Material

Microbiologists identified two strains of the bacterium *Streptococcus pneumoniae*. The R-strain produced rough colonies on a bacterial plate, while the other S-strain was smooth (Figure 1.2.2). More importantly, the S-strain bacteria caused fatal infections when injected into mice, while the R-strain did not (top, Figure 1.2.3). Neither did "heat-treated" S-strain cells. **Griffith** in 1929 noticed that upon mixing "heat-treated" S-strain cells together with some R-type bacteria (neither should kill the mice), the mice died and there were S-strain, pathogenic cells recoverable. Thus, some non-living component from the S-type strains contained genetic information that could be transferred to and **transform** the living R-type strain cells into S-type cells.

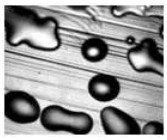


Figure 1.2.2: Colonies of Rough (top) and Smooth (bottom) strains of S. pneumoniae. (J. Exp.Med.98:21, 1953-R. Austrian-Pending)

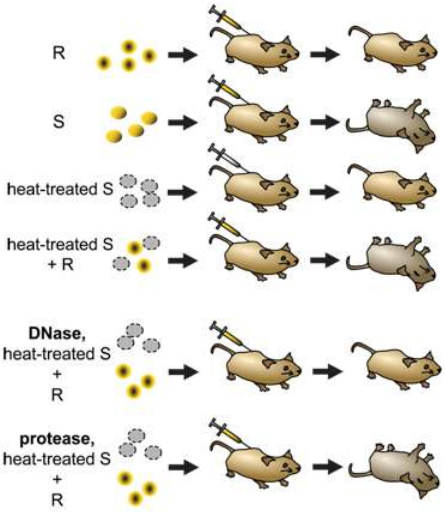


Figure 1.2.3: Experiments of Griffith and of Avery , MacLeod and McCarty. R strains of S. pneumoniae do not cause lethality. However, DNA-containing extracts from pathogenic S strains are sufficient to make R strains pathogenic. (Original-Deyholos-CC:AN)

This page titled 1.2: DNA is the Genetic Material is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **1.2: DNA is the Genetic Material** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

1.3: DNA Structure

A. Early Clues and Ongoing Misconceptions

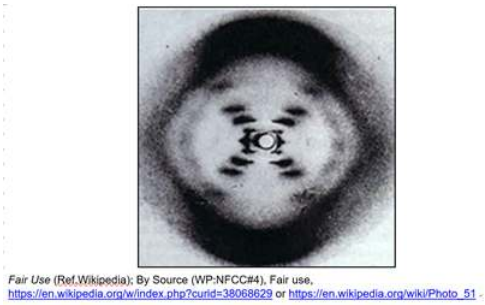
By 1878, a substance in the pus of wounded soldiers derived from cell nuclei (called *nuclein*) was shown to be composed of 5 bases (the familiar ones of DNA and RNA). The four bases known to make up DNA (as part of nucleotides) were thought to be connected through the phosphate groups in short repeating chains of four nucleotides. By the 1940s, we knew that DNA was a long polymer. Nevertheless, it was still considered too simple to account for genes (see above). After the Hershey and Chase experiments, only a few holdouts would not accept DNA as the genetic material. So, the question remaining was how such a "simple" molecule could account for all the genes, even in so simple an organism as a bacterium. The answer to this question was to lie at least in part in an understanding of the physical structure of DNA, made possible by the advent of ***X-Ray Crystallography***.

If a substance can be crystallized, the crystal will diffract X-rays at angles revealing regular (repeating) structures of the crystal. William Astbury demonstrated that high molecular weight DNA had just such a regular structure. His ***crystallographs*** suggested DNA to be a linear polymer of stacked bases (nucleotides), each nucleotide separated from the next by 0.34 nm. Astbury is also remembered for coining the term "*molecular biology*" to describe his studies. The term now covers as all aspects of biomolecular structure, as well as molecular functions (e.g. replication, transcription, translation, gene regulation…).

In an irony of history, the Russian biologist Nikolai Koltsov had already intuited in 1927 that the basis of genetic transfer of traits would be a "giant hereditary molecule" made up of "two mirror strands that would replicate in a semi-conservative fashion using each strand as a template". A pretty fantastic inference if you think about it since it was proposed long before Watson and Crick and their colleagues worked out the structure of the DNA double-helix!

B. Wilkins, Franklin, Watson & Crick

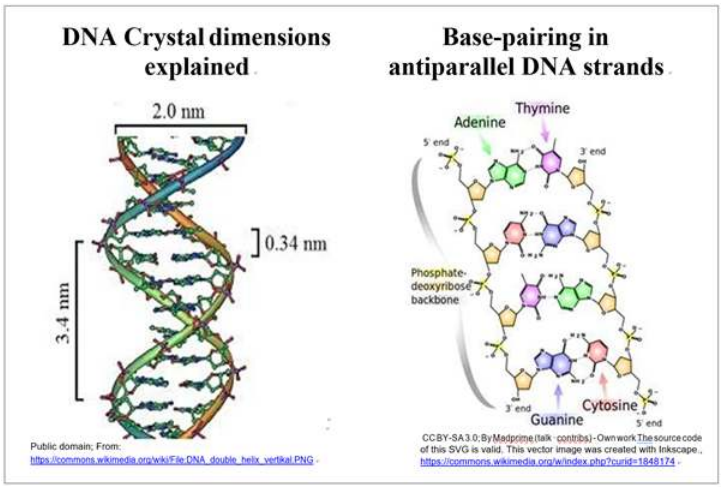
Maurice Wilkins, an English biochemist, was the first to isolate highly pure, high molecular weight DNA. Working in Wilkins laboratory, Rosalind Franklin was able to crystalize this DNA and produce very high-resolution X-Ray diffraction images of the DNA crystals. Franklin's most famous (and definitive) crystallography was "Photo 51", shown below.



Fair Use (Ref.Wikipedia); By Source (WP:NFCC#4), Fair use, https://en.wikipedia.org/w/index.php?curid=38068629 or https://en.wikipedia.org/wiki/Photo_51

This image confirmed Astbury's **0.34 nm** repeat dimension and revealed two more numbers, **3.4 nm** and **2 nm**, reflecting additional repeat structures in the DNA crystal. When James Watson and Francis Crick got hold of these numbers, they used them along with other data to build DNA models out of nuts, bolts and plumbing. Their models eventually revealed DNA to be a pair of **antiparallel complementary** of nucleic acid polymers…, shades of Koltsov's mirror-image macromolecules! Each strand is a string of nucleotides linked by *phosphodiester bonds*, the two strands held together in a *double helix* by complementary H-bond interactions.

Let's look at the evidence for these conclusions and as we do, refer to the two illustrations of the double helix below.



Recalling that Astbury's 0.34 nm dimension was the ***distance between successive nucleotides*** in a DNA strand, Watson and Crick surmised that the 3.4 nm repeat was a structurally meaningful 10-fold multiple of Astbury's number. When they began building their DNA models, they realized from the bond angles connecting the nucleotides that the strand was forming a helix, from which they concluded that the 3.4 nm repeat was the ***pitch*** of the helix, i.e., the distance of one complete turn of the helix. This meant that there were 10 bases per turn of the helix. They further reasoned that the 2.0 nm number might reflect the diameter of helix. When their scale model of a single stranded DNA helix predicted a helical diameter much less than 2.0 nm, they were able to model a ***double helix*** that more nearly met the 2.0 nm diameter requirement. In building their double helix, Watson and Crick realized that bases in opposing strands would come together to form H-bonds, holding the helix together. However, for their double helix to have a constant diameter of 2.0 nm, they also realized that the smaller **pyrimidine** bases, ***Thymine*** (**T**) and ***Cytosine*** (**C**), would have to H-bond to the larger **purine** bases, ***Adenine*** (**A**) and ***Guanosine*** (**G**).

Now to the question of how a "simple" DNA molecule could have the structural diversity needed to encode thousands of different polypeptides and proteins. In early studies, purified *E. coli* DNA was chemically hydrolyzed down to nucleotide monomers. The hydrolysis products contained nearly equal amounts of each base, reinforcing the notion that DNA was that simple molecule that could not encode genes. But Watson and Crick had private access to revealing data from Erwin Chargaff. Chargaff had determined the base composition of DNA isolated from different species, including *E. coli*. He found that the base composition of DNA from different species was not always *equimolar*, meaning that for some species, the DNA was not composed of equal amounts of each of the four bases (see some of this data below).

Base Compositions of DNA from Different Organisms

base		*human*	*yeast*	*fly*
purine	A	27%	21%	15%
pyrimidine	C	23%	32%	35%
purine	G	23%	32%	35%
pyrimidine	T	27%	21%	15%

The mere fact that DNA from some species could have base compositions that deviated from equimolarity put to rest the argument that DNA had to be a very simple sequence. Finally, it was safe to accept that to accept the obvious, namely that DNA was indeed the "stuff of genes".

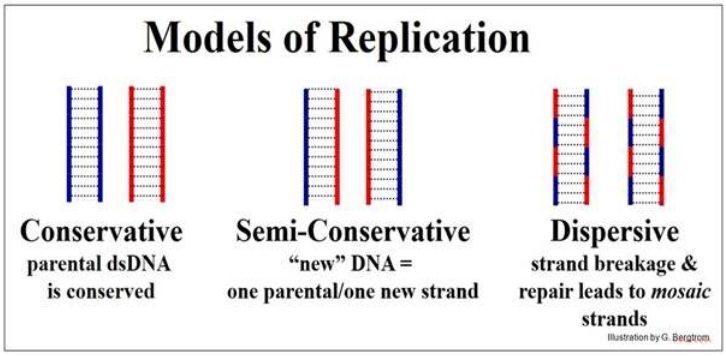
Chargaff's data also showed a unique pattern of base ratios. Although base compositions could vary between species, the **A/T** and **G/C** *ratio* was always one, for every species. Likewise the ratio of **(A+C)/(G+T)** and **(A+G)/(C+T)**. From this information, Watson and Crick inferred that **A** (a purine) would H-bond with **T (a pyrimidine)**, and **G** (a purine) would H-bond with **C** (a pyrimidine) in the double helix. When building their model with this new information, they also found H-bonding between the complementary bases would be maximal only if the two DNA strands were **antiparallel**, leading to the most stable structure of the double helix.

Watson and Crick published their conclusions about the structure of DNA in 1953 (Click here to read their seminal article: Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Their article is also famous for predicting a semi-conservative mechanism of replication, something that had been predicted by Koltsov 26 years earlier, albeit based on intuition… and much less evidence!

Watson, Crick and Wilkins shared a Nobel Prize in 1962 for their work on DNA structure. Unfortunately, Franklin died in 1958 and Nobel prizes were not awarded posthumously. There is still controversy about why Franklin did not get appropriate credit for her role in the work. But she has been getting well-deserved, long-delayed recognition, including a university in Chicago named in her honor!

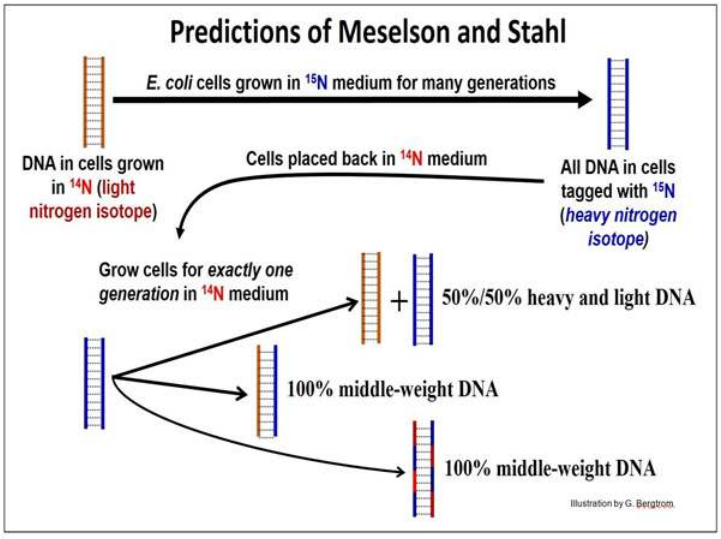
169 Unraveling the Structure of DNA

Confirmation of Watson & Crick's suggestion of semiconservative replication came from Meselson and Stahl's very elegant experiment, which tested the three possible models of replication shown below.



In their experiment, *E. coli* cells were grown for in medium containing 15N, a 'heavy' nitrogen isotope. After many generations, all of the DNA in the cells had become labeled with the heavy isotope. At that point, the 15N-tagged cells were placed back in medium containing the more common, 'light' 14N isotope and allowed to grow for exactly one generation.

Meselson and Stahl's predictions and their experimental design are shown below.



Meselson and Stahl knew that 14N-labeled and 15N-labeled DNA would form separate bands after centrifugation on *CsCl chloride density gradients*. They tested their predictions by purifying and centrifuging the DNA from the 15N-labeled cells grown in 14N medium for one generation. They found that this DNA formed a single band with a density between that of 15N-labeled DNA and 14N-labeled DNA. This result *eliminated* a conservative model of DNA replication (as Watson and Crick also predicted. That left two possibilities: replication was either semiconservative or dispersive. The dispersive model was *eliminated* when DNA isolated from cells grown for a 2nd generation on 14N were shown to contain two bands of DNA on the CsCl density gradients.

Chromosomes

We understood from the start of the 20th century that chromosomes contained genes. Therefore, it becomes necessary to understand the relationship between chromosomes, chromatin, DNA and genes. As noted earlier, chromosomes are a specialized, *condensed* version of chromatin, with key structural features shown below.

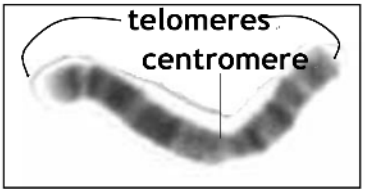
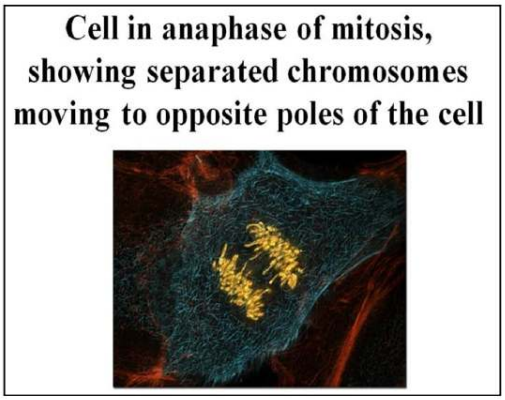


Illustration by G. Bergtrom

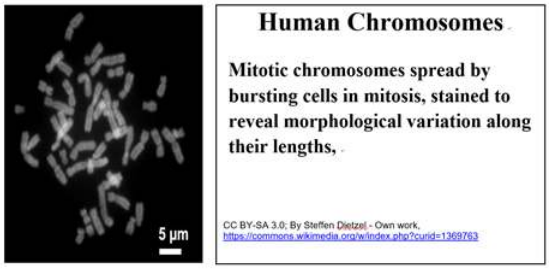
Public Domain. Adapted from National Human Genome Research Institute - Cropped from File:Human male karyotype high resolution.jpg. https://commons.wikimedia.org/w/index.php?curid=40985706

We now know that the compact structure of a chromosome prevents damage to the DNA during cell division. This damage can occur when forces on centromeres generated by mitotic or meiotic spindle fibers pull chromatids apart. As the nucleus breaks down during mitosis or meiosis, late 19th century microscopists saw chromosomes condense from the dispersed cytoplasmic background. These chromosomes remained visible as they separated, moving to opposite poles of the cell during cell division. Such observations of chromosome behavior during cell division pointed to their role in heredity. A computer- colorized cell in mitosis is shown below.

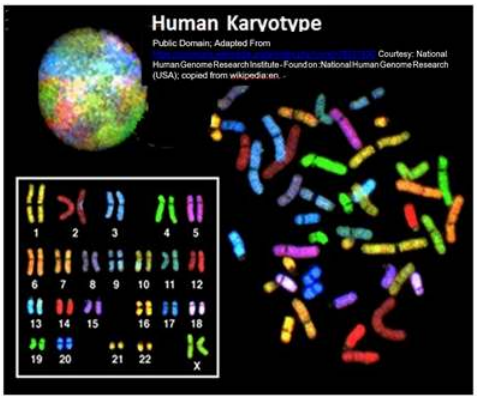


CC-BY-SA-NC; From: https://fascience.wikispaces.com/10+Cell+Growth+and+Division

It is possible to distinguish one chromosome from another by ***karyotyping***. When cells in metaphase of mitosis are placed under pressure, they burst and the chromosomes spread apart. Such a chromosome spread is shown below.



By the early 1900s, the number, sizes and shapes of chromosomes were shown to be species-specific. What's more, a close look at chromosome spreads revealed that chromosomes came in morphologically matched pairs. This was so reminiscent of Gregor Mendel's paired hereditary factors that chromosomes were then widely accepted as the structural seat of genetic inheritance. Cutting apart micrographs like the one above and pairing the chromosomes by their morphology generates a ***karyotype***. Paired human homologs are easily identified in the colorized micrograph below.



Captured in mitosis, all dividing human cells contain 23 pairs of homologous chromosomes. The ***karyotype*** is from a female; note the pair of homologous sex ("X") chromosomes (lower right of the inset). X and Y chromosomes in males are not truly homologous. Chromosomes in the original spread and in the aligned karyotype stained with fluorescent antibodies to chromosome-specific DNA sequences, 'light up' the different chromosomes.

171 DNA, Chromosomes, Karyotypes & Gene Maps

This page titled 1.3: DNA Structure is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **8.3: DNA Structure** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

1.4: Genes are the Basic Units of Inheritance

Blending vs Particulate inheritance

The once prevalent (but now discredited) concept of **blending inheritance** proposed that some undefined essence, in its entirety, contained all of the heritable information for an individual. It was thought that mating combined the essences from each parent, much like the mixing of two colors of paint. Once blended together, the individual characteristics of the parents could not be separated again. However, **Gregor Mendel** (Fig 1.10) was one of the first to take a quantitative, scientific approach to the study of heredity.



Figure 1.4.10: Gregor Mendel. (Original-unknown-PD)

He started with well-characterized strains, repeated his experiments many times, and kept careful records of his observations. Working with peas, Mendel showed that white-flowered plants could be produced by crossing two purple-flowered plants, but only if the purple-flowered plants themselves had at least one white-flowered parent (Fig 1.11). This was evidence that the genetic factor that produced white-flowers had *not* blended irreversibly with the factor for purple-flowers. Mendel's observations disprove blending inheritance and favor an alternative concept, called **particulate inheritance**, in which heredity is the product of discrete factors that control independent traits.

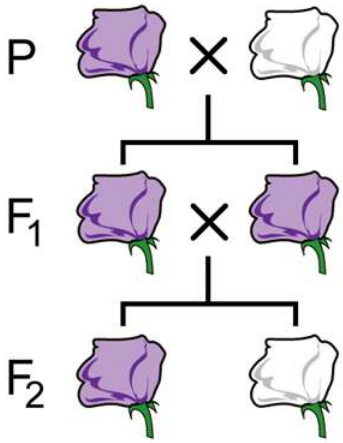


Figure 1.4.11: Inheritance of flower color in peas. Mendel observed that a cross between pure breeding, white and purple peas (generation P) produced only progeny (generation F_1) with purple flowers. However, white flowered plant reappeared among the F_2 generation progeny of a mating between two F_1 plants. The symbols P, F_1 and F_2 are abbreviations for parental, first filial, and second filial generations, respectively. (Original-Deyholos-CC:AN)

This page titled 1.4: Genes are the Basic Units of Inheritance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **1.4: Genes are the Basic Units of Inheritance** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

1.5: Genes and Genomes

Source: BiochemFFA_7_1.pdf. The entire textbook is available for free from the authors at http://biochem.science.oregonstate.edu/content/biochemistry-free-and-easy



Introduction

For many years, scientists wondered about the nature of the information that directed the activities of cells. What kind of molecules carried the information, and how was the information passed on from one generation to the next? Key experiments, done between the 1920s and the 1950s, established convincingly that this genetic information was carried by DNA. In 1953, with the elucidation of the structure of DNA, it was possible to begin investigating how this information is passed on, and how it is used.

Genomes

We use the word “genome” to describe all of the genetic material of the cell. That is, a genome is the entire sequence of nucleotides in the DNA that is in all of the chromosomes of a cell. When we use the term genome without further qualification, we are generally referring to the chromosomes in the nucleus of a eukaryotic cell. As you know, eukaryotic cells have organelles like mitochondria and chloroplasts that have their own DNA (Figure 7.1 & 7.2). These are referred to as the mitochondrial or chloroplast genomes to distinguish them from the nuclear genome.

Starting in the 1980s, scientists began to determine the complete sequence of the genomes of many organisms, in the hope of better understanding how the DNA sequence specifies cellular functions. Today, the complete genome sequences have been determined for thousands of species from all domains of life, and many more are in the process of being worked out by groups of scientists across the world.

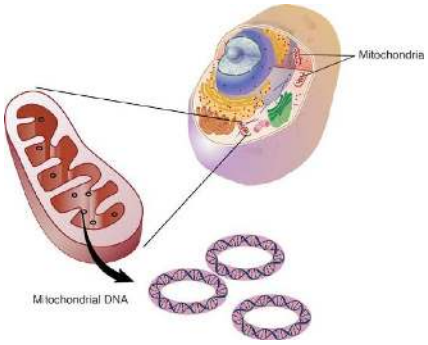


Figure 7.1 - Mitochondria carry their own genome

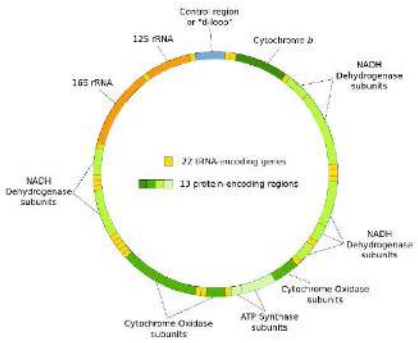


Figure 7.2 - The human mitochondrial genome - Wikipedia

Global genome initiative

The Global Genome Initiative, a collaborative effort to sequence at least one species from each of the 9,500 described invertebrate, vertebrate, and plant families is one of many such ventures. The information from these various efforts is collected in enormous online repositories, so that it is freely available to scientists. As the sequence databases compile ever more information, the fields of computational biology and bioinformatics have arisen, to analyze and organize the data in a way that helps biologists understand what the information in DNA means in the cellular context.

Genes

It has been known for many years that phenotypic traits are controlled by specific regions of the DNA that were termed “genes”. Thus, DNA was envisioned as a long string of nucleotides, in which certain regions, the genes, were separated by non-coding regions that were simply referred to as intergenic sequences (inter=between; genic=of genes). Early experiments in molecular biology suggested a simple relationship between the DNA sequence of a gene and its product, and led scientists to believe that each gene carried the information for a single protein. Changes, or mutations in the base sequence of a gene would be reflected in changes in the gene product, which in turn, would manifest itself in the phenotype or observable trait. This simple picture, while still useful, has been modified by subsequent discoveries that demonstrated that the use of genetic information by cells is somewhat more complicated. Our definition of a gene is also evolving to take new knowledge into consideration.

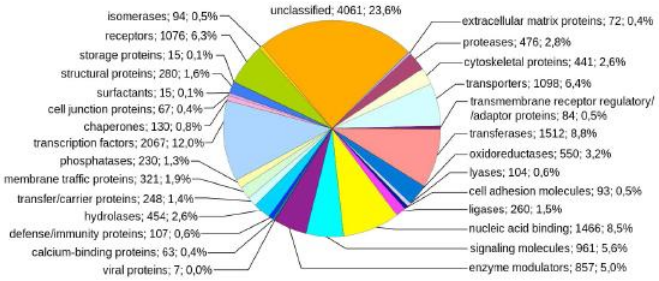


Figure 7.3 - From genomes to genes - Wikipedia

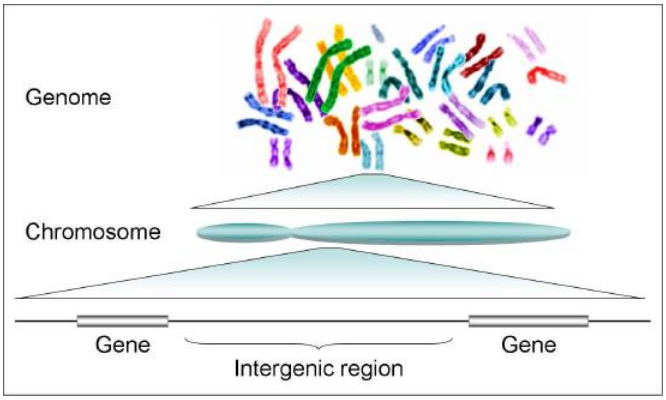


Figure 7.4 - Human genes sorted by class

Matters of size

A common-sense assumption about genomes would be that if genes specify proteins, then the more proteins an organism made, the more genes it would need to have, and thus, the larger its genome would be. Comparison of various genomes shows, surprisingly, that there is not necessarily a direct relationship between the complexity of an organism and the size of its genome (Figure 7.5). To understand how this could be true, it is necessary to recognize that while genes are made up of DNA, all DNA does not consist of genes (for purposes of our discussion, we define a gene as a section of DNA that encodes an RNA or protein product). In the human genome, less than 2% of the total DNA seems to be the sort of coding sequence that directs the synthesis of proteins. For many years, non-coding DNA in genomes was believed to be useless, and was described as "junk DNA" although it was perplexing that there seemed to be so much "useless" sequence. Recent discoveries have, however, demonstrated that much of this so-called junk DNA may play important roles in evolution, as well as in regulation of gene expression.

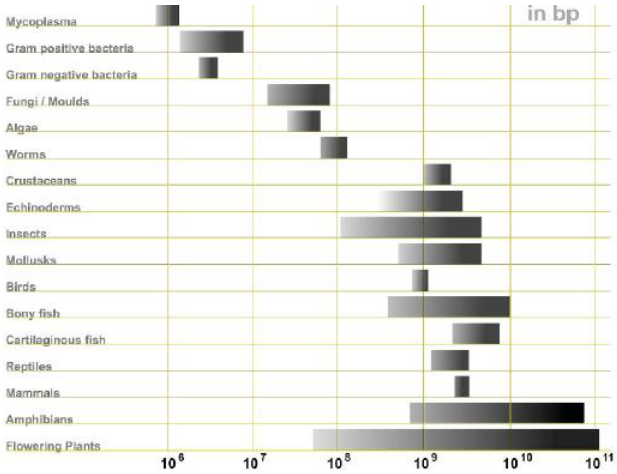


Figure 7.5 - Sizes of various genomes - Wikipedia

Introns

So, what is all the non-coding DNA doing there? We know that even coding regions in our DNA are interrupted by non-coding sequences called introns. This is true of most eukaryotic genomes. An examination of genes in eukaryotes shows that non-coding intron sequences can be much longer than the coding sections of the gene, or exons. Most exons are relatively small, and code for fewer than a hundred amino acids, while introns can vary in size from several hundred base-pairs to many kilobase-pairs (thousands of base-pairs) in length. For many genes in humans, there is much more of intron sequence than coding (a.k.a. exon) sequence. Intron sequences account for roughly a quarter of the genome in humans.

Other non-coding sequences

What other kinds of non-coding sequences are there? One function for some DNA sequences that do not encode RNA or proteins is in specifying when and to what extent a gene is used, or expressed. Such regions of DNA are called regulatory regions and each gene has one or more regulatory sequences that control its expression. However, regulatory sequences do not account for all the rest of the DNA in our genomes, either.

Transposable sequences

Surprisingly, almost half of the human genome appears to consist of several kinds of repetitive sequences. Many of the repetitive sequences are known to be transposable elements (transposons), sections of DNA that can move around within the genome. Sometimes referred to as "jumping genes" these transposable elements can move from one chromosomal location to another, either through a simple "cut and paste" mechanism that cuts the sequence out of one region of the DNA and inserts it into another location, or through a process called retrotransposition involving an RNA intermediate.

LINES & SINES

There are millions of copies of each of two major classes of such transposable elements, the LINEs (Long Interspersed Elements) and SINEs (Short Interspersed Elements) in our genomes.

LINEs and SINEs are both a kind of transposable element called retrotransposons, sequences that are copied into RNA, then reverse transcribed back into DNA before being inserted into new locations. This movement is typically not sequence specific, meaning that the transposons can be inserted randomly in the genome, in many cases within coding regions. As might be expected, this can disrupt the function of the gene. Transposons may also insert within regulatory regions, and change the expression of the genes they control. As a major cause of mutation in genomes, transposons play an important role in evolution.

Finally, recent findings have shown that much of the genome is transcribed into RNAs, even though only about 2% encodes proteins. What are the RNAs that do not encode proteins? Ribosomal RNAs (Figure 7.7) and transfer RNAs, together with the small nuclear RNAs that function in splicing, account for some of these non-translated transcripts, but not all. The remaining RNAs are regulatory RNAs, small molecules that play an important role in regulating gene expression. As we understand more about genomes, it is becoming evident that the so-called “junk” DNA is anything but.

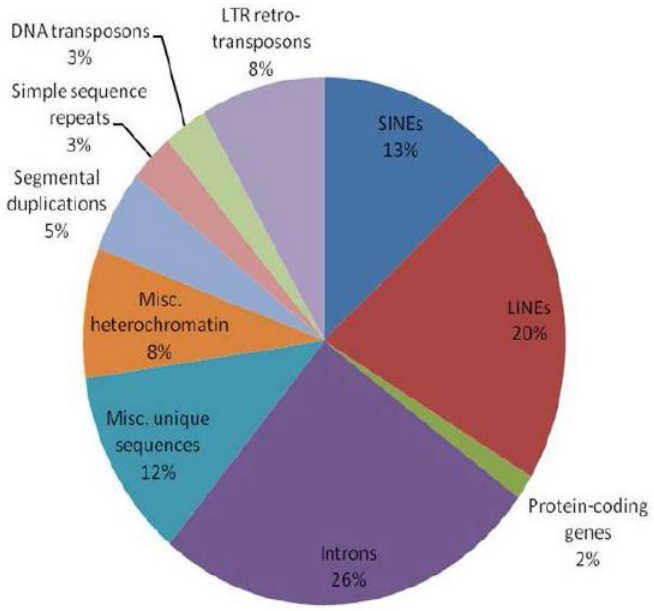


Figure 7.6 - Components of the human genome - Wikipedia

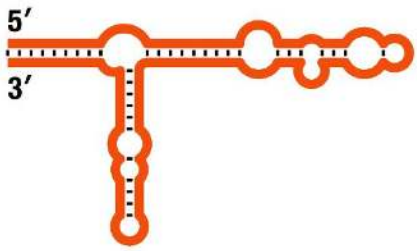


Figure 7.7 - 5S rRNA structure

This page titled 1.5: Genes and Genomes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.2: Genes and Genomes** by Kevin Ahern, Indira Rajagopal, & Taralyn Tan is licensed CC BY-NC-SA 4.0. Original source: http://biochem.science.oregonstate.edu/content/biochemistry-free-and-easy.

1.6: Model Organisms Facilitate Genetic Advances

Model organisms

Many of the great advances in genetics were made using species that are not especially important from a medical, economic, or even ecological perspective. Geneticists, from Mendel onwards, have sought the best organisms for their experiments. Today, a small number of species are widely used as **model organisms** in genetics (Fig 1.17). All of these species have specific characteristics that make large number of them easy to grow and analyze in laboratories: (1) they are small, (2) fast growing with a short generation time, (3) produce lots of progeny from matings that can be easily controlled, (4) have small genomes (small C-value), and (5) are diploid (i.e. chromosomes are present in pairs).

The most commonly used model organism are:

- The prokaryote bacterium, ***Escherichia coli,*** is the simplest genetic model organism and is often used to clone DNA sequences from other model species.
- Yeast (***Saccharomyces cerevisiae***) is a good general model for the basic functions of eukaryotic cells.
- The roundworm, ***Caenorhabditis elegans*** is a useful model for the development of multicellular organisms, in part because it is transparent throughout its life cycle, and its cells undergo a well-characterized series of divisions to produce the adult body.
- The fruit fly (***Drosophila melanogaster***) has been studied longer, and probably in more detail, than any of the other genetic model organisms still in use, and is a useful model for studying development as well as physiology and even behavior.
- The mouse (***Mus musculus***) is the model organism most closely related to humans, however there are some practical difficulties working with mice, such as cost, slow reproductive time, and ethical considerations.
- The zebrafish (***Danio rerio***) has more recently been developed by researchers as a genetic model for vertebrates.Unlike mice, zebrafish embryos develop quickly and externally to their mothers, and are transparent, making it easier to study the development of internal structures and organs.
- Finally, a small weed, ***Arabidopsis thaliana***, is the most widely studied plant genetic model organism. This provides knowledge that can be applied to other plant species, such as wheat, rice, and corn.



Figure 1.6.17: Some of the most important genetic model organisms in use today. Clockwise from top left: yeast, fruit fly, arabidopsis, mouse, roundworm, zebrafish. (Original/Flickr/Wikipedia – Deyholos/M.Westby/D.Joly/Z.F.Altun/Masur/Azul – CC:AS/ANS/AN &GFDL)

This page titled 1.6: Model Organisms Facilitate Genetic Advances is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **1.7: Model Organisms Facilitate Genetic Advances** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

1.E: Overview, DNA, and Genes (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

1.1 How would the results of the cross in Figure 1.11 have been different if heredity worked through blending inheritance rather than particulate inheritance?

1.2 Imagine that astronauts provide you with living samples of multicellular organisms discovered on another planet. These organisms reproduce with a short generation time, but nothing else is known about their genetics.

a) How could you define laws of heredity for these organisms?

b) How could you determine what molecules within these organisms contained genetic information?

c) Would the mechanisms of genetic inheritance likely be similar for all organisms from this planet?

d) Would the mechanisms of genetic inheritance likely be similar to organisms from earth?

1.3 It is relatively easy to extract DNA and protein from cells; biochemists had been doing this since at least the 1800's. Why then did Hershey and Chase need to use radioactivity to label DNA and proteins in their experiments?

1.4 Compare Watson and Crick's discovery with Avery, MacLeod and McCarty's discovery.

a) What did each discover, and what was the impact of these discoveries on biology?

b) How did Watson and Crick's approach generally differ from Avery, MacLeod and McCarty's?

c) Briefly research Rosalind Franklin on the internet. Why is her contribution to the structure of DNA controversial?

1.5 Starting with mice and R and S strains of *S. pneumoniae*, what experiments in additional to those shown in Figure 1.3 to demonstrate that DNA is the genetic material?

1.6 List the information that Watson and Crick used to deduce the structure of DNA.

1.7 Refer to Watson and Crick'

a) List the defining characteristics of the structure of a DNA molecule.

b) Which of these characteristics are most important to replication?

c) Which characteristics are most important to the Central Dogma?

1.8 Compare Figure 1.13 and Table 1.1. Which of the mutants (#1, #2, #3) shown in Figure 1.13 matches each of the phenotypes expected for mutations in genes A, B,C?

1.9 Refer to Table 1.2

a) What is the relationship between DNA content of a genome, number of genes, gene density, and chromosome number?

b) What feature of genomes explains the c-value paradox?

c) Do any of the numbers in Table 1.2 show a correlation with organismal complexity?

1.10 a) List the characteristics of an ideal model organism.

b) Which model organism can be used most efficiently to identify genes related to:

i) eye development

ii) skeletal development

iii) photosynthesis

iii) cell division

iv) cell differentiation

v) cancer

1.11 Refer to Figure 1.8

a) Identify the part of the DNA molecule that would be radioactively labeled in the manner used by Hershey & Chase

b) DNA helices that are rich in G-C base pairs are harder to separate (e.g. by heating) than A-T rich helices. Why?

This page titled 1.E: Overview, DNA, and Genes (Exercises) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

1.S: Overview, DNA, and Genes (Summary)

- Mendel demonstrated that heredity involved discrete, heritable factors that affected specific traits.
- A gene can be defined abstractly as a unit of inheritance.
- The ability of DNA from bacteria and viruses to transfer genetic information into bacteria demonstrated that DNA is the genetic material.
- DNA is a double helix made of two anti-parallel strands of bases on a sugar-phosphate backbone.
- Specific bases on opposite strands pair through hydrogen bonding, ensuring complementarity of the strands.
- The Central Dogma explains how DNA dictates heritable traits.
- Not all DNA in an organism contains genes.
- Model organisms accelerate the use of genetics in basic and applied research in biology, agriculture and medicine.

Key Terms

blending inheritance
particulate inheritance
Mendel
gene
allele
trait
P, F1, F2
Griffith
Avery, MacLeod, & McCarty
Hershey and Chase
Meselson & Stahl
DNase
proteinase
^{35}S
^{32}P
bacteriophage
semi-conservative
conservative
dispersive
E. coli
Nitrogen-14
Nitrogen-15
heavy vs light
CsCl gradient
Beadle & Tatum
auxotroph
prototroph
metabolic pathway
Neurospora crassa
Chargaff's Rules
Watson and Crick
DNA bases
sugar-phosphate backbone
anti-parallel
complementary
hydrogen bond
minor groove
major groove
adenine
cytosine
thymine
guanine
Central Dogma
transcription
reverse transcription
translation
RNA
prion
one-gene:one-enzyme
minimal medium
complete medium
arginine
genetic screen
nuclear genome
c-value paradox
model organism
Saccharomyces cerevisiae
Caenorhabditis elegans
Drosophila melanogaster
Mus musculus
Danio rerio
Arabidopsis thaliana
Escherichia coli

This page titled 1.S: Overview, DNA, and Genes (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **1.S: Overview, DNA, and Genes (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

2: Cells and Mitosis

2.1: Introduction
2.2: Studying Cells
2.3: DNA is Packaged into Chromatin
2.4: The Cell Cycle and Changes in DNA Content
2.5: Karyotypes Describe Chromosome Number and Structure
2.6: Key Words and Terms

Thumbnail: Life cycle of the cell. (CC BY-SA 4.0; BruceBlaus).

This page titled 2: Cells and Mitosis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

2.1: INTRODUCTION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Describe the role of cells in organisms
- Compare and contrast light microscopy and electron microscopy
- Summarize cell theory

Close your eyes and picture a brick wall. What is the wall's basic building block? It is a single brick. Like a brick wall, cells are the building blocks that make up your body.

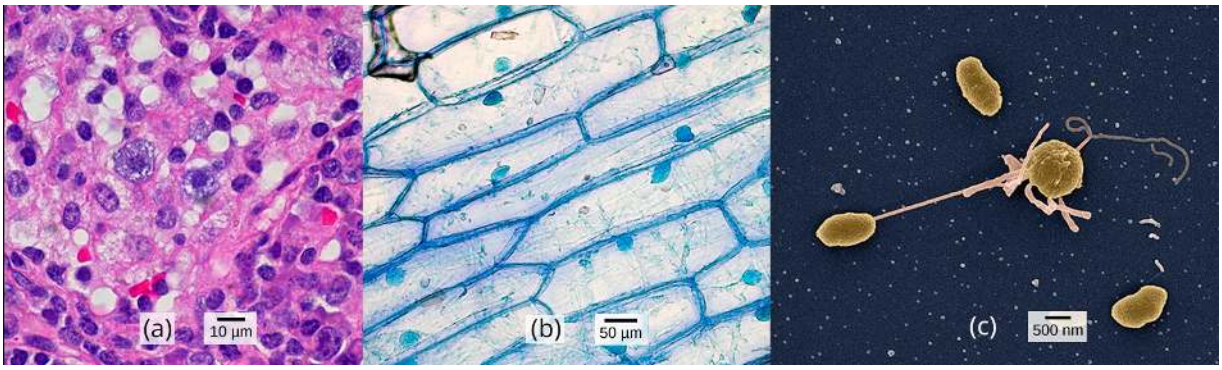


Figure **2.1.1**: Different cells. (a) Nasal sinus cells (viewed with a light microscope), (b) onion cells (viewed with a light microscope), and (c) *Vibrio tasmaniensis* bacterial cells (seen through a scanning electron microscope) are from very different organisms, yet all share certain basic cell structure characteristics. (credit a: modification of work by Ed Uthman, MD; credit b: modification of work by Umberto Salvagnin; credit c: modification of work by Anthony D'Onofrio, William H. Fowle, Eric J. Stewart, and Kim Lewis of the Lewis Lab at Northeastern University; scale-bar data from Matt Russell)

Your body has many kinds of cells, each specialized for a specific purpose. Just as we use a variety of materials to build a home, the human body is constructed from many cell types. For example, epithelial cells protect the body's surface and cover the organs and body cavities within. Bone cells help to support and protect the body. Immune system cells fight invading bacteria. Additionally, blood and blood cells carry nutrients and oxygen throughout the body while removing carbon dioxide. Each of these cell types plays a vital role during the body's growth, development, and day-to-day maintenance. In spite of their enormous variety, however, cells from all organisms—even ones as diverse as bacteria, onion, and human—share certain fundamental characteristics.

A cell is the basic unit for living organism. According to the Cell Theory, all living organisms are made of one or more cells. A bacterium or a yeast is made of one cell, and is called unicellular organism. A human being or an elephant is made of many cells, and is called a multicellular organism.

In the human body, there are more than 200 different types of cells, including muscle cells, nerve cells called neurons, bone cells and liver cells. Each cell type has specialized functions: muscle cells contract, allowing you to move your body; neurons conduct nerve impulse, allowing them to communicate with each other and other cells, etc. These mature cells that have specialized functions are called our body (somatic) cells.

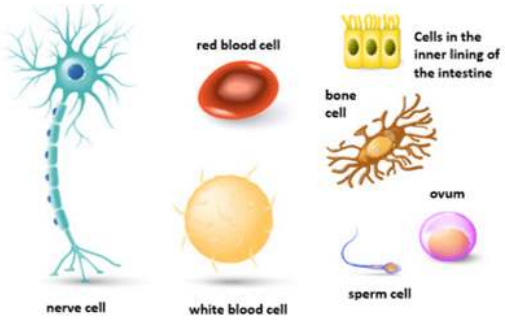


Figure **2.1.1**: Cells in the human body have different features to help them perform different functions. (Copyright; author via source)

This page titled 2.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

2.2: STUDYING CELLS

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Describe the role of cells in organisms
- Compare and contrast light microscopy and electron microscopy
- Summarize cell theory

A cell is the smallest unit of a living thing. Whether comprised of one cell (like bacteria) or many cells (like a human), we call it an organism. Thus, cells are the basic building blocks of all organisms.

Several cells of one kind that interconnect with each other and perform a shared function form tissues. These tissues combine to form an organ (your stomach, heart, or brain), and several organs comprise an organ system (such as the digestive system, circulatory system, or nervous system). Several systems that function together form an organism (like a human being). Here, we will examine the structure and function of cells.

There are many types of cells, which scientists group into one of two broad categories: prokaryotic and eukaryotic. For example, we classify both animal and plant cells as eukaryotic cells; whereas, we classify bacterial cells as prokaryotic. Before discussing the criteria for determining whether a cell is prokaryotic or eukaryotic, we will first examine how biologists study cells.

MICROSCOPY

Cells vary in size. With few exceptions, we cannot see individual cells with the naked eye, so scientists use microscopes (micro- = “small”; -scope = “to look at”) to study them. A microscope is an instrument that magnifies an object. We photograph most cells with a microscope, so we can call these images micrographs.

The optics of a microscope’s lenses change the image orientation that the user sees. A specimen that is right-side up and facing right on the microscope slide will appear upside-down and facing left when one views through a microscope, and vice versa. Similarly, if one moves the slide left while looking through the microscope, it will appear to move right, and if one moves it down, it will seem to move up. This occurs because microscopes use two sets of lenses to magnify the image. Because of the manner by which light travels through the lenses, this two lens system produces an inverted image (binocular, or dissecting microscopes, work in a similar manner, but include an additional magnification system that makes the final image appear to be upright).

LIGHT MICROSCOPES

To give you a sense of cell size, a typical human red blood cell is about eight millionths of a meter or eight micrometers (abbreviated as eight μm) in diameter. A pin head is about two thousandths of a meter (two mm) in diameter. That means about 250 red blood cells could fit on a pinhead.

Most student microscopes are light microscopes (Figure 4.2a). Visible light passes and bends through the lens system to enable the user to see the specimen. Light microscopes are advantageous for viewing living organisms, but since individual cells are generally transparent, their components are not distinguishable unless they are colored with special stains. Staining, however, usually kills the cells.

Light microscopes that undergraduates commonly use in the laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the process of enlarging an object in appearance. Resolving power is the microscope's ability to distinguish two adjacent structures as separate: the higher the resolution, the better the image's clarity and detail. When one uses oil immersion lenses to study small objects, magnification usually increases to 1,000 times. In order to gain a better understanding of cellular structure and function, scientists typically use electron microscopes.



(a) (b)

Figure 4.2 (a) Most light microscopes in a college biology lab can magnify cells up to approximately 400 times and have a resolution of about 200 nanometers. (b) Electron microscopes provide a much higher magnification, 100,000x, and have a resolution of 50 picometers. (credit a: modification of work by "GcG"/Wikimedia Commons; credit b: modification of work by Evan Bench)

ELECTRON MICROSCOPES

In contrast to light microscopes, electron microscopes (Figure 4.2b) use a beam of electrons instead of a beam of light. Not only does this allow for higher magnification and, thus, more detail (Figure 4.3), it also provides higher resolving power. The method to prepare the specimen for viewing with an electron microscope kills the specimen. Electrons have short wavelengths (shorter than photons) that move best in a vacuum, so we cannot view living cells with an electron microscope.

In a scanning electron microscope, a beam of electrons moves back and forth across a cell's surface, creating details of cell surface characteristics. In a transmission electron microscope, the electron beam penetrates the cell and provides details of a cell's internal structures. As you might imagine, electron microscopes are significantly more bulky and expensive than light microscopes.



Figure 4.3 (a) These *Salmonella* bacteria appear as tiny purple dots when viewed with a light microscope. (b) This scanning electron microscope micrograph shows *Salmonella* bacteria (in red) invading human cells (yellow). Even though subfigure (b) shows a different *Salmonella* specimen than subfigure (a), you can still observe the comparative increase in magnification and detail. (credit a: modification of work by CDC/Armed Forces Institute of Pathology, Charles N. Farmer, Rocky Mountain Laboratories; credit b: modification of work by NIAID, NIH; scale-bar data from Matt Russell)

LINK TO LEARNING

LINK TO LEARNING

For another perspective on cell size, try the HowBig interactive at this site.

CELL THEORY

The microscopes we use today are far more complex than those that Dutch shopkeeper Antony van Leeuwenhoek, used in the 1600s. Skilled in crafting lenses, van Leeuwenhoek observed the movements of single-celled organisms, which he collectively termed "animalcules."

In the 1665 publication *Micrographia*, experimental scientist Robert Hooke coined the term "cell" for the box-like structures he observed when viewing cork tissue through a lens. In the 1670s, van Leeuwenhoek discovered bacteria and protozoa. Later advances in lenses, microscope construction, and staining techniques enabled other scientists to see some components inside cells.

By the late 1830s, botanist Matthias Schleiden and zoologist Theodor Schwann were studying tissues and proposed the unified cell theory, which states that one or more cells comprise all living things, the cell is the basic unit of life, and new cells arise from existing cells. Rudolf Virchow later made important contributions to this theory.

CAREER CONNECTION

CAREER CONNECTION

CytotechnologistHave you ever heard of a medical test called a Pap smear (Figure 4.4)? In this test, a doctor takes a small sample of cells from the patient's uterine cervix and sends it to a medical lab where a cytotechnologist stains the cells and examines them for any changes that could indicate cervical cancer or a microbial infection.

Cytotechnologists (cyto- = "cell") are professionals who study cells via microscopic examinations and other laboratory tests. They are trained to determine which cellular changes are within normal limits and which are abnormal. Their focus is not limited to cervical cells. They study cellular specimens that come from all organs. When they notice abnormalities, they consult a pathologist, a medical doctor who interprets and diagnoses changes that disease in body tissue and fluids cause.

Cytotechnologists play a vital role in saving people's lives. When doctors discover abnormalities early, a patient's treatment can begin sooner, which usually increases the chances of a successful outcome.

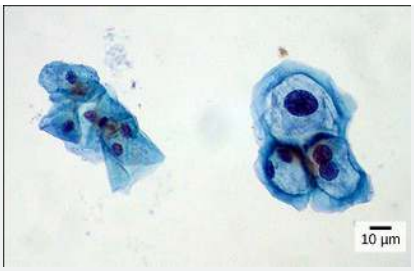


Figure 4.4 These uterine cervix cells, viewed through a light microscope, are from a Pap smear. Normal cells are on the left. The cells on the right are infected with human papillomavirus (HPV). Notice that the infected cells are larger. Also, two of these cells each have two nuclei instead of one, the normal number. (credit: modification of work by Ed Uthman, MD; scale-bar data from Matt Russell)

This page titled 2.2: Studying Cells is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

2.3: DNA is Packaged into Chromatin

DNA can be highly compacted

If stretched to its full length, the DNA molecule of the largest human chromosome would be 85mm. Yet during mitosis and meiosis, this DNA molecule is compacted into a chromosome approximately 5µm long. Although this compaction makes it easier to transport DNA within a dividing cell, it also makes DNA less accessible for other cellular functions such as DNA synthesis and transcription. Thus, chromosomes vary in how tightly DNA is packaged, depending on the stage of the cell cycle and also depending on the level of gene activity required in any particular region of the chromosome.

Levels of compaction

There are several different levels of structural organization in eukaryotic chromosomes, with each successive level contributing to the further compaction of DNA (Figure 2.3.2). For more loosely compacted DNA, only the first few levels of organization may apply. Each level involves a specific set of proteins that associate with the DNA to compact it. First, proteins called the **core histones** act as spool around which DNA is coiled twice to form a structure called the **nucleosome**. Nucleosomes are formed at regular intervals along the DNA strand, giving the molecule the appearance of "beads on a string". At the next level of organization, **histone H1** helps to compact the DNA strand and its nucleosomes into a **30nm fibre**. Subsequent levels of organization involve the addition of **scaffold proteins** that wind the 30nm fibre into coils, which are in turn wound around other scaffold proteins.

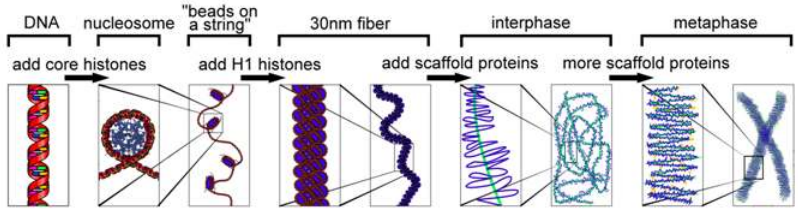


Figure 2.3.2: Successive stages of chromosome compaction depend on the introduction of additional proteins. (Wikipedia-R. Wheeler-GFDL)

Chromatin Packaging Varies inside the Nucleus: Euchromatin & Heterochromatin

Chromosomes stain with some types of dyes, which is how they got their name (chromosome means "colored body"). Certain dyes stain some regions along a chromosome more intensely than others, giving some chromosomes a banded appearance. The material that makes up chromosomes, which we now know to be proteins and DNA, is called **chromatin**. Classically, there are two major types of chromatin, but these are more the ends of a continous and varied spectrum. **Euchromatin** is more loosely packed, and tends to contain genes that are being transcribed, when compared to the more densely compacted **heterochromatin,** which is rich in repetitive sequences and tends not to be transcribed. Heterochromatin sequences include short, highly-repetitive sequences called **satellite DNA**, which acquired their name because their bouyant density is distictly different from the main band of DNA following ultracentrifugation.

Morphological features of Chromosomes

Chromosomes also contain other distinctive features such as centromeres and telomeres. Both of these are usually heterochromatin. In most cases, each chromosome contains one **centromere**. These sequences are bound by centromeric proteins that link the centromere to microtubules that transport chromosomes during cell division. Under the microscope, centromeres of metaphase chromosomes can sometimes appear as constrictions in the body of the chromosome (Figure 2.3.3) and are called primary (1°) constrictions. If a centromere is located near the middle of a chromosome, it is said to be **metacentric**, while an **acrocentric** centromere is closer to one end of a chromosome, and a **telocentric** chromsome is at, or near, the very end. In contrast, some species have a **holocentric** centromere, where no single centromere can be defined and the entire chromsome acts as the centromere. **Telomeres** are repetitive sequences near the ends of linear chromosomes, and are important in maintaining the length of the chromosomes during replication, and protecting the ends of the chromosomes from alterations.



Figure 2.3.3: A pair of metacentric chromosomes. The arrow shows a centromeric region (primary constriction). (Unkonwn-Unknown-pending)

It is essential to describe the similarity between chromosomes using appropriate terminology (Fig 2.4). **Homologous chromosomes** are typically pairs of similar, but non-identical, chromosomes in which one member of the pair comes from the male parent, and the other comes from the female parent. Homologs contain the same gene loci but not necessarily the same alleles. **Non-homologous chromosomes** contain different gene loci, and are usually distinguishable based on cytological features such as length, centromere position, and banding patterns.

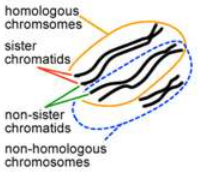


Figure 2.3.4: Relationships between chromosomes and chromatids. (Original-Deyholos-CC:AN)

An **unreplicated chromosomes** can undergo replication, to produce a **replicated chromosome** that has two **sister chromatids,** which are physically connected to each other at the centromere and remain joined until cell division. Because a pair of sister chromatids is produced by the replication of a single DNA molecule, their sequences are essentially identical (same alleles), differing only because of DNA replication errors. On the other hand, **non-sister chromatids** come from two separate, but homologous chromosomes, and therefore usually contain the same gene loci in the same order, but do not necessarily have identical DNA sequences (allelic differences).

The decondensed chromosomes are not randomly arranged in within the interphase nucleus. They often have specific locations within the nucleus and relative to one another (Figure 2.3.5)

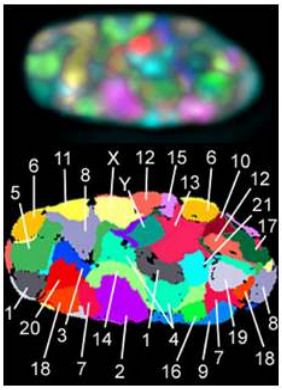


Figure 2.3.5: Top: FISH (Fluorescence in situ hybridization) labeling of all 24 different human chromosomes (1 - 22, X, and Y) in a fibroblast nucleus, each with a different combination of in total seven fluorochromes. **Bottom:** False color representation of all chromosome territories visible in this mid-section after computer classification. (BMC-Unknown-pending)

Chromosome and DNA Replication

When the cell enters *S-phase* in the cell cycle (G1-S-G2-M) all the chromosomal DNA is replicated. This is done by enzymes called **DNA polymerases**. All DNA polymerases synthesize new strands by adding nucleotides to the 3'OH group present on the previous nucleotide. For this reason they are said to work in a 5' to 3' direction. DNA polymerases use a single strand of DNA as a template upon which it will synthesize the complementary sequence. This works fine for the middle of chromosomes - DNA-directed DNA polymerases travel along the original DNA strands making complementary strands (Figure 2.3.6a).

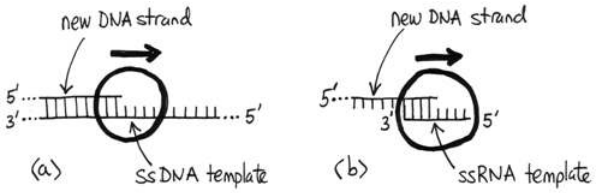


Figure 2.3.6: DNA polymerases make new strands in a 5' to 3' direction. (a) Regular DNA polymerases are proteins or protein complexes that use a single strand of DNA as a template. For example, the main human DNA polymerase, Pol α, is large protein complex made of four polypeptides. (b) Telomerases use their own RNA as a template. The human telomerase is a complex made of one polypeptide and one RNA molecule.(Original-Harrington-CC:AN)

DNA replication in both prokaryotes and eukaryotes begins at an **Origin of Replication** (Ori). Origins are specific sequences on specific positions on the chromosome. In *E. coli*, the *OriC* origin is ~245 bp in size. Chromosome replication begins with the binding of the DnaA initiator protein to an AT-rich 9-mer in *OriC* and melts the two strands. Then DnaC loader protein helps DnaB helicase protein extend the single stranded regions such that the DnaG primase can initiate the synthesis of an RNA primer, from which the DNA polymerases can begin DNA synthesis at the two replication forks. The forks continue in opposite directions until they meet another fork or the end of the chromosome (Figure 2.3.7).

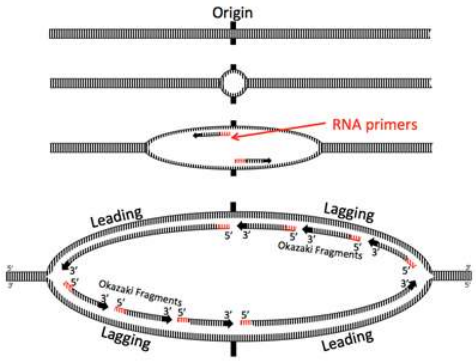


Figure 2.3.7: An origin of replication. The sequence specific DNA duplex is melted then the primase synthesizes RNA primers from which bidirectional DNA replication begins as the two replication forks head off in opposite directions. The leading and lagging strands are shown along with Okazaki fragments. Note the 5' and 3' orientation of all strands. (Original-Locke-CC:AN)

The ends of linear chromosomes present a problem – at each end one strand cannot be completely replicated because there is no primer to extend. Although the loss of such a small sequence would not be a problem, the continued rounds of replication would result in the continued loss of sequence from the chromosome end to a point were it would begin to loose essential gene sequences. Thus, this DNA must be replicated. Most eukaryotes solve the problem of synthesizing this unreplicated DNA with a specialized DNA polymerase called **telomerase,** in combination with a regular polymerase. Telomerases are RNA-directed DNA polymerases. They are a **riboprotein,** as they are composed of both protein and RNA. As Figure 2.3.6b shows, these enzymes contain a small piece of RNA that serves as a portable and reusable template from which the complementary DNA is synthesized. The RNA in human telomerases uses the sequence 3-AAUCCC-5' as the template, and thus our telomere DNA has the complementary sequence 5'-TTAGGG-3' repeated over and over 1000's of times. After the telomerase has made the first strand a primase synthesizes an RNA primer and a regular DNA polymerase can then make a complementary strand so that the telomere DNA will ultimately be double stranded to the original length (Figure 2.3.8). Note: the number of repeats, and thus the size of the telomere, is not set. It fluctuates after each round of the cell cycle. Because there are many repeats at the end, this fluctuation maintains a length buffer – sometimes it's longer, sometimes it's shorter – but the average length will be maintained over the generations of cell replication.

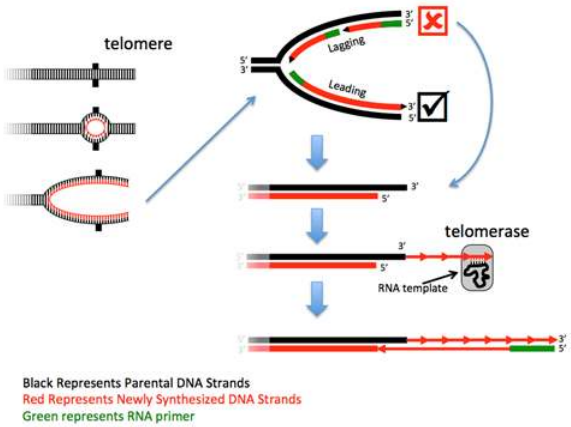


Figure 2.3.8: Telomere replication showing the completion of the leading strand and incomplete replication of the lagging strand. The gap is replicated by the extension of the 3' end by telomerase and then filled in by extension of an RNA primer. (Original-Locke-CC:AN)

In the absence of telomerase, as is the case in human somatic cells, repeated cell division leads to the "**Hayflick limit**", where the telomeres shorten to a critical limit and then the cells enter a senescence phase of non-growth. The activation of telomerase expression permits a cell and its descendants to become immortal and bypass the Hayflick limit. This happens in cancer cells, which can form tumours as well as in cells in culture, such as **HeLa cells**, which can be propagated essentially indefinitely. HeLa cells have been kept in culture since 1951.

Eukaryote Chromosomes have Multiple Origins of Replication

In prokaryotes, with a small, simple, circular chromosome, only one origin of replication is needed to replicate the whole genome. For example, *E. coli* has a ~4.5 Mb genome (chromosome) that can be duplicated in ~40 minutes assuming a single origin, bi-directional replication, and a speed of ~1000 bases/second/fork for the polymerase.

However, in larger, more complicated eukaryotes, with multiple linear chromosomes, more than one origin of replication is required per chromosome to duplicate the whole chromosome set in the 8-hours of S-phase of the cell cycle. For example, the human diploid genome has 46 chromosomes (6 x 10^9 basepairs). The shortest chromosomes are ~50 Mbp long and so could not possibly be replicated from one origin. Additionally, the rate of replication fork movement is slower, only ~100 base/second. Thus, eukaryotes contain multiple origins of replication distributed over the length of each chromosome to enable the duplication of each chromosome within the observed time of S-phase (Fig 2.9).

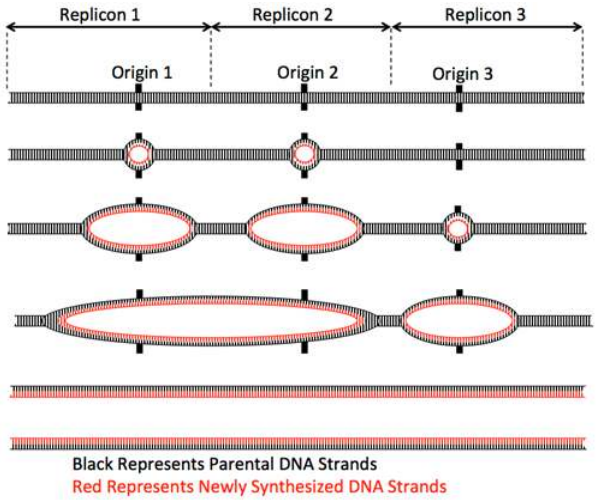


Figure 2.3.9: Part of a eukaryote chromosome showing multiple Origins (1, 2, 3) of Replication, each defining a **replicon** (1, 2, 3). Replication may start at different times in S-phase. Here #1 and #2 begin first then #3. As the replication forks proceed bi-directionally, they create what are referred to as "**replication bubbles**" that meet and form larger bubbles. The end result is two semi-conservatively replicated duplex DNA strands. (Original-Locke-CC:AN)

This page titled 2.3: DNA is Packaged into Chromatin is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **2.1: DNA is Packaged into Chromatin** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

2.4: The Cell Cycle and Changes in DNA Content

Four stages of a typical cell cycle

The life cycle of eukaryotic cells can generally be divided into four stages and a typical cell cycle is shown in Figure 2.4.13. When a cell is produced through fertilization or cell division, there is usually a lag before it undergoes DNA synthesis (replication). This lag period is called Gap 1 (**G_1**), and ends with the onset of the DNA synthesis (**S**) phase, during which each chromosome is replicated. Following replication, there may be another lag, called Gap 2 (**G_2**), before mitosis (**M**). Cells undergoing meiosis do not usually have a G_2 phase. **Interphase** is as term used to include those phases of the cell cycle excluding mitosis and meiosis. Many variants of this generalized cell cycle also exist. Some cells never leave G_1 phase, and are said to enter a permanent, non-dividing stage called **G_0**. On the other hand, some cells undergo many rounds of DNA synthesis (S) without any mitosis or cell division, leading to endoreduplication. Understanding the control of the cell cycle is an active area of research, particularly because of the relationship between cell division and cancer.

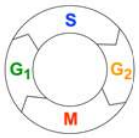


Figure 2.4.13: A typical eukaryotic cell cycle.(Original-Deyholos-CC:AN)

Measures of DNA content and chromosome content

The amount of DNA within a cell changes following each of the following events: fertilization, DNA synthesis, mitosis, and meiosis (Fig 2.14). We use "**c**" to represent the DNA content in a cell, and "**n**" to represent the number of complete sets of chromosomes. In a gamete (i.e. sperm or egg), the amount of DNA is 1c, and the number of chromosomes is 1n. Upon fertilization, both the DNA content and the number of chromosomes doubles to 2c and 2n, respectively. Following DNA replication, the DNA content doubles again to 4c, but each pair of sister chromatids is still counted as a single chromosome (a **replicated chromosome**), so the number of chromosomes remains unchanged at 2n. If the cell undergoes mitosis, each daughter cell will return to 2c and 2n, because it will receive half of the DNA, and one of each pair of sister chromatids. In contrast, the 4 cells that come from meiosis of a 2n, 4c cell are each 1c and 1n, since each pair of sister chromatids, and each pair of homologous chromosomes, divides during meiosis.

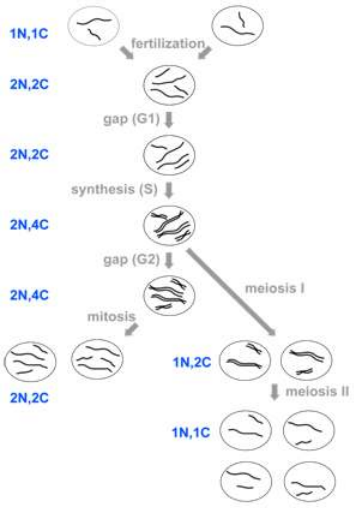


Figure 2.4.14: Changes in DNA and chromosome content during the cell cycle. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation.*(Original-Deyholos-CC:AN)*

This page titled 2.4: The Cell Cycle and Changes in DNA Content is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **2.4: The Cell Cycle and Changes in DNA Content** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

2.5: Karyotypes Describe Chromosome Number and Structure

Karyograms are images of real chromosomes

Each eukaryotic species has its nuclear genome divided among a number of chromosomes that is characteristic of that species. For example, a haploid human nucleus (i.e. sperm or egg) normally has 23 chromosomes (n=23), and a diploid human nucleus has 23 pairs of chromosomes (2n=46). A **karyotype** is the complete set of chromosomes of an individual. The cell was in metaphase so each of the 46 structures is a replicated chromosome even though it is hard to see the two sister chromatids for each chromosome at this resolution. As expected there are 46 chromosomes. Note that the chromosomes have different lengths. In fact, human chromosomes were named based upon this feature. Our largest chromosome is called 1, our next longest is 2, and so on. By convention the chromosomes are arranged into the pattern shown in Figure 2.5.15 and the resulting image is called a **karyogram**. A karyogram allows a geneticist to determine a person's karyotype - a written description of their chromosomes including anything out of the ordinary.

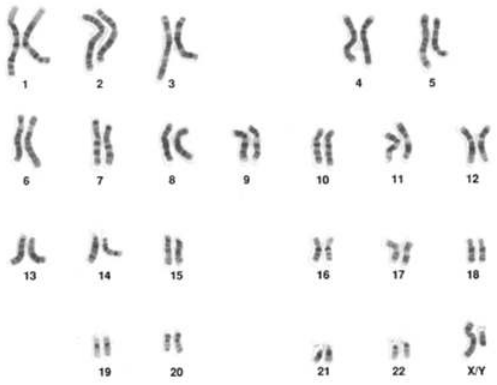


***Figure** 2.5.15**:** Karyogram of a normal human male karytype.(Wikipedia-NHGRI-PD)*

Various stains and fluorescent dyes are used to produce characteristic banding patterns to distinguish all 23 chromosomes. The number of chromosomes varies between species, but there appears to be very little correlation between chromosome number and either the complexity of an organism or its total amount genomic DNA.

Autosomes and Sex Chromosomes

In the figure above note that most of the chromosomes are paired (same length, centromere location, and banding pattern). These chromosomes are called **autosomes**. However note that two of the chromosomes, the X and the Y do not look alike. These are **sex chromosomes**. In humans males have one of each while females have two X chromosomes. Autosomes are those chromosomes present in the same number in males and females while sex chromosomes are those that are not. When sex chromosomes were first discovered their function was unknown and the name X was used to indicate this mystery. The next ones were named Y, then Z, and then W.

The combination of sex chromosomes within a species is associated with either male or female individuals. In mammals, fruit flies, and some flowering plants embryos, those with two X chromosomes develop into females while those with an X and a Y become males. In birds, moths, and butterflies males are ZZ and females are ZW. Because sex chromosomes have arisen multiple times during evolution the molecular mechanism(s) through which they determine sex differs among those organisms. For example, although humans and Drosophila both have X and Y sex chromosomes, they have different mechanisms for determining sex .

In mammals, the sex chromosomes evolved just after the divergence of the monotreme lineage from the lineage that led to placental and marsupial mammals. Thus nearly every mammal species uses the same sex determination system. During embryogenesis the gonads will develop into either ovaries or testes. A gene present only on the Y chromosome called *TDF* encodes a protein that makes the gonads mature into testes. XX embryos do not have this gene and their gonads mature into ovaries instead (default). Once formed the testes produce sex hormones that direct the rest of the developing embryo to become male, while the

ovaries make different sex hormones that promote female development. The testes and ovaries are also the organs where gametes (sperm or eggs) are produced.

How do the sex chromosome behave during meiosis? Well, in those individuals with two of the same chromosome (i.e. **homogametic** sexes: XX females and ZZ males) the chromosomes pair and segregate during meiosis I the same as autosomes do. During meiosis in XY males or ZW females (**heterogametic** sexes) the sex chromosomes pair with each other (Figure 2.5.16). In mammals the consequence of this is that all egg cells will carry an X chromosome while the sperm cells will carry either an X or a Y chromosome. Half of the offspring will receive two X chromosomes and become female while half will receive an X and a Y and become male.

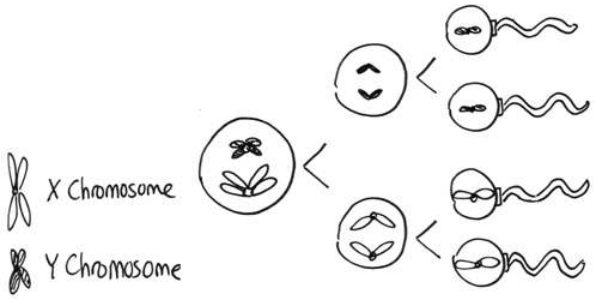


***Figure** 2.5.16**:** Meiosis in an XY mammal. The stages shown are anaphase I, anaphase II, and mature sperm. Note how half of the sperm contain Y chromosomes and half contain X chromosomes. (Original-Harrington-CC:AN)*

Aneuploidy - Changes in Chromosome Number

Analysis of karyotypes can identify chromosomal abnormalities, including **aneuploidy**, which is the addition or subtraction of a chromosome from a pair of homologs. More specifically, the absence of one member of a pair of homologous chromosomes is called **monosomy** (only one remains). On the other hand, in a **trisomy**, there are three, rather than two (disomy), homologs of a particular chromosome. Different types of aneuploidy are sometimes represented symbolically; if **2n** symbolizes the normal number of chromosomes in a cell, then **2n-1** indicates monosomy and **2n+1** represents trisomy. The addition or loss of a whole chromosome is a mutation, a change in the genotype of a cell or organism.

The most familiar human aneuploidy is trisomy-21 (i.e. three copies of chromosome 21), which is one cause of **Down syndrome**. Most (but not all) other human aneuploidies are lethal at an early stage of embryonic development. Note that aneuploidy usually affects only one set of homologs within a karyotype, and is therefore distinct from **polyploidy**, in which the entire chromosome set is duplicated (see below). Aneuploidy is almost always deleterious, whereas polyploidy appears to be beneficial in some organisms, particularly many species of food plants.

Aneuploidy can arise due to a **non-disjunction** event, which is the failure of at least one pair of chromosomes or chromatids to segregate during mitosis or meiosis. Non-disjunction will generate gametes with extra and missing chromosomes.

Chromosomal abnormalities

Structural defects in chromosomes are another type of abnormality that can be detected in karyotypes (Fig 2.17). These defects include **deletions, duplications**, and **inversions**, which all involve changes in a segment of a single chromosome. **Insertions** and **translocations** involve two non-homologous chromosomes. In an insertion, DNA from one chromosome is moved to a non-homologous chromosome in a unidirectional manner. In a translocation, the transfer of chromosomal segments is bidirectional and reciprocal – a reciprocal translocation.

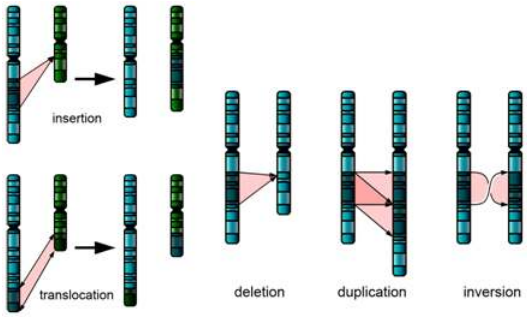


Figure 2.5.17***:*** *Structural abberations in chromosomes.(Wikipedia-Zephyris-GFDL)*

Structural defects affect only part of a chromosome (a subset of genes), and so tend to be less harmful than aneuploidy. In fact, there are many examples of ancient chromosomal rearrangements in the genomes of species including our own. Duplications of some small chromosomal segments, in particular, may have some evolutionary advantage by providing extra copies of some genes, which can then evolve in new, potentially beneficial, ways.

Chromosomal abnormalities arise in many different ways, some of which can be traced to rare errors in natural cellular processes such as DNA replication. **Chromosome breakage** also occurs infrequently as the result of physical damage (such as ionizing radiation), movement of some types of transposons, and other factors. During the repair of a broken chromosome, deletions, insertions, translocations and even inversions can be introduced.

This page titled 2.5: Karyotypes Describe Chromosome Number and Structure is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **2.5: Karyotypes Describe Chromosome Number and Structure** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

2.6: Key Words and Terms

anaphase	G_2 phase	mTOR signaling
apoptosis	*Guardian of the Genome*	necrosis
ATM kinase	immunotherapy	oncogenic viruses
benign tumors	intergrin	p14ARF
cancer cells	interphase	p21
CDKs	invasive tumors	p53
cell cycle	LFS	
cell cycle checkpoints	Li-Fraumeni Syndrome	PD-L1
chemotherapy	M checkpoint	programmed cell death
Chk2	M phase of the cell cycle	prophase
colchicine	malignant tumors	protein phosphorlyation
cyclin	MAPK	proteasome
cyclin level in cell cycle	maturation	radiation therapy
cytokinesis	maturation promoting	S phase
dideoxyNTP	Mdm2	signal transduction
elephant p53 genes	metaphase	SV40
FAK	metastasis	T antigens
G_0 of the cell cycle	mitosis	taxol
G_1 checkpoint	mitosis promoting factor	telophase
G_1 phase	mitotic phases	tumor suppressor protein
G_2 checkpoint	MPF	ubiquitination

This page titled 2.6: Key Words and Terms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **19.6: Key Words and Terms** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

CHAPTER OVERVIEW

3: Meiosis - Sexual Reproduction

Learning Objectives

Course Outcomes for this section:

Apply biological theories and concepts to solve problems related to classical and molecular genetics

1. Describe the molecular basis of inheritance.

The ability to reproduce *in kind* is a basic characteristic of all living things. *In kind* means that the offspring of any organism closely resembles its parent or parents. Hippopotamuses give birth to hippopotamus calves; Monterey pine trees produce seeds from which Monterey pine seedlings emerge; and adult flamingos lay eggs that hatch into flamingo chicks. *In kind* does not generally mean *exactly the same*. While many single-celled organisms and a few multicellular organisms can produce genetically identical clones of themselves through mitotic cell division, many single-celled organisms and most multicellular organisms reproduce regularly using another method.



Figure 3.1**:** Each of us, like these other large multicellular organisms, begins life as a fertilized egg. After trillions of cell divisions, each of us develops into a complex, multicellular organism. (credit a: modification of work by Frank Wouters; credit b: modification of work by Ken Cole, USGS; credit c: modification of work by Martin Pettitt)

Sexual reproduction is the production by parents of sex cells and the fusion of two sex cells to form a single, unique cell. In multicellular organisms, this new cell will then undergo mitotic cell divisions to develop into an adult organism. A type of cell division called meiosis leads to the cells that are part of the sexual reproductive cycle. Sexual reproduction, specifically meiosis and fertilization, introduces variation into offspring that may account for the evolutionary success of sexual reproduction. The vast majority of eukaryotic organisms can or must employ some form of meiosis and fertilization to reproduce.

Sexual reproduction was an early evolutionary innovation after the appearance of eukaryotic cells. The fact that most eukaryotes reproduce sexually is evidence of its evolutionary success. In many animals, it is the only mode of reproduction. And yet, scientists recognize some real disadvantages to sexual reproduction. On the surface, offspring that are genetically identical to the parent may appear to be more advantageous. If the parent organism is successfully occupying a habitat, offspring with the same traits would be similarly successful. There is also the obvious benefit to an organism that can produce offspring by asexual budding, fragmentation, or asexual eggs. These methods of reproduction do not require another organism of the opposite sex. There is no need to expend energy finding or attracting a mate. That energy can be spent on producing more offspring. Indeed, some organisms that lead a solitary lifestyle have retained the ability to reproduce asexually. In addition, asexual populations only have female individuals, so every individual is capable of reproduction. In contrast, the males in sexual populations (half the population) are not producing offspring themselves. Because of this, an asexual population can grow twice as fast as a sexual population in theory. This means that in competition, the asexual population would have the advantage. All of these advantages to asexual reproduction, which are also disadvantages to sexual reproduction, should mean that the number of species with asexual reproduction should be more common.

However, multicellular organisms that exclusively depend on asexual reproduction are exceedingly rare. Why is sexual reproduction so common? This is one of the important questions in biology and has been the focus of much research from the latter

half of the twentieth century until now. A likely explanation is that the variation that sexual reproduction creates among offspring is very important to the survival and reproduction of those offspring. The only source of variation in asexual organisms is mutation. This is the ultimate source of variation in sexual organisms. In addition, those different mutations are continually reshuffled from one generation to the next when different parents combine their unique genomes, and the genes are mixed into different combinations by the process of **meiosis**. Meiosis is the division of the contents of the nucleus that divides the chromosomes among gametes. Variation is introduced during meiosis, as well as when the gametes combine in fertilization.

3.1: Overview of Meiosis
3.2: Meiosis I
3.3: Meiosis II
3.4: Meiosis
3.5: Comparing Meiosis and Mitosis
3.6: Errors in Meiosis

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:qOUtHXNY@3/Sexual-Reproduction

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:1Q8z96mT@4/Meiosis

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:6-3MVU-j@4/Errors-in-Meiosis

Thumbnail: Pollen mother cells dividing during meiosis. (CC BY-SA 3.0; Josef Reischig via Wikimedia Commons).

This page titled 3: Meiosis - Sexual Reproduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

3.1: Overview of Meiosis

Sexual reproduction requires **fertilization**, a union of two cells from two individual organisms. If those two cells each contain one set of chromosomes, then the resulting cell contains two sets of chromosomes. The number of sets of chromosomes in a cell is called its ploidy level (Figure 3.1.1). **Haploid** cells contain one set of chromosomes. Cells containing two sets of chromosomes are called **diploid**. If the reproductive cycle is to continue, the diploid cell must somehow reduce its number of chromosome sets before fertilization can occur again, or there will be a continual doubling in the number of chromosome sets in every generation. So, in addition to fertilization, sexual reproduction includes a nuclear division, known as meiosis, that reduces the number of chromosome sets.

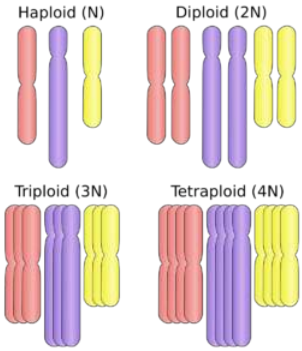


Figure 3.1.1**:** Number of chromosomes in a haploid and diploid cell. Note that triploid and tetraploid are not normal numbers of chromosomes in humans. (CC BY-SA 3.0; Ehamberg via Wikimedia Commons)

Most animals and plants are diploid, containing two sets of chromosomes; in each **somatic cell** (the non-reproductive cells of a multicellular organism), the nucleus contains two copies of each chromosome that are referred to as homologous chromosomes. Somatic cells are sometimes referred to as “body” cells. **Homologous chromosomes** are matched pairs containing genes for the same traits in identical locations along their length (Figure 3.1.2). Diploid organisms inherit one copy of each homologous chromosome from each parent; all together, they are considered a full set of chromosomes. In animals, haploid cells containing a single copy of each homologous chromosome are found only within gametes. Gametes fuse with another haploid gamete to produce a diploid cell.

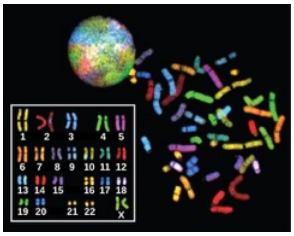


Figure 3.1.2**:** A karyotype displaying all of the chromosomes in the human genome. Note that there are two copies of each chromosome. These are the homologous chromosomes (one from each parent).

Nearly all animals employ a diploid-dominant life-cycle strategy in which the only haploid cells produced by the organism are the gametes. Early in the development of the embryo, specialized diploid cells, called germ cells, are produced within the gonads, such as the testes and ovaries. Germ cells are capable of mitosis to perpetuate the cell line and meiosis to produce gametes. Once the haploid gametes are formed, they lose the ability to divide again. There is no multicellular haploid life stage. Fertilization occurs with the fusion of two gametes, usually from different individuals, restoring the diploid state (Figure 3.1.3).

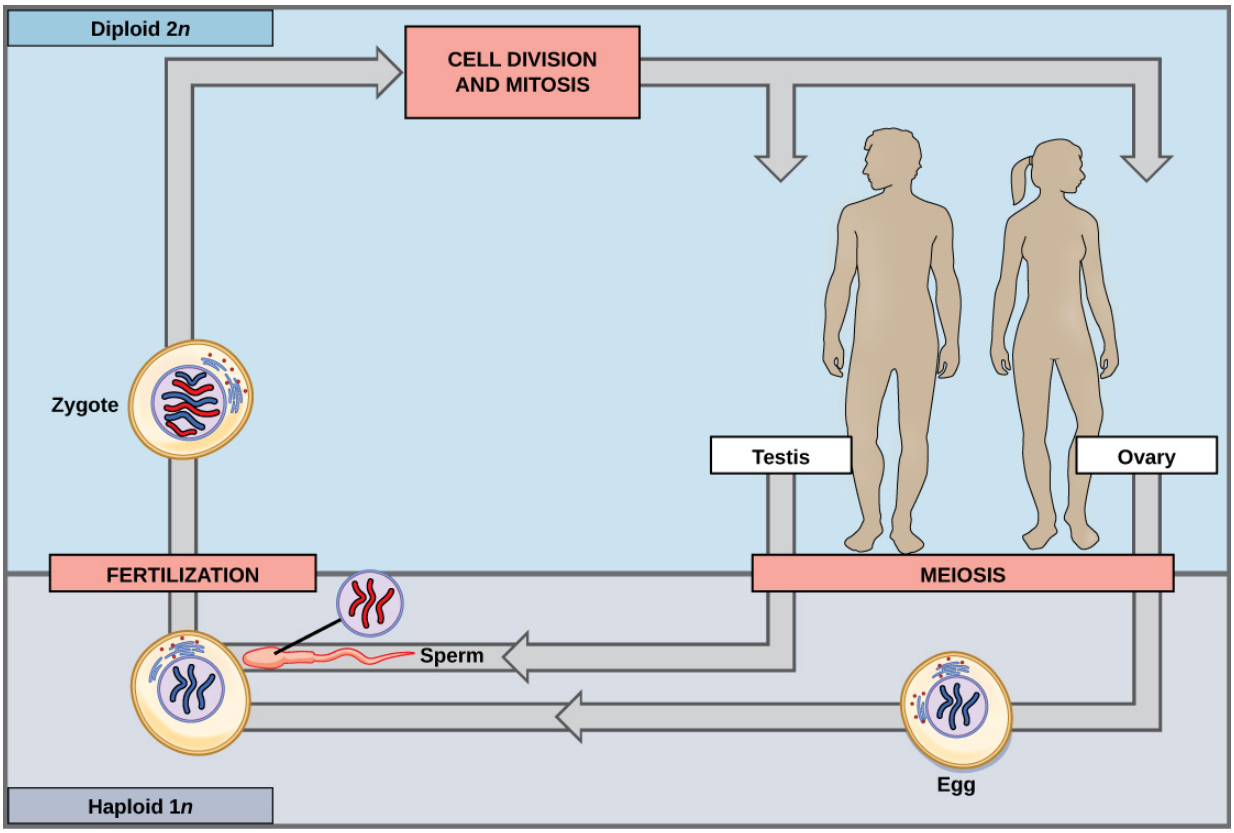


Figure 3.1.3**:** In animals, sexually reproducing adults form haploid gametes from diploid germ cells. Fusion of the gametes gives rise to a fertilized egg cell, or zygote. The zygote will undergo multiple rounds of mitosis to produce a multicellular offspring. The germ cells are generated early in the development of the zygote.

The nuclear division that forms haploid cells, which is called meiosis, is related to mitosis. As you have learned, mitosis is part of a cell reproduction cycle that results in identical daughter nuclei that are also genetically identical to the original parent nucleus. In mitosis, both the parent and the daughter nuclei contain the same number of chromosome sets—diploid for most plants and animals. Meiosis employs many of the same mechanisms as mitosis. However, the starting nucleus is always diploid and the nuclei that result at the end of a meiotic cell division are haploid. To achieve the reduction in chromosome number, meiosis consists of one round of chromosome duplication and two rounds of nuclear division.

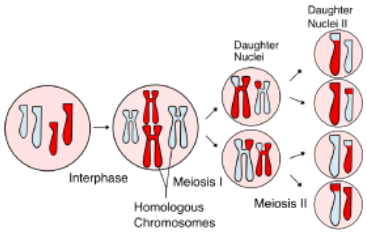


Figure 3.1.4**:** An overview of meiosis. Two sets of homologous chromosomes are shown. One set is comprised of a long red and a long blue chromosome. The second set is the two shorter chromosomes. During interphase, the chromosomes are duplicated so that in the second cell they look like X's. These two connected copies are called sister chromatids. Photo credit Rdbickel; Wikimedia.

Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds of division, the stages are designated with a "I" or "II." Thus, **meiosis I** is the first round of meiotic division and reduces the number of chromosome sets from two to one (Figure 3.1.4). The genetic information is also mixed during this division to create unique recombinant chromosomes. **Meiosis II**, in which the second round of meiotic division takes place in a way that is similar to mitosis, separates the sister chromatids (the identical copies of each chromosome produced during DNA replication that are attached at the centromere).

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:1Q8z96mT@4/Meiosis

This page titled 3.1: Overview of Meiosis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

3.2: Meiosis I

Interphase

Meiosis is preceded by an interphase which is nearly identical to the interphase preceding mitosis. During interphase, the DNA of the chromosomes is replicated (during S phase). After DNA replication, each chromosome becomes composed of two identical copies (called sister chromatids) that are held together at the centromere until they are pulled apart during meiosis II (**Figure** 3.2.1).

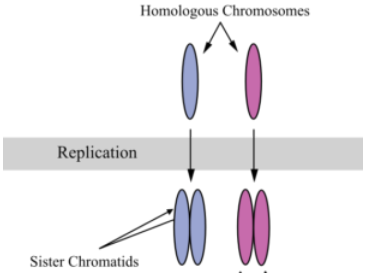


Figure 3.2.1**:** Sister chromatids are identical copies of a chromosome that are held together at the centromere. They are produced during DNA replication. (Credit: User:SyntaxError55, from Wikimedia)

Meiosis I

Meiosis is preceded by an interphase consisting of the G_1, S, and G_2 phases, which are nearly identical to the phases preceding mitosis. The G_1 phase, which is also called the first gap phase, is the first phase of the interphase and is focused on cell growth. The S phase is the second phase of interphase, during which the DNA of the chromosomes is replicated. Finally, the G_2 phase, also called the second gap phase, is the third and final phase of interphase; in this phase, the cell undergoes the final preparations for meiosis.

During DNA duplication in the S phase, each chromosome is replicated to produce two identical copies, called sister chromatids, that are held together at the centromere by cohesin proteins. Cohesin holds the chromatids together until anaphase II. The centrosomes, which are the structures that organize the microtubules of the meiotic spindle, also replicate. This prepares the cell to enter prophase I, the first meiotic phase.

Prophase I

Early in prophase I, the chromosomes can be seen clearly microscopically. As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. The tight pairing of the homologous chromosomes is called **synapsis** (Figure 3.2.2). In synapsis, the genes on the chromatids of the homologous chromosomes are precisely aligned with each other. Recall that synapsis does NOT occur during mitosis.

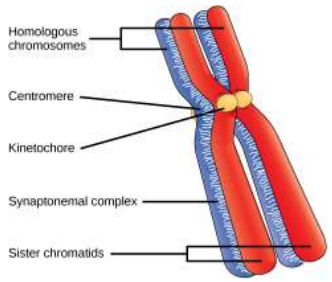


Figure 3.2.2**:** Early in prophase I, homologous chromosomes come together to form a synapse. The chromosomes are bound tightly together and in perfect alignment by a protein lattice called a synaptonemal complex and by cohesin proteins at the centromere.

In synapsis, the genes on the chromatids of the homologous chromosomes are aligned precisely with each other. An exchange of chromosome segments between non-sister homologous chromatids occurs and is called **crossing over (Figure** 3.2.3**)**. The crossover events are the first source of genetic variation produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete, it will carry some DNA from one parent of the individual and some DNA from

the other parent. The **recombinant** sister chromatid has a combination of maternal and paternal genes that did not exist before the crossover.

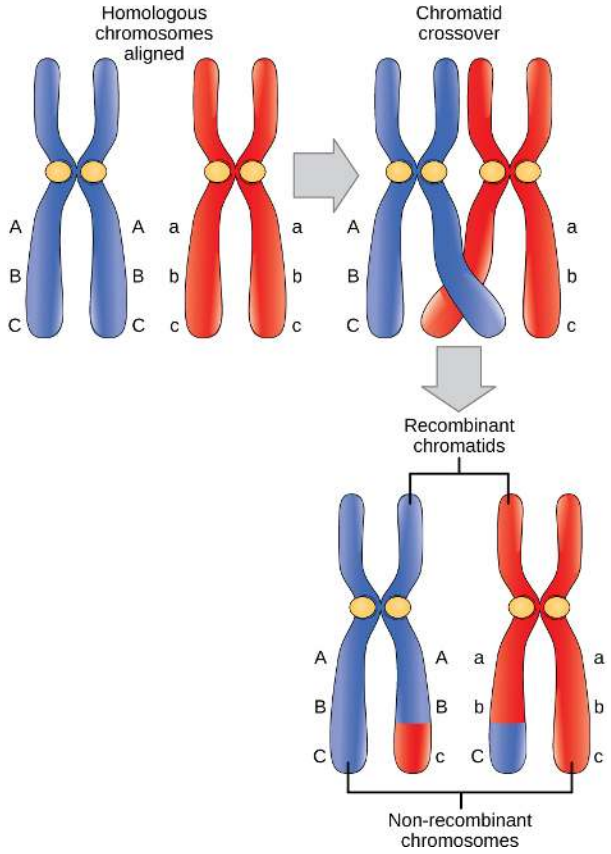


Figure 3.2.3**:** In this illustration of the effects of crossing over, the blue chromosome came from the individual's father and the red chromosome came from the individual's mother. Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes. The chromosomes that have a mixture of maternal and paternal sequence are called recombinant and the chromosomes that are completely paternal or maternal are called non-recombinant.

Prometaphase I

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. Kinetochore proteins are multiprotein complexes that bind the centromeres of a chromosome to the microtubules of the mitotic spindle. Microtubules grow from centrosomes placed at opposite poles of the cell. The microtubules move toward the middle of the cell and attach to one of the two fused homologous chromosomes. The microtubules attach at each chromosomes' kinetochores. With each member of the homologous pair attached to opposite poles of the cell, in the next phase, the microtubules can pull the homologous pair apart. A spindle fiber that has attached to a kinetochore is called a kinetochore microtubule. At the end of prometaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome facing each pole. The homologous chromosomes are still held together at chiasmata. In addition, the nuclear membrane has broken down entirely.

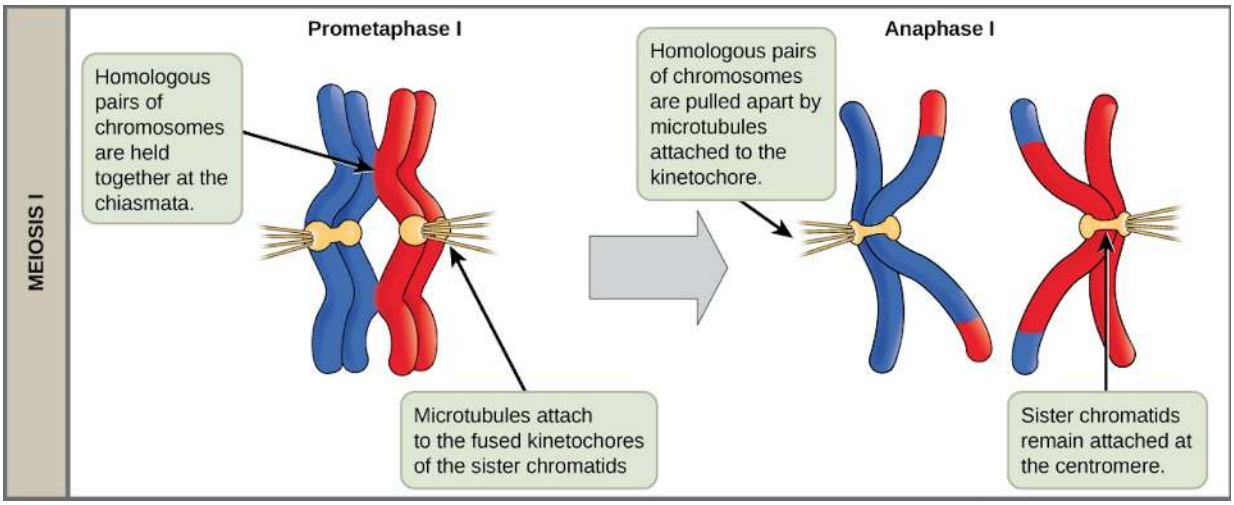


Figure 3.2.4**:** In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes, and the homologous chromosomes are arranged at the midpoint of the cell in metaphase I. In anaphase I, the homologous chromosomes are separated.

Metaphase I

During metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The orientation of each pair of homologous chromosomes at the center of the cell is random. This randomness, called **independent assortment**, is the physical basis for the generation of the second form of genetic variation in offspring (Figure 3.2.5). Consider that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets, one from each parent in the egg and the sperm. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the egg. In metaphase I, these pairs line up at the midway point between the two poles of the cell. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Any maternally inherited chromosome may face either pole. Any paternally inherited chromosome may also face either pole. The orientation of each tetrad is independent of the orientation of the other 22 tetrads.

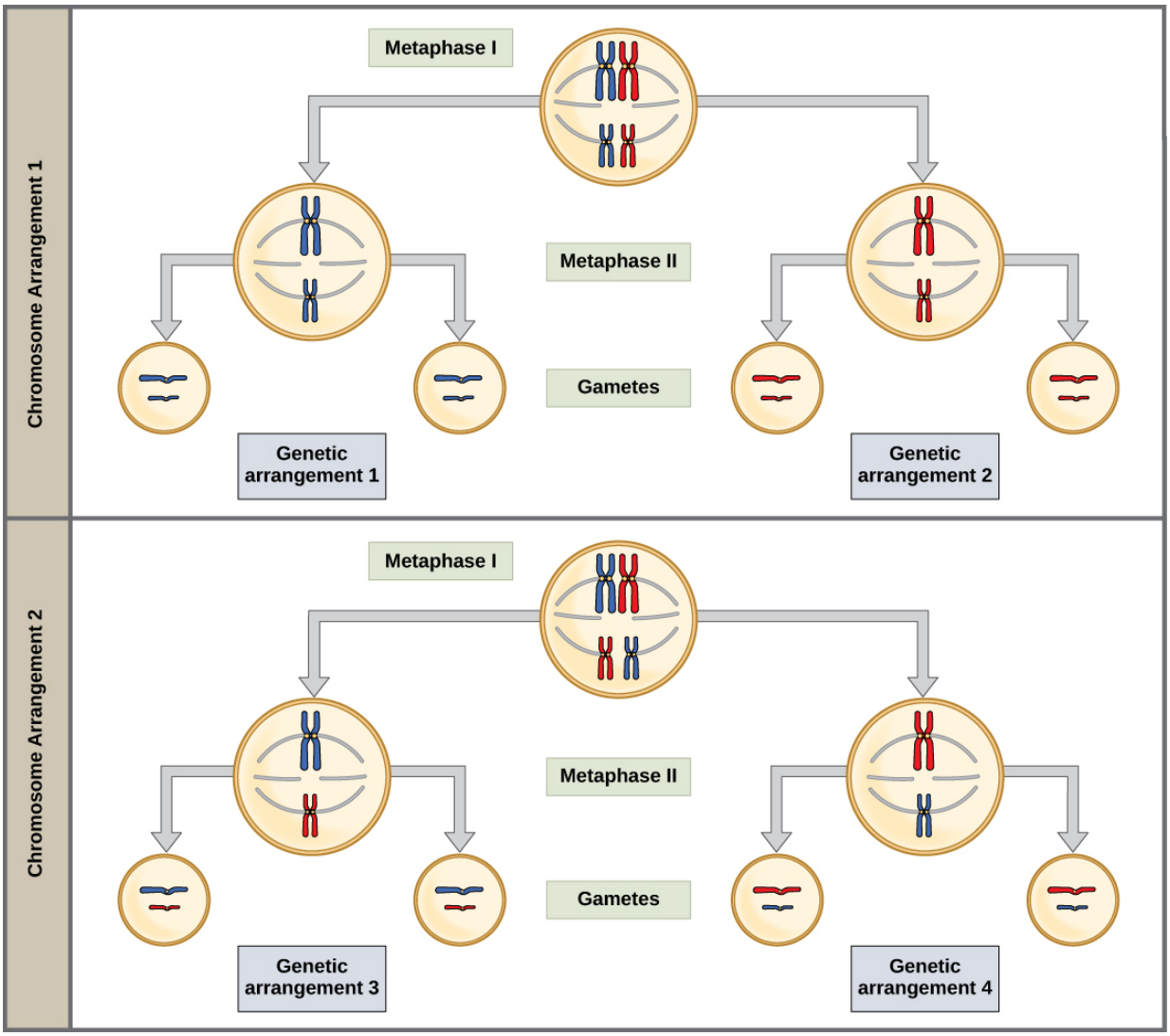


Figure 3.2.5**:** Random, independent assortment during metaphase I can be demonstrated by considering a cell with a set of two chromosomes (n = 2). In this case, there are two possible arrangements at the equatorial plane in metaphase I. The total possible number of different gametes is 2^n, where n equals the number of chromosomes in a set. In this example, there are four possible genetic combinations for the gametes. With n = 23 in human cells, there are over 8 million possible combinations of paternal and maternal chromosomes.

In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations depends on the number of chromosomes making up a set. There are two possibilities for orientation (for each tetrad); thus, the possible number of alignments equals 2^n where n is the number of chromosomes per set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possibilities. This number does not include the variability previously created in the sister chromatids by crossover. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition (**Figure** 3.2.5).

To summarize the genetic consequences of meiosis I: the maternal and paternal genes are recombined by crossover events occurring on each homologous pair during prophase I; in addition, the random assortment of tetrads at metaphase produces a unique combination of maternal and paternal chromosomes that will make their way into the gametes.

Anaphase I

In anaphase I, the microtubules pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. The chiasmata are broken in anaphase I as the microtubules attached to the fused kinetochores pull the homologous chromosomes apart (Figure 3.2.4).

Telophase I and Cytokinesis I

In telophase, the separated chromosomes arrive at opposite poles. The remainder of the typical telophase events may or may not occur, depending on the species. In some organisms, the chromosomes decondense and nuclear envelopes form around the chromatids in telophase I. In other organisms, cytokinesis—the physical separation of the cytoplasmic components into two daughter cells—occurs without reformation of the nuclei. In nearly all species of animals and some fungi, cytokinesis separates the cell contents via a cleavage furrow (constriction of the actin ring that leads to cytoplasmic division). In plants, a cell plate is formed during cell cytokinesis by Golgi vesicles fusing at the metaphase plate. This cell plate will ultimately lead to the formation of cell walls that separate the two daughter cells.

Two haploid cells are the end result of the first meiotic division. The cells are haploid because at each pole, there is just one of each pair of the homologous chromosomes. Therefore, only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, even though each homolog still consists of two sister chromatids. Recall that sister chromatids are merely duplicates of one of the two homologous chromosomes (except for changes that occurred during crossing over). In meiosis II, these two sister chromatids will separate, creating four haploid daughter cells.

Summary of Meiosis I

The chromosomes are copied during interphase (prior to meiosis I). This forms two identical sister chromatids that are attached together at the centromere. During prophase I, crossing over introduces genetic variation by swapping pieces of homologous chromosomes. Additional genetic variation is introduced by independent assortment, which takes into account how the homologous chromosomes line up during metaphase I. At the end of meiosis I, two haploid cells (where each chromosome still consists of two sister chromatids) are produced.

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Biology. OpenStax CNX. January 2, 2017 http://cnx.org/contents/s8Hh0oOc@9.10:1Q8z96mT@4/Meiosis

This page titled 3.2: Meiosis I is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

3.3: Meiosis II

In some species, cells enter a brief interphase, or interkinesis, before entering meiosis II. Interkinesis lacks an S phase, so chromosomes are **not** duplicated. The two cells produced in meiosis I go through the events of meiosis II at the same time. During meiosis II, the sister chromatids within the two daughter cells separate, forming four new haploid gametes. The mechanics of meiosis II is similar to mitosis, except that each dividing cell has only one set of homologous chromosomes. Therefore, each cell has half the number of sister chromatids to separate out as a diploid cell undergoing mitosis.

Prophase II

If the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The centrosomes that were duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed.

Prometaphase II

The nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches to microtubules from opposite poles.

Metaphase II

The sister chromatids are maximally condensed and aligned at the equator of the cell.

Anaphase II

The sister chromatids are pulled apart by the kinetochore microtubules and move toward opposite poles (Figure 3.3.1). Non-kinetochore microtubules elongate the cell.

In meiosis II, the connected sister chromatids remaining in the haploid cells from meiosis I will be split to form four haploid cells. The two cells produced in meiosis I go through the events of meiosis II in synchrony. Overall, meiosis II resembles the mitotic division of a haploid cell. During meiosis II, the sister chromatids are pulled apart by the spindle fibers and move toward opposite poles.

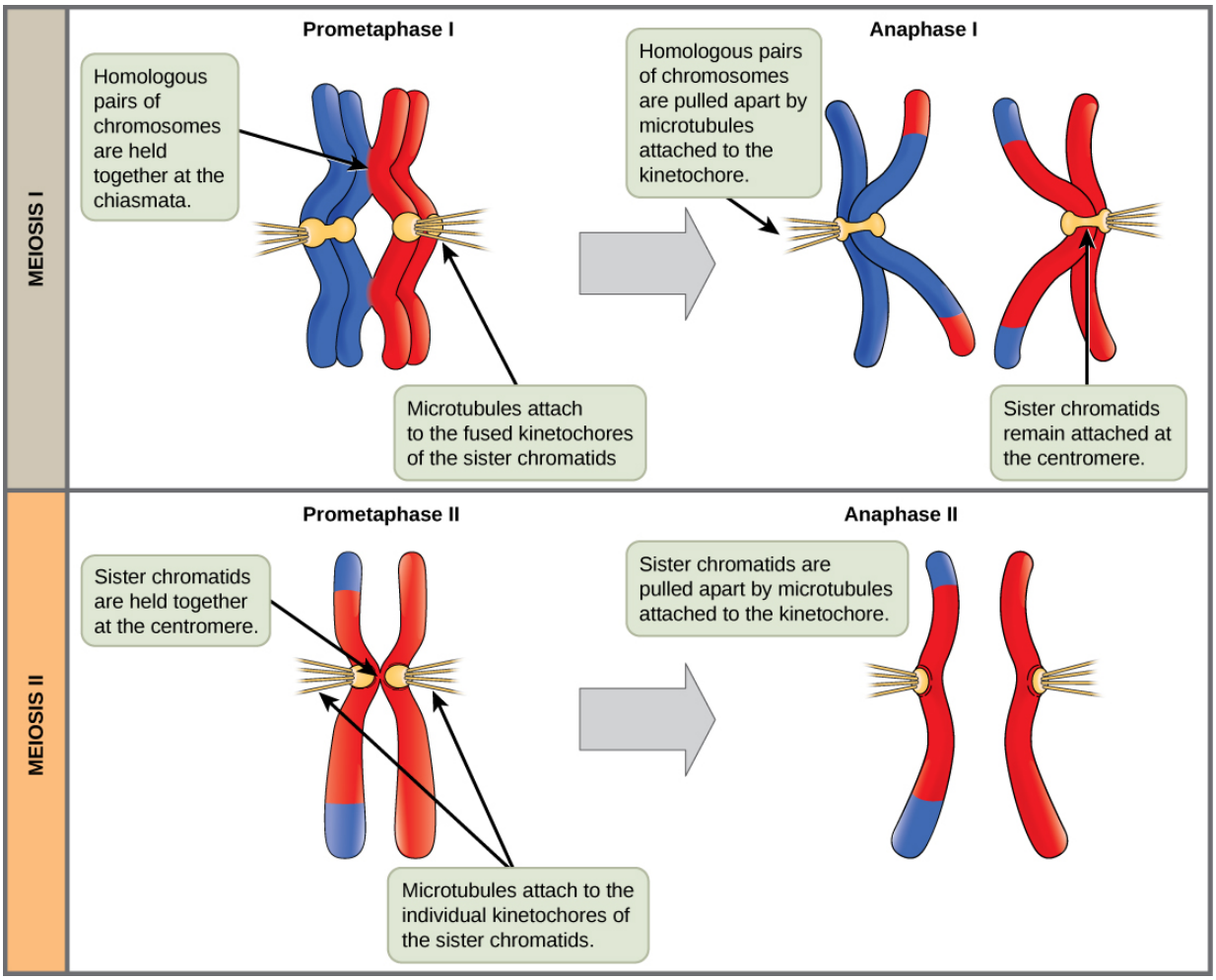
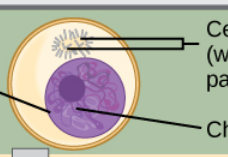
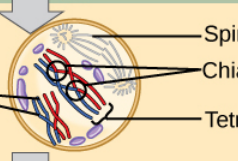
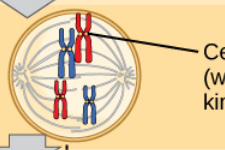
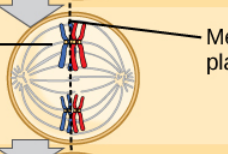
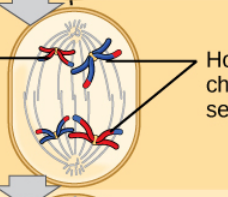
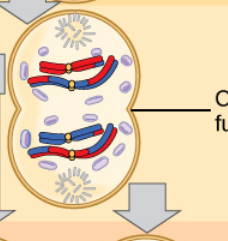
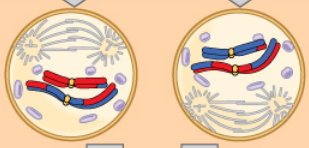
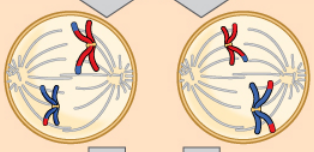
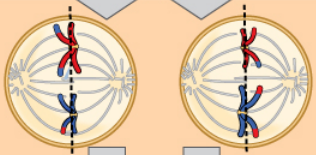
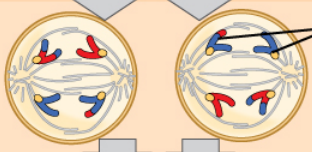


Figure 3.3.1**:** In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to individual kinetochores of sister chromatids. In anaphase II, the sister chromatids are separated.

Telophase II and Cytokinesis

The chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. Cytokinesis separates the two cells into four unique haploid cells. At this point, the newly formed nuclei are both haploid and have only one copy of the single set of chromosomes. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and because of the recombining of maternal and paternal segments of chromosomes (with their sets of genes) that occurs during crossover.

The entire process of meiosis is outlined in Figure 3.3.2.

	Stage	Event	Outcome
INTERPHASE	S phase		Chromosomes are duplicated during interphase. The resulting sister chromatids are held together at the centromere. The centrosomes are also duplicated.
MEIOSIS I	Prophase I		Chromosomes condense, and the nuclear envelope fragments. Homologous chromosomes bind firmly together along their length, forming a tetrad. Chiasmata form between non-sister chromatids. Crossing over occurs at the chiasmata. Spindle fibers emerge from the centrosomes.
	Prometaphase I		Homologous chromosomes are attached to spindle microtubules at the fused kinetochore shared by the sister chromatids. Chromosomes continue to condense, and the nuclear envelope completely disappears.
	Metaphase I		Homologous chromosomes randomly assemble at the metaphase plate, where they have been maneuvered into place by the microtubules.
	Anaphase I		Spindle microtubules pull the homologous chromosomes apart. The sister chromatids are still attached at the centromere.
	Telophase I and Cytokinesis		Sister chromatids arrive at the poles of the cell and begin to decondense. A nuclear envelope forms around each nucleus and the cytoplasm is divided by a cleavage furrow. The result is two haploid cells. Each cell contains one duplicated copy of each homologous chromosome pair.
MEIOSIS II	Prophase II		Sister chromatids condense. A new spindle begins to form. The nuclear envelope starts to fragment.
	Prometaphase II		The nuclear envelope disappears, and the spindle fibers engage the individual kinetochores on the sister chromatids.
	Metaphase II		Sister chromatids line up at the metaphase plate.
	Anaphase II		Sister chromatids are pulled apart by the shortening of the kinetochore microtubules. Non-kinetochore microtubules lengthen the cell.
	Telophase II and		Chromosomes arrive at the poles of the cell and decondense. Nuclear envelopes surround the four

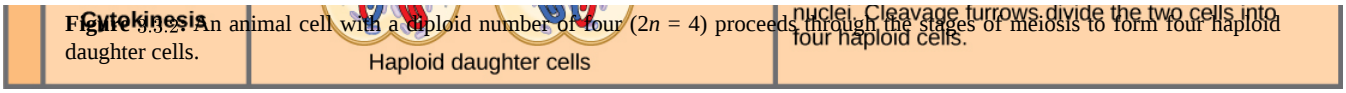


Figure 3.3.2: An animal cell with a diploid number of four ($2n = 4$) proceeds through the stages of meiosis to form four haploid daughter cells.

Summary of Meiosis II

Meiosis II begins with the 2 haploid cells where each chromosome is made up of two connected sister chromatids. DNA replication does NOT occur at the beginning of meiosis II. The sister chromatids are separated, producing 4 genetically different haploid cells.

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:1Q8z96mT@4/Meiosis

This page titled 3.3: Meiosis II is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

3.4: Meiosis

Most eukaryotes replicate sexually - a cell from one individual joins with a cell from another to create the next generation. For this to be successful, the cells that fuse must contain half the number of chromosomes as in the adult organism. Otherwise, the number of chromosomes would double with each generation! The reduction in chromosome number is achieved by the process of **meiosis**. In meiosis, there are usually two steps, Meiosis I and II. In Meiosis I homologous chromosomes segregate, while in Meiosis II sister chromatids segregate. Most multicellular organisms use meiosis to produce **gametes**, the cells that fuse to make offspring. Some single celled eukaryotes such as yeast also use meiosis.

Meiosis I and II

Meiosis begins similarly to mitosis (a cell has replicated its chromosomes and grown large enough to divide), but requires two rounds of division (Figure 3.4.8). In the first, known as meiosis I, the homologous chromosomes separate and segregate. During meiosis II the sister chromatids separate and segregate. Note how meosis I and II are both divided into prophase, metaphase, anaphase, and telophase. After two rounds of cytokinesis, four cells will be produced, each with a single copy of each chromosome.

Cells that will undergo the process of meiosis are called **meiocytes** and are diploid (2N). Meiosis is divided into two stages designated by the roman numerals I and II. Meiosis I is called a **reductional** division, because it reduces the number of chromosomes inherited by each of the daughter cells. Meiosis I is further divided into Prophase I, Metaphase I, Anaphase I, and Telophase I, which are roughly similar to the corresponding stages of mitosis, except that in Prophase I and Metaphase I, homologous chromosomes pair with each other, or **synapse**, and are called **bivalents** (Figs. 2.12). This is an important difference between mitosis and meiosis, because it affects the segregation of alleles, and also allows for recombination to occur through crossing-over, as described later. During Anaphase I, one member of each pair of homologous chromosomes migrates to each daughter cell (1N). Meiosis II resembles mitosis, with one sister chromatid from each chromosome separating to produce two daughter cells. Because Meiosis II, like mitosis, results in the segregation of sister chromatids, Meiosis II is called an **equational** division.

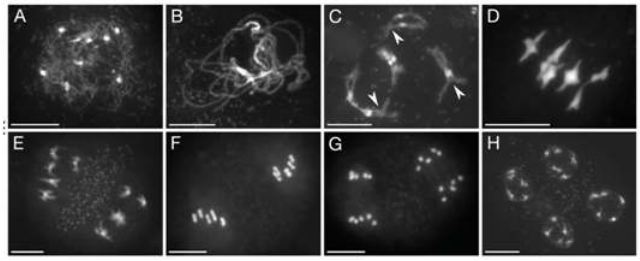


Figure 3.4.12: Meiosis in Arabidopsis (n=5). Panels A-C show different stages of prophase I, each with an increasing degree of chromosome condensation. Subsequent phases are shown: metaphase I (D), telophase I (E), metaphase II (F), anaphase II (G), and telophase II (H). (PLoS Genetics-Chelysheva, L. et al (2008) PLoS Genetics-PD)

In meiosis I replicated, homologous chromosomes **pair up**, or **synapse**, during prophase I, lining up in the middle of the cell during metaphase I, and separating during anaphase I. For this to happen the homologous chromosomes need to be brought together while they condense during prophase I. These attachments are formed in two ways. Proteins bind to both homologous chromosomes along their entire length and form the **synaptonemal complex** (synapse means junction). These proteins hold the chromosomes in a transient structure called a **bivalent**. The proteins are released when the cell enters anaphase I.

Within the synaptonemal complex a second event, **crossingover**, occurs. These are places where DNA repair enzymes break the DNA two non-sister chromatids in similar locations and then covalently reattach non-sister chromatids together to create a crossover between non-sister chromatids. This reorganization of chromatids will persist for the remainder of meiosis and result in recombination of alleles in the gametes.

Crossovers function to hold homologous chromosomes together during meiosis I so they segregate successfully and they also cause the reshuffling of gene/allele combinations to create genetic diversity, which can have an important effect on evolution (see Chapter 7).

Stages of Prophase I

In meiosis, Prophase I is divided up into five visual stages, that are steps along a continuum of events. Leptotene, zygotene, pachytene, diplotene and diakinesis. From interphase, a cell enters **leptotene** as the nuclear material begins to condense into long visible threads (chromosomes). During **Zygotene** homologous chromosomes begin to pair up (synapse) and form an elaborate structure called the synaptonemal complex along their length. At **pachytene** homologous chromosomes are fully synapsed (two chromosomes and four chromatids) to form **bivalents**. **Crossing over** takes place in pachytene. After this, the pairing begins to loosen and individual chromatids become apparent in **diplotene**. This is when the consequences of each crossing over event can be seen as a **chiasma** (plural: **chiasmata**). **Diakinesis** follows as the chromosomes continue to condense and individualize. This is followed by metaphase I were the paired chromosomes orient on the metaphase plate in preparation for segregation (reductional).

Meiosis II and Gamete Maturation

At the completion of meiosis I there are two cells, each with one, replicated copy of each chromosome (1N). Because the number of chromosomes per cell has decreased (2->1), meiosis I is called a **reductional cell division**. In the second part of meiosis the chromosomes will once again be brought to the middle of the cell, but this time it is the sister chromatids that will segregate during anaphase.

After cytokinesis there will be four cells, each containing only one unreplicated chromosome of each type. Meiosis II resembles mitosis in that the number of chromosomes per cell is unchanged - both are **equational cell divisions –** but in meiosis II all cells won't have the same genetic composition. There will be allelic differences among the gametes.

In animals and plants the cells produced by meiosis need to mature before they become functional gametes. In male animals the four products of meiosis are called spermatids. They grow tails and become functional sperm cells. In female animals the gametes are eggs. In order that each egg contains the maximum amount of nutrients only one of the four products of meiosis becomes an egg. The other three cells end up as tiny disposable cells called **polar bodies**. In plants the products of meiosis reproduce a few times using mitosis as they develop into functional male or female gametes.

This page titled 3.4: Meiosis is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **2.3: Meiosis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.5: Comparing Meiosis and Mitosis

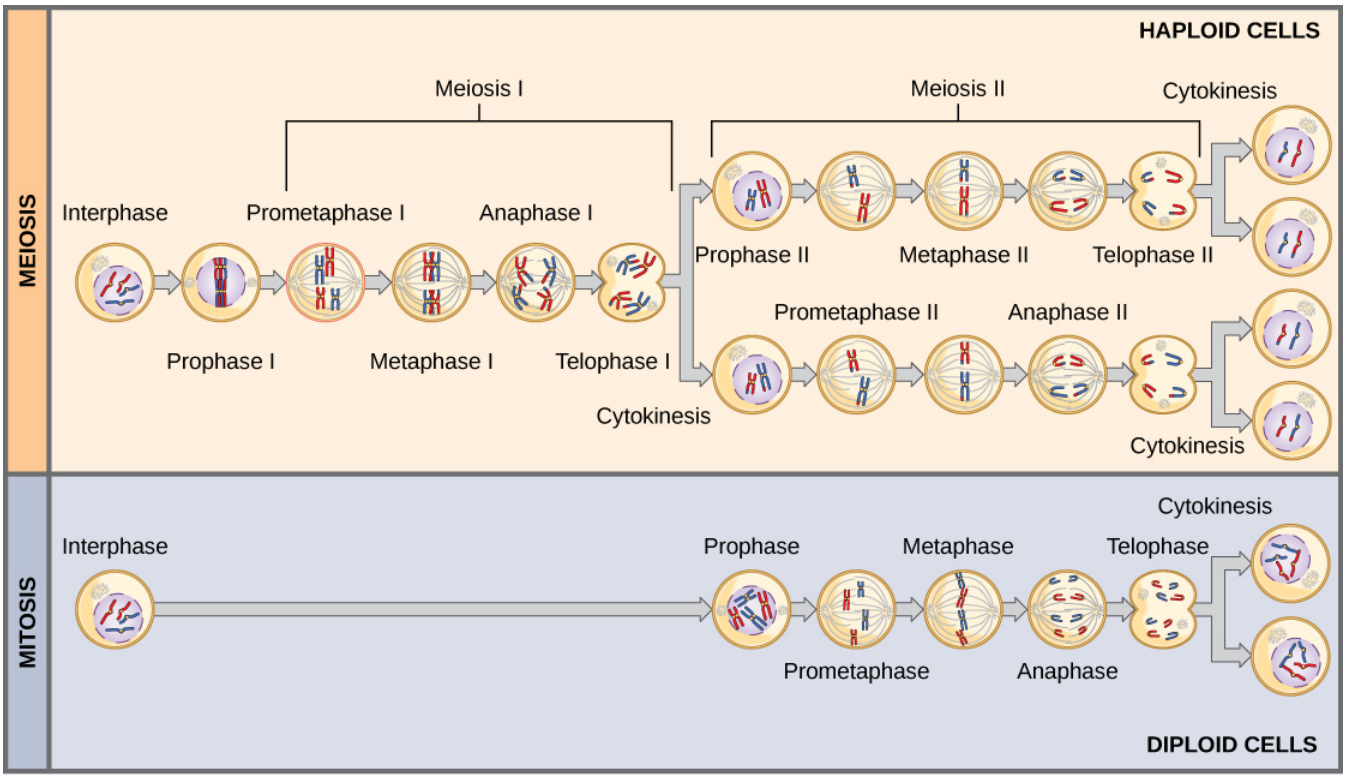
Mitosis and meiosis, which are both forms of division of the nucleus in eukaryotic cells, share some similarities, but also exhibit distinct differences that lead to their very different outcomes. Mitosis is a single nuclear division that results in two nuclei, usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original. They have the same number of sets of chromosomes: one in the case of haploid cells, and two in the case of diploid cells. On the other hand, meiosis is two nuclear divisions that result in four nuclei, usually partitioned into four new cells. The nuclei resulting from meiosis are never genetically identical, and they contain one chromosome set only—this is half the number of the original cell, which was diploid.

The differences in the outcomes of meiosis and mitosis occur because of differences in the behavior of the chromosomes during each process. Most of these differences in the processes occur in meiosis I, which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs become associated with each other, are bound together, experience chiasmata and crossover between sister chromatids, and line up along the metaphase plate in tetrads with spindle fibers from opposite spindle poles attached to each kinetochore of a homolog in a tetrad. All of these events occur only in meiosis I, never in mitosis.

Homologous chromosomes move to opposite poles during meiosis I so the number of sets of chromosomes in each nucleus-to-be is reduced from two to one. For this reason, meiosis I is referred to as a **reduction division**. There is no such reduction in ploidy level in mitosis.

Meiosis II is much more analogous to a mitotic division. In this case, duplicated chromosomes (only one set of them) line up at the center of the cell with divided kinetochores attached to spindle fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid is pulled to one pole and the other sister chromatid is pulled to the other pole. If it were not for the fact that there had been crossovers, the two products of each meiosis II division would be identical as in mitosis; instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because, although there are fewer copies of the genome in the resulting cells, there is still one set of chromosomes, as there was at the end of meiosis I.

Cells produced by mitosis will function in different parts of the body as a part of growth or replacing dead or damaged cells. They may even be involved in asexual reproduction in some organisms. Cells produced by meiosis in a diploid-dominant organism such as an animal will only participate in sexual reproduction.



						OUTCOME
PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Figure 3.5.1**:** Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:1Q8z96mT@4/Meiosis

This page titled 3.5: Comparing Meiosis and Mitosis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

3.6: Errors in Meiosis

Inherited disorders can arise when chromosomes behave abnormally during meiosis. Chromosome disorders can be divided into two categories: abnormalities in chromosome number and chromosome structural rearrangements. Because even small segments of chromosomes can span many genes, chromosomal disorders are characteristically dramatic and often fatal.

Disorders in Chromosome Number

The isolation and microscopic observation of chromosomes forms the basis of cytogenetics and is the primary method by which clinicians detect chromosomal abnormalities in humans. A **karyotype** is the number and appearance of chromosomes, including their length, banding pattern, and centromere position. To obtain a view of an individual's karyotype, cytologists photograph the chromosomes and then cut and paste each chromosome into a chart, or **karyogram** (**Figure** 3.6.1).

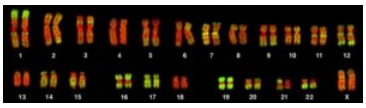


Figure 3.6.1**:** This karyogram shows the chromosomes of a normal female human immune cell during mitosis. (Credit: Andreas Bolzer, et al)

By observing a karyogram, geneticists can actually visualize the chromosomal composition of an individual to confirm or predict genetic abnormalities in offspring even before birth.

Geneticists Use Karyograms to Identify Chromosomal Aberrations

Although Mendel is referred to as the "father of modern genetics," he performed his experiments with none of the tools that the geneticists of today routinely employ. One such powerful cytological technique is karyotyping, a method in which traits characterized by chromosomal abnormalities can be identified from a single cell. To observe an individual's karyotype, a person's cells (like white blood cells) are first collected from a blood sample or other tissue. In the laboratory, the isolated cells are stimulated to begin actively dividing. A chemical called colchicine is then applied to cells to arrest condensed chromosomes in metaphase. Cells are then made to swell using a hypotonic solution so the chromosomes spread apart. Finally, the sample is preserved in a fixative and applied to a slide.

The geneticist then stains chromosomes with one of several dyes to better visualize the distinct and reproducible banding patterns of each chromosome pair. Following staining, the chromosomes are viewed using bright-field microscopy. A common stain choice is the Giemsa stain. Giemsa staining results in approximately 400–800 bands (of tightly coiled DNA and condensed proteins) arranged along all of the 23 chromosome pairs; an experienced geneticist can identify each band. In addition to the banding patterns, chromosomes are further identified on the basis of size and centromere location. To obtain the classic depiction of the karyotype in which homologous pairs of chromosomes are aligned in numerical order from longest to shortest, the geneticist obtains a digital image, identifies each chromosome, and manually arranges the chromosomes into this pattern (Figure 3.6.1).

At its most basic, the karyogram may reveal genetic abnormalities in which an individual has too many or too few chromosomes per cell. Examples of this are Down Syndrome, which is identified by a third copy of chromosome 21, and Turner Syndrome, which is characterized by the presence of only one X chromosome in women instead of the normal two. Geneticists can also identify large deletions or insertions of DNA. For instance, Jacobsen Syndrome—which involves distinctive facial features as well as heart and bleeding defects—is identified by a deletion on chromosome 11. Finally, the karyotype can pinpoint translocations, which occur when a segment of genetic material breaks from one chromosome and reattaches to another chromosome or to a different part of the same chromosome. Translocations are implicated in certain cancers, including chronic myelogenous leukemia.

During Mendel's lifetime, inheritance was an abstract concept that could only be inferred by performing crosses and observing the traits expressed by offspring. By observing a karyogram, today's geneticists can actually visualize the chromosomal composition of an individual to confirm or predict genetic abnormalities in offspring, even before birth.

Of all the chromosomal disorders, abnormalities in chromosome number are the most easily identifiable from a karyogram. Disorders of chromosome number include the duplication or loss of entire chromosomes, as well as changes in the number of complete sets of chromosomes. They are caused by **nondisjunction**, which occurs when pairs of homologous chromosomes or sister chromatids fail to separate during meiosis. The risk of nondisjunction increases with the age of the parents.

Nondisjunction can occur during either meiosis I or II, with different results (**Figure** 3.6.2). If homologous chromosomes fail to separate during meiosis I, the result is two gametes that lack that chromosome and two gametes with two copies of the chromosome. If sister chromatids fail to separate during meiosis II, the result is one gamete that lacks that chromosome, two normal gametes with one copy of the chromosome, and one gamete with two copies of the chromosome.

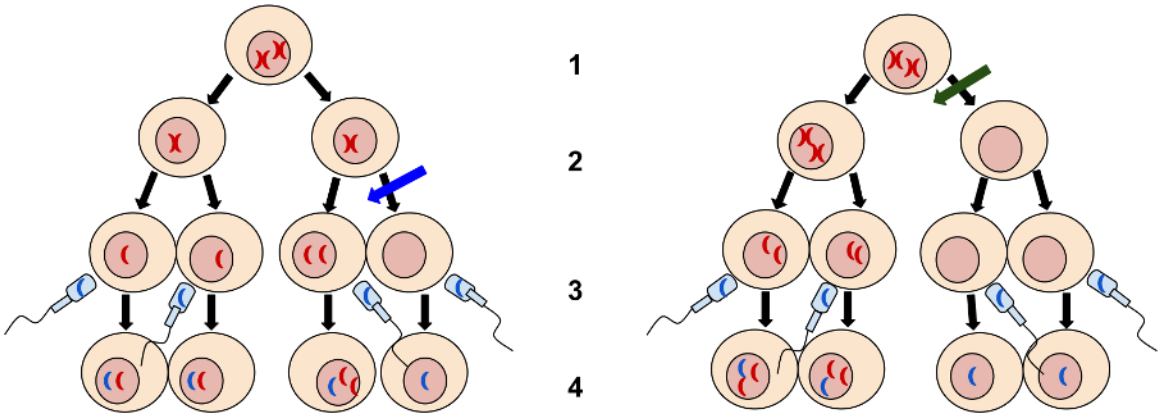


Figure 3.6.2**:** Nondisjunction occurs when homologous chromosomes or sister chromatids fail to separate during meiosis, resulting in an abnormal chromosome number. Nondisjunction may occur during meiosis I or meiosis II. Photo credit Tweety207; Wikimedia.

An individual with the appropriate number of chromosomes for their species is called **euploid**; in humans, euploidy corresponds to 22 pairs of **autosomes** and one pair of sex chromosomes (such as is seen in the karyotype in Figure 3.6.1). An individual with an error in chromosome number is described as **aneuploid**, a term that includes **monosomy** (loss of one chromosome) or **trisomy** (gain of an extraneous chromosome). Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they have only one copy of essential genes. Most autosomal trisomies also fail to develop to birth; however, duplications of some of the smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years. Trisomic individuals suffer from a different type of genetic imbalance: an excess in gene dose. Cell functions are calibrated to the amount of gene product produced by two copies (doses) of each gene; adding a third copy (dose) disrupts this balance. The most common trisomy is that of chromosome 21, which leads to Down syndrome. Individuals with this inherited disorder have characteristic physical features and developmental delays in growth and cognition.

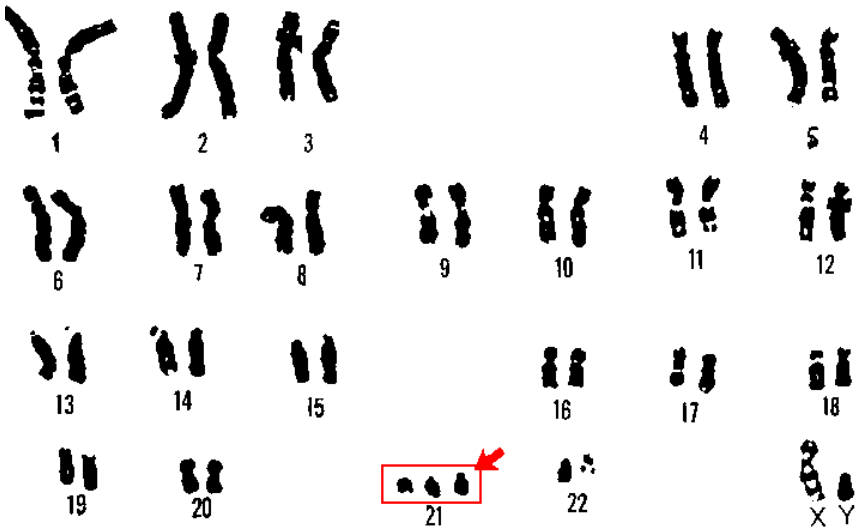


Figure 3.6.3**:** Karyotype of an individual with Down Syndrome. Photo credit U.S. Department of Energy Human Genome Program. Wikimedia.

The incidence of Down syndrome is correlated with maternal age, such that older women are more likely to give birth to children with Down syndrome (Figure 3.6.4).

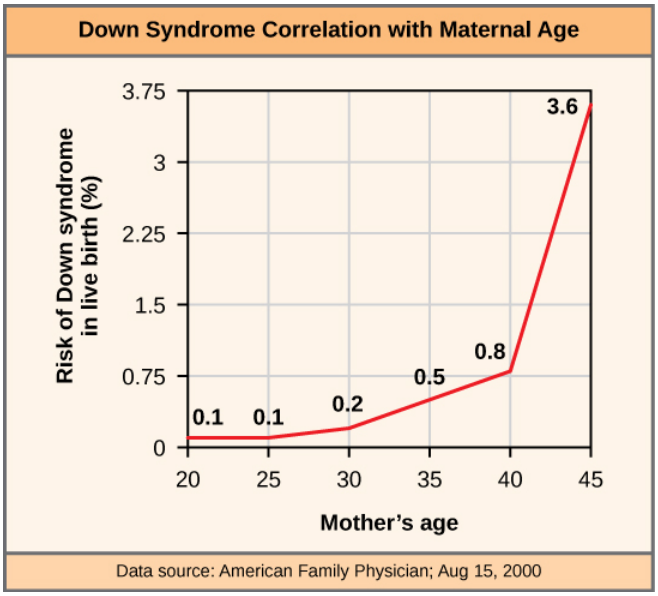


Figure 3.6.4**:** The incidence of having a fetus with trisomy 21 increases dramatically with maternal age.

An individual with more than the correct number of chromosome sets (two for diploid species) is called **polyploid**. For instance, fertilization of an abnormal diploid egg with a normal haploid sperm would yield a triploid zygote. Polyploid animals are extremely rare, with only a few examples among the flatworms, crustaceans, amphibians, fish, and lizards. Triploid animals are sterile because meiosis cannot proceed normally with an odd number of chromosome sets. In contrast, polyploidy is very common in the plant kingdom, and polyploid plants tend to be larger and more robust than euploids of their species (Figure 3.6.5).



Figure 3.6.5**:** As with many polyploid plants, this triploid orange daylily (*Hemerocallis fulva*) is particularly large and robust, and grows flowers with triple the number of petals of its diploid counterparts. (credit: Steve Karg)

Sex Chromosome Nondisjunction

Humans display dramatic deleterious effects with autosomal trisomies and monosomies. Therefore, it may seem counterintuitive that human females and males can function normally, despite carrying different numbers of the X chromosome. In part, this occurs because of a process called **X inactivation**. Early in development, when female mammalian embryos consist of just a few thousand cells, one X chromosome in each cell inactivates by condensing into a structure called a Barr body. The genes on the inactive X chromosome are not expressed. The particular X chromosome (maternally or paternally derived) that is inactivated in each cell is random, but once the inactivation occurs, all cells descended from that cell will have the same inactive X chromosome. By this process, females compensate for their double genetic dose of X chromosome.

In so-called "tortoiseshell" cats, X inactivation is observed as coat-color variegation (**Figure** 3.6.6). Females heterozygous for an X-linked coat color gene will express one of two different coat colors over different regions of their body, corresponding to whichever X chromosome is inactivated in the embryonic cell progenitor of that region. When you see a tortoiseshell cat, you will know that it has to genetically be a female.



Figure 3.6.6**:** Embryonic inactivation of one of two different X chromosomes encoding different coat colors gives rise to the tortoiseshell phenotype in cats. (credit: Michael Bodega)

In an individual carrying an abnormal number of X chromosomes, cellular mechanisms will inactivate all but one X in each of her cells. As a result, X-chromosomal abnormalities are typically associated with mild mental and physical defects, as well as sterility. If the X chromosome is absent altogether, the individual will not develop.

Several errors in sex chromosome number have been characterized. Individuals with three X chromosomes, called triplo-X, appear female but express developmental delays and reduced fertility. The XXY chromosome complement, corresponding to one type of Klinefelter syndrome, corresponds to male individuals with small testes, enlarged breasts, and reduced body hair. The extra X chromosome undergoes inactivation to compensate for the excess genetic dosage. Turner syndrome, characterized as an X0 chromosome complement (i.e., only a single sex chromosome), corresponds to a female individual with short stature, webbed skin in the neck region, hearing and cardiac impairments, and sterility.

Chromosome Structural Rearrangements

Cytologists have characterized numerous structural rearrangements in chromosomes, including partial duplications, deletions, inversions, and translocations. Duplications and deletions often produce offspring that survive but exhibit physical and mental abnormalities. Cri-du-chat (from the French for "cry of the cat") is a syndrome associated with nervous system abnormalities and identifiable physical features that results from a deletion of most of the small arm of chromosome 5 (**Figure** 3.6.7). Infants with this genotype emit a characteristic high-pitched cry upon which the disorder's name is based.



Figure 3.6.7**:** This individual with cri-du-chat syndrome is shown at various ages: (A) age two, (B) age four, (C) age nine, and (D) age 12. (credit: Paola Cerruti Mainardi)

Chromosome inversions and translocations can be identified by observing cells during meiosis because homologous chromosomes with a rearrangement in one of the pair must contort to maintain appropriate gene alignment and pair effectively during prophase I.

A **chromosome inversion** is the detachment, 180° rotation, and reinsertion of part of a chromosome (**Figure** 3.6.8). Unless they disrupt a gene sequence, inversions only change the orientation of genes and are likely to have more mild effects than aneuploid errors.

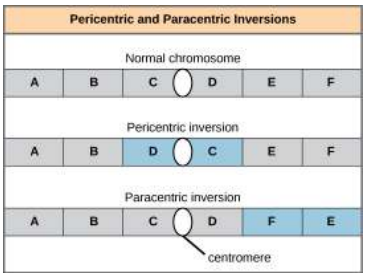


Figure 3.6.8**:** An inversion occurs when a chromosome segment breaks from the chromosome, reverses its orientation, and then reattaches in the original position.

A translocation occurs when a segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome. Translocations can be benign or have devastating effects, depending on how the positions of genes are altered with respect to regulatory sequences. Notably, specific translocations have been associated with several cancers and with schizophrenia. Reciprocal translocations result from the exchange of chromosome segments between two nonhomologous chromosomes such that there is no gain or loss of genetic information (**Figure** 3.6.9).

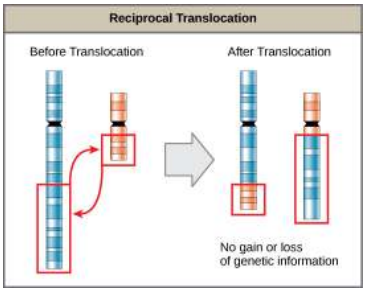


Figure 3.6.9**:** A reciprocal translocation occurs when a segment of DNA is transferred from one chromosome to another, nonhomologous chromosome. (credit: modification of work by National Human Genome Research/USA)

Philadelphia Chromosome

One specific example of a chromosomal translocation – the “Philadelphia chromosome” – is found in people who suffer from chronic myeloid leukemia (CML). In this translocation, a piece of chromosome 9 is swapped with a section of chromosome 22. This connects two genes on chromosome 22; one that was originally from chromosome 9 and one that was from chromosome 22. This translocation produces the BCR-ABL fusion protein, which causes white blood cells to divide out of control. BCR-ABL positive cancers can be treated with the drug Gleevac.

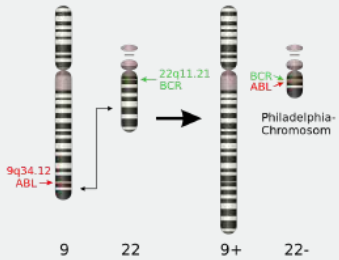


Figure 3.6.10**:** "Philadelphia chromosome" showing the location of the BCR-ABL fusion protein. Photo credit A Obeidat; Wikimedia.

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:6-3MVU-j@4/Errors-in-Meiosis

This page titled 3.6: Errors in Meiosis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

CHAPTER OVERVIEW

4: Mendelian Genetics

Before Mendel, the basic rules of heredity were not understood. For example, it was known that green-seeded pea plants occasionally produced offspring that had yellow seeds; but were the hereditary factors that controlled seed color somehow changing from one generation to the next, or were certain factors disappearing and reappearing? And did the same factors that controlled seed color also control things like plant height?

4.1: Introduction
4.2: Introduction
4.3: Mendel's Experiments
4.4: Punnett Squares and Test Crosses
4.5: Laws of Inheritance
4.6: Mendel's First Law
4.7: Relationships Between Genes, Genotypes and Phenotypes
4.8: Pedigree Analysis
- 4.8.1: Symbols used in Pedigree Charts
- 4.8.2: Modes of Inheritance
- 4.8.3: Sporadic and Non-Heritable Diseases
- 4.8.4: Calculating Probabilities Using Pedigree Charts
- 4.8.5: Chapter 4 Summary
- 4.8.6: Key Terms in Chapter 4
- 4.8.7: Chapter 4 Study Questions
- 4.8.8: Chapter 4 Answers

4.9: Pedigrees and Punnett Squares
4.10: Biochemical Basis of Dominance
4.11: Black fur color - a dominant trait
4.12: Yellow fur color - a recessive trait
4.13: Sex-Linkage- An Exception to Mendel's First Law
4.14: Phenotypes May Not Be As Expected from the Genotype
4.15: Phenotypic Ratios May Not Be As Expected
4.16: Extensions of the Laws of Inheritance
4.E: Genetic Analysis of Single Genes (Exercises)
4.S: Genetic Analysis of Single Genes (Summary)

Thumbnail: Pea plants were used by Gregor Mendel to discover some fundamental laws of genetics. (Flicker-Christian Guthier-CC:A)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 4: Mendelian Genetics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Introduction

Learning Objectives

- Describe the different laws of genetics developed by Mendel.
- Utilize Punnett square to perform and follow a monohybrid cross.
- Explain how pedigree trees are constructed and how to use them to study genetic diseases.

We all have traits we inherit from our parents, things such as your hair color or your eye color. But what determines which traits we take after our mother versus our father? Some genetic diseases such as cystic fibrosis run in families. How do we know if a newborn baby is at risk of having the disease? And if he/she is at risk, can we calculate the chance of the risk?

About 150 years ago, an Austrian monk named **Johann Gregor Mendel** (1822-1884) (Figure 4.1.1) bred peas in his garden. He was interested in studying several characteristics of pea plants: plant height, flower color, seed color, seed shape, etc. He studied them carefully, and bred them for generations, and worked out several rules of inheritance, now referred to as the **Laws of Inheritance**. Mendel's work was not highly regarded when he was alive, but was rediscovered in the early 20th century when several scientists independently verified Mendel's experimental findings. Today, Mendel is considered as the **father of modern genetics**.



Figure 4.1.1**:** Johann Gregor Mendel set the framework for the study of genetics.

In this chapter we will explore Mendel's experiments with pea plants, which laid the groundwork for modern genetics. It covers key concepts like the Law of Segregation and the Law of Independent Assortment, explaining how traits are inherited through dominant and recessive alleles. Monohybrid and dihybrid crosses are explored using Punnett squares, with examples of genotype and phenotype ratios.

Definition: Term

Definition text

Image Long Description

- **Figure 4.1.1** Johann Gregor Mendel (1822–1884) was a lifelong learner, teacher, scientist, and man of faith. As a young adult, he joined the Augustinian Abbey of St. Thomas in Brno in what is now the Czech Republic. Supported by the monastery, he taught physics, botany, and natural science courses at the secondary and university levels. In 1856, he began a decade-long research pursuit involving inheritance patterns in honeybees and plants, ultimately settling on pea plants as his primary model system (a system with convenient characteristics used to study a specific biological phenomenon to be applied to other

systems). In 1865, Mendel presented the results of his experiments with nearly 30,000 pea plants to the local Natural History Society. He demonstrated that traits are transmitted from parents to offspring independently of other traits and in dominant and recessive patterns. In 1866, he published his work, Experiments in Plant Hybridization, in the proceedings of the Natural History Society of Brünn.

This page titled 4.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **8.1: Mendel's Experiments** by OpenStax is licensed CC BY 4.0.

4.2: Introduction

Learning Objectives

- Discuss Gregor Mendel's discovery of the Principles of Heredity and comment on why he was successful.
- Establish and analyze monohybrid crosses.
- Describe how monohybrid crosses reveal the Principle of Segregation and the Concept of Dominance.
- Apply a test cross to identify genotypes in individuals expressing one dominant trait.
- Apply probability to predict genotypic and phenotypic ratios in monohybrid crosses.

Gregor Mendel's Principles of Inheritance

Gregor Mendel's principles of inheritance form the cornerstone of modern genetics. Mendel (**Figure 1.1.1**) was an Augustinian monk and plant breeder who conducted a series of simple, yet elegant, experiments in 1865. He was one of the first to take a quantitative, scientific approach to the study of heredity.



Figure 1.1.1 Gregor Johann Mendel (1822-1884). Mendel was an Augustinian Friar, who lived in Moravia (now part of the Czech Republic). He published his work in 1866 on what has become known as the laws of Mendelian Inheritance. [Long description]

Mendel started with well-characterized strains, repeated his experiments many times, and kept careful records of his observations. He used the garden pea plant (*Pisum sativum* – **Figure 1.1.2**) with which to conduct his studies.



Figure 1.1.2 Pea Plants Were Used by Gregor Mendel to Discover Fundamental Laws of Genetics [Long description]

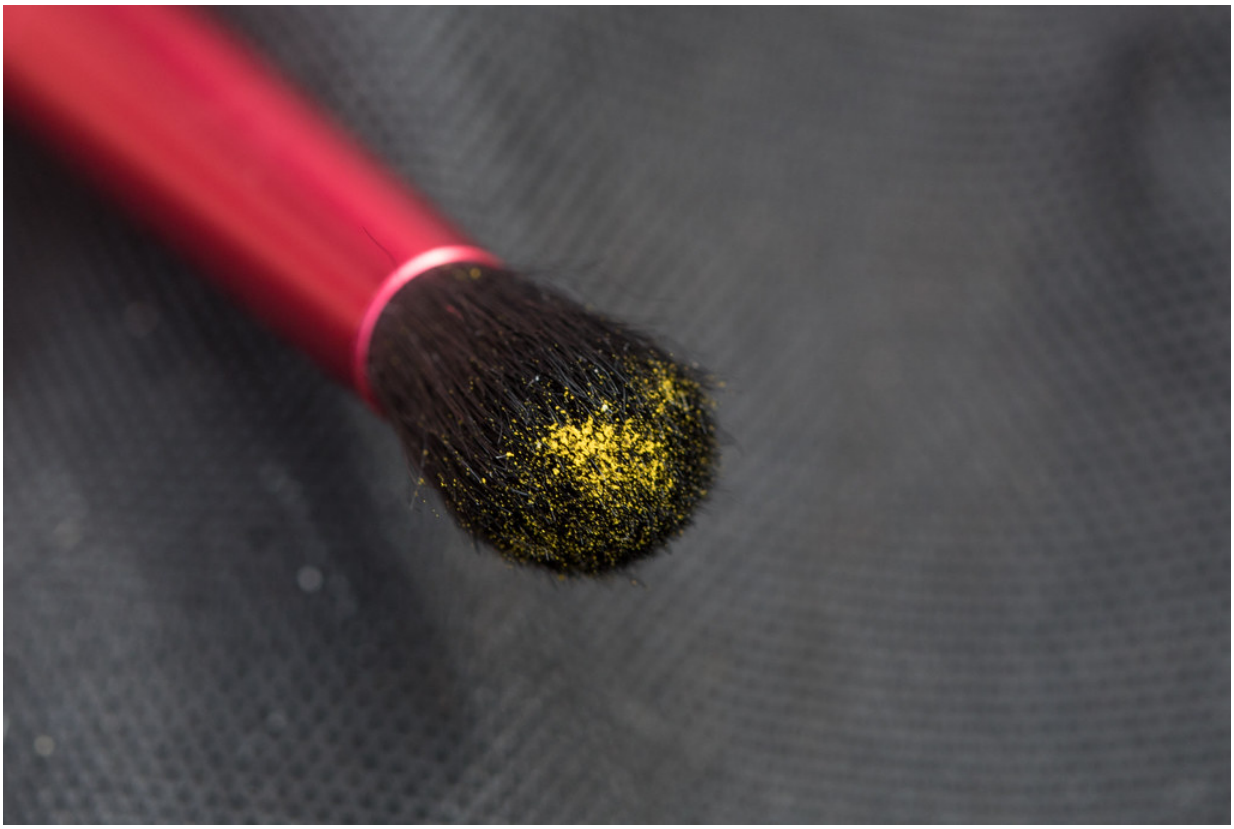


Figure 1.1.3 Closeup of Pollen on a Cosmetic Brush for Use in Hand-Pollinating [Long description]

This garden pea plant was an excellent choice for Mendel, for the following four reasons:

1. Peas had been shown to be true-breeding (all offspring will have the same characteristic generation after generation).
2. Peas exhibit a variety of contrasting traits (purple vs. white flowers; round vs. wrinkled seeds).
3. The shape of the pea flower protected it from foreign pollen. Peas usually reproduce by self-pollination, in which pollen produced by a flower fertilizes eggs in the same flower.
4. Pea plants grow quickly and do not require much space.

The seven (7) traits that Mendel studied are as follows:

- Form of ripe seed (R) — smooth or wrinkled
- Colour of seed albumen (Y) — yellow or green
- Colour of flower (P) — purple or white
- Form of ripe pods (I) — inflated or constricted
- Colour of unripe pods (G) — green or yellow
- Position of flowers (A) — axial or terminal
- Length of stem (T) — tall or dwarf

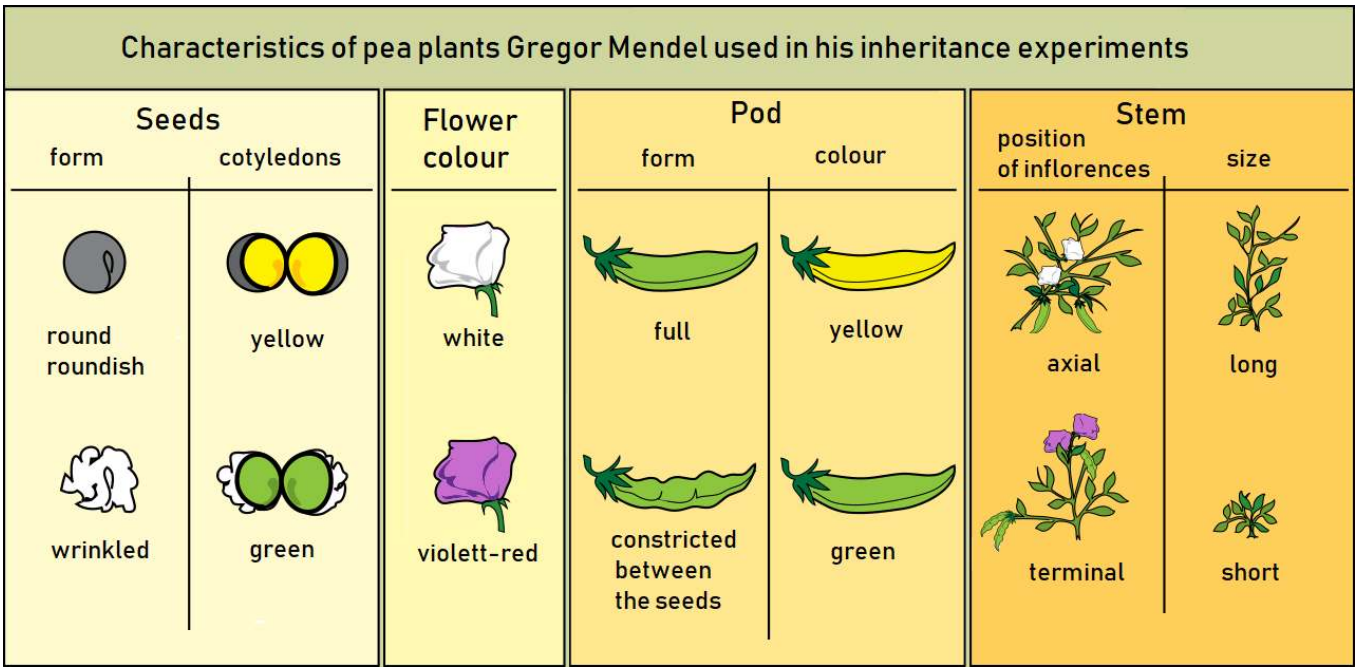


Figure 1.1.4 Characteristics of Pea Plants Gregor Mendel Used in Inheritance Experiments. These are the seven traits in garden peas that Gregor Mendel, the founder of genetics, experimented with as he was studying the principles of genetics in the 1850s & 1860s. [Long description]

After eight years of tedious experiments with these plants, Mendel proposed his foundational principles of inheritance. Mendel showed that white-flowered plants could be produced by crossing two purple-flowered plants, but only if the purple-flowered plants themselves had at least one white-flowered parent (Figure 1.1.5). This was evidence that a discrete genetic factor that produced white-flowers had *not* blended irreversibly with the factor for purple-flowers. Mendel's observations disproved blending inheritance and favoured an alternative concept, called **particulate inheritance**, in which heredity is the product of discrete factors that control independent traits.

Through careful study of patterns of inheritance, Mendel recognized that a single trait could exist in different versions, or **alleles** , even within an individual plant or animal. For example, he found two allelic forms of a gene for seed colour: one allele gave green seeds, and the other gave yellow seeds. Mendel also observed that although different alleles could influence a single trait, they remained indivisible and could be inherited separately. This is the basis of **Mendel's First Law** , also called **The Law of Equal Segregation** , which states that, during gamete formation the two alleles at a gene locus segregate from each other, and each gamete has an equal probability of containing either allele.

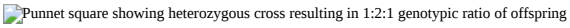


Figure 1.1.5 Punnett Square Demonstrating Mendel's Cross Between Purple and White Flowers. This Punnett Square illustrates a 3:1 ratio of purple: white flowers in the offspring (F1 generation). [Long description]

Genes

Mendel first made his discoveries of inheritance in the 1850s. In his 1866 publication, *Experiments on Plant Hybridization*, he didn't use the word "gene" as the fundamental unit of heredity because it wasn't coined until 1909 by Danish botanist Wilhelm Johannsen. Thomas Hunt Morgan proposed that genes resided on chromosomes in 1910, and occupied distinct regions on those chromosomes. DNA as a substance was discovered in the 1860s, but it took until the 1940s to realize that DNA was the molecule that contained the genetic information. Then in the 1950s Watson and Crick discovered the structure of DNA. Take a look at the video, "A History of Research on Genetics" by Sigma Documentaries on YouTube, which summarizes some of these landmark studies.

Watch this video, *History of Research on Genetics*, by Sigma Documentaries (2017) on YouTube.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=190

Key Milestones in Genetics and Molecular Biology

An interactive or media element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=190

Note: If you are not using the online version of this guide, you can find the timeline in Appendix A:Key Milestones in Genetics and Molecular Biology Timeline

Media Attributions

- **Figure 1.1.1**Portrait of Mendel in Oval, by Hugo Iltis/Wellcome, CC BY 4.0, via Wikipedia
- **Figure 1.1.2**File:Peultjes peultjes Pisum sativum mange-tout by Rasbak, CC BY-SA 3.0 via Wikimedia Commons
- **Figure 1.1.3** closeup of Ka‘ū silversword pollen by Wei/ Hawai‘i Volcanoes NPS (2017), is CC PDM 1.0 (public domain mark), via Flickr
- **Figure 1.1.4**Gregor Mendel, Characteristics of Pea Plants (reworked by Sciencia58), LadyofHats, CC0, via Wikimedia Commons
- **Figure 1.1.5**Punnett square Mendel flowers by Madprime is licensed under CC BY-SA 3.0, via Wikimedia Commons

References

A&E Television Networks. (2009, November 24). James D. Watson and Francis H.C. Crick (image). *Chemical structure of DNA discovered* . History.com. https://www.history.com/this-day-in-history/watson-and-crick-discover-chemical-structure-of-dna.

Hurmerinta S. (2014, August). Day 227/366 [image]. Flickr. https://www.flickr.com/photos/explodingfish/8156844439/in/photolist-dqMXEX

Hugo Iltis/Wellcome. (2020, September 9). File:Portrait of Mendel in oval. Wellcome M0014816.jpg [image]. *Wikimedia Commons*. commons.wikimedia.org/w/index.php?title=File:Portrait_of_Mendel_in_oval._Wellcome_M0014816.jpg&oldid=453284612.

John Murray (Publisher). (2020, November 24). File:Origin of Species title page [digital image]. *Wikimedia Commons*. commons.wikimedia.org/w/index.php?title=File:Origin_of_Species_title_page.jpg&oldid=514806088.

Madprime. (2020, September 15). File:Punnett square mendel flowers.svg [image]. *Wikimedia Commons*. commons.wikimedia.org/w/index.php?title=File:Punnett_square_mendel_flowers.svg&oldid=460082081.

Maramorosch, K. *(1953).* Martha Chase and Alfred D. Hershey [photograph]. Linus *Pauling and the Race for DNA: A Documentary History.* Special Collections & Archives Research Center, Oregon State University Libraries. http://scarc.library.oregonstate.edu/coll/pauling/dna/pictures/portrait-hersheychase.html

Ruiz Villarreal, M. [user: ladyofhats]. (2018, October 9). File:Gregor Mendel – characteristics of pea plants – english [digital image – reworked by user: Sciencia58]. *Wikimedia Commons*. commons.wikimedia.org/wiki/File:Gregor_Mendel_-_characteristics_of_pea_plants_-_english.png

Rasbak. (2020, October 13). File:Peultjes peultjes Pisum sativum mange-tout.jpg. *Wikimedia Commons*. commons.wikimedia.org/w/index.php?title=File:Peultjes_peultjes_Pisum_sativum_mange-tout.jpg&oldid=488730288.

Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., Yael Avissar, Y. (2016, October 21). Figure 13.3 Inheritance patterns of unlinked and linked genes [digital image]. In *Biology*. OpenStax. https://openstax.org/books/biology/pages/13-1-chromosomal-theory-and-genetic-linkage. Access for free at https://openstax.org/books/biology/pages/1-introduction

Sciencia58. (2021, January 4). File:Gregor Mendel – characteristics of pea plants – english.png (image). *Wikimedia Commons*.commons.wikimedia.org/w/index.php?title=File:Gregor_Mendel_-_characteristics_of_pea_plants_-_english.png&oldid=524031116 (Original by Mariana Ruiz Villarreal/ LadyofHats)

Sigma Documentaries. (2017, January 29). A History of Research on Genetics (video). YouTube. https://www.youtube.com/watch?v=LMEy9uiu-Wk&feature=youtu.be.

Wei, J./Hawaii Volcanoes National Park. (2017). Closeup of Ka‘ū silversword pollen on a cosmetic brush for use in hand-pollinating (NPS photo). Flickr. https://www.flickr.com/photos/144356245@N06/35215975004

Wikipedia contributors. (2021, February 16). Thomas Hunt Morgan. *Wikipedia*.en.Wikipedia.org/w/index.php?title=Thomas_Hunt_Morgan&oldid=1007173283

Wikipedia contributors. (2020, December 10). Wilhelm Johannsen. *Wikipedia*. en.Wikipedia.org/w/index.php?title=Wilhelm_Johannsen&oldid=993478431

Long Descriptions

- **Figure 1.1.1** Headshot of Gregor Mendel, the Augustinian monk, pictured with a receding hairline, wearing a black robe, spectacles and a cross around his neck. [Back to Figure 1.1.1]
- **Figure 1.1.2** Garden peas appear with many green leaves, tall green stems, green pea pods, and a few white flowers. [Back to Figure 1.1.2]
- **Figure 1.1.3** A medium bristled paint brush with yellow pollen dust on the tip of the brush is used for hand-pollination. [Back to Figure 1.1.3]
- **Figure 1.1.4** The seven traits in garden peas Mendel experimented with are: 1) round and wrinkled seeds; 2) yellow and green seeds; 3) white and purple flowers; 4) full and constricted pods; 5) yellow and green pods; 6) axial and terminal stems; and 7) long and short stems. [Back to Figure 1.1.4]
- **Figure 1.1.5** A Punnett square illustrates the segregation of alleles in male and female parents in a monohybrid cross, to produce four offspring in a 3:1 phenotypic ratio. A purple flowered male is crossed with a purple flowered female. Both parents are heterozygous (Bb). Offspring are 1 homozygous dominant (BB), 2 heterozygous (BB) and 1 homozygous recessive (bb). [Back to Figure 1.1.5]

This page titled 4.2: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **1.1: Introduction** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

4.3: Mendel's Experiments

Learning Objectives

- Describe the different laws of genetics developed by Mendel.
- Utilize Punnett square to perform and follow a monohybrid cross.
- Explain how pedigree trees are constructed and how to use them to study genetic diseases.

Mendel's Crosses

Mendel's seminal work was accomplished using the garden pea, *Pisum sativum*, to study inheritance. This species naturally self-fertilizes, meaning that pollen encounters ova within the same flower. The flower petals remain sealed tightly until pollination is completed to prevent the pollination of other plants. The result is highly inbred, or "**true-breeding**," pea plants. These are plants that always produce offspring that look like the parent. By experimenting with true-breeding pea plants, Mendel avoided the appearance of unexpected traits in offspring that might occur if the plants were not true breeding. The garden pea also grows to maturity within one season, meaning that several generations could be evaluated over a relatively short time. Finally, large quantities of garden peas could be cultivated simultaneously, allowing Mendel to conclude that his results did not come about simply by chance.

Monohybrid Crosses

Mendel performed **hybridizations**, which involve mating two true-breeding individuals that have different traits. In the pea, which is naturally self-pollinating, this is done by manually transferring pollen from the anther of a mature pea plant of one variety to the stigma of a separate mature pea plant of the second variety.

Plants used in first-generation crosses were called **P**, or **parental generation**, plants (Figure 4.3.1). Mendel collected the seeds produced by the P plants that resulted from each cross and grew them the following season. These offspring were called the $\mathbf{F_1}$, or the **first filial** (filial = daughter or son), generation. Once Mendel examined the characteristics in the F_1 generation of plants, he allowed them to self-fertilize naturally. He then collected and grew the seeds from the F_1 plants to produce the $\mathbf{F_2}$, or **second filial**, generation. Mendel's experiments extended beyond the F_2 generation to the F_3 generation, F_4 generation, and so on, but it was the ratio of characteristics in the P, F_1, and F_2 generations that were the most intriguing and became the basis of Mendel's postulates.

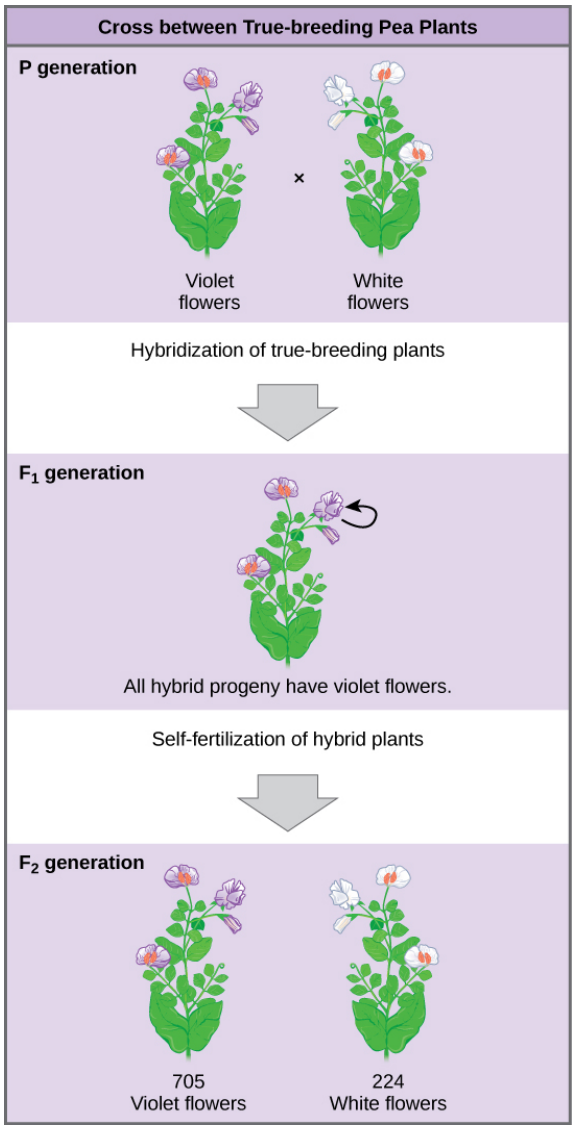


Figure 4.3.1**:** Mendel's process for performing crosses included examining flower color.

A **monohybrid cross** is one in which both parents are heterozygous (or a hybrid) for a single (mono) trait. The figure above shows the trait petal colour.

The ideal gas law is easy to remember and apply in solving problems, as long as you get the **proper values a**

Example 4.3.1

Add example text here.

Solution

Add example text here.

Definition: Term

F1: first filial generation in a cross; the offspring of the parental generation.

F2: second filial generation produced when F_1 individuals are self-crossed or fertilized with each other.

Hybridization: process of mating two individuals that differ with the goal of achieving a certain characteristic in their offspring.

Monohybrid: result of a cross between two true-breeding parents that express different traits for only one characteristic.

P or P0: parental generation in a cross.

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).
-

This page titled 4.3: Mendel's Experiments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **8.1: Mendel's Experiments** by OpenStax is licensed CC BY 4.0.
- **3.4: Crossing Techniques Used in Classical Genetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.4: Punnett Squares and Test Crosses

Punnett Squares

Given the genotypes of any two parents, we can predict all of the possible genotypes of the offspring. Furthermore, if we also know the dominance relationships for all of the alleles, we can predict the phenotypes of the offspring. A convenient method for calculating the expected genotypic and phenotypic ratios from a cross was invented by Reginald Punnett. A **Punnett square** is a matrix in which all of the possible gametes produced by one parent are listed along one axis, and the gametes from the other parent are listed along the other axis. Each possible combination of gametes is listed at the intersection of each row and column. The F_1 cross from Figure 4.4.6b would be drawn as in Figure 4.4.7. Punnett squares can also be used to calculate the frequency of offspring. The frequency of each offspring is the frequency of the male gametes multiplied by the frequency of the female gamete.

	A	***a***
A	*AA*	*Aa*
a	*Aa*	*aa*

Figure 4.4.1***:*** *A Punnett square showing a monohybrid cro*ss. (Original-Deholos (Fireworks)-CC:AN)

Test Crosses

Knowing the genotypes of an individual is usually an important part of a genetic experiment. However, genotypes cannot be observed directly; they must be inferred based on phenotypes. Because of dominance, it is often not possible to distinguish between a heterozygote and a homozgyote based on phenotype alone (e.g. see the purple-flowered F_2 plants in Figure 4.4.2b). To determine the genotype of a specific individual, a **test cross** can be performed, in which the individual with an uncertain genotype is crossed with an individual that is homozygous recessive for all of the loci being tested.

For example, if you were given a pea plant with purple flowers it might be a homozygote (*AA*) or a heterozygote (*Aa*). You could cross this purple-flowered plant to a white-flowered plant as a **tester**, since you know the genotype of the tester is *aa*. Depending on the genotype of the purple-flowered parent (Figure 4.4.2), you will observe different phenotypic ratios in the F_1 generation. If the purple-flowered parent was a homozgyote, all of the F_1 progeny will be purple. If the purple-flowered parent was a heterozygote, the F_1 progeny should segregate purple-flowered and white-flowered plants in a 1:1 ratio.

	A	***A***
a	*Aa*	*Aa*
a	*Aa*	*Aa*

	A	***a***
a	*Aa*	*aa*
a	*Aa*	*aa*

Figure 4.4.2***:*** *Punnett Squares showing the two possible outcomes of a test cross*. (Original-Deholos (Fireworks)-CC:AN)

This page titled 4.4: Punnett Squares and Test Crosses is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.4: Crossing Techniques Used in Classical Genetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.5: Laws of Inheritance

The seven characteristics that Mendel evaluated in his pea plants were each expressed as one of two versions, or traits. Mendel deduced from his results that each individual had two discrete copies of the characteristic that are passed individually to offspring. We now call those two copies genes, which are carried on chromosomes. The reason we have two copies of each gene is that we inherit one from each parent. In fact, it is the chromosomes we inherit and the two copies of each gene are located on paired chromosomes. Recall that in meiosis these chromosomes are separated out into haploid gametes. This separation, or segregation, of the homologous chromosomes means also that only one of the copies of the gene gets moved into a gamete. The offspring are formed when that gamete unites with one from another parent and the two copies of each gene (and chromosome) are restored.

For cases in which a single gene controls a single characteristic, a diploid organism has two genetic copies that may or may not encode the same version of that characteristic. For example, one individual may carry a gene that determines white flower color and a gene that determines violet flower color. Gene variants that arise by mutation and exist at the same relative locations on homologous chromosomes are called alleles. Mendel examined the inheritance of genes with just two allele forms, but it is common to encounter more than two alleles for any given gene in a natural population.

Phenotypes and Genotypes

Two alleles for a given gene in a diploid organism are expressed and interact to produce physical characteristics. The observable traits expressed by an organism are referred to as its phenotype. An organism's underlying genetic makeup, consisting of both the physically visible and the non-expressed alleles, is called its genotype. Mendel's hybridization experiments demonstrate the difference between phenotype and genotype. For example, the phenotypes that Mendel observed in his crosses between pea plants with differing traits are connected to the diploid genotypes of the plants in the P, F_1, and F_2 generations. We will use a second trait that Mendel investigated, seed color, as an example. Seed color is governed by a single gene with two alleles. The yellow-seed allele is dominant and the green-seed allele is recessive. When true-breeding plants were cross-fertilized, in which one parent had yellow seeds and one had green seeds, all of the F_1hybrid offspring had yellow seeds. That is, the hybrid offspring were phenotypically identical to the true-breeding parent with yellow seeds. However, we know that the allele donated by the parent with green seeds was not simply lost because it reappeared in some of the F_2 offspring (Figure 4.5.1). Therefore, the F_1 plants must have been genotypically different from the parent with yellow seeds.

The P plants that Mendel used in his experiments were each homozygous for the trait he was studying. Diploid organisms that are homozygous for a gene have two identical alleles, one on each of their homologous chromosomes. The genotype is often written as *YY* or *yy*, for which each letter represents one of the two alleles in the genotype. The dominant allele is capitalized and the recessive allele is lower case. The letter used for the gene (seed color in this case) is usually related to the dominant trait (yellow allele, in this case, or "*Y*"). Mendel's parental pea plants always bred true because both produced gametes carried the same allele. When P plants with contrasting traits were cross-fertilized, all of the offspring were heterozygous for the contrasting trait, meaning their genotype had different alleles for the gene being examined. For example, the F_1 yellow plants that received a *Y* allele from their yellow parent and a *y* allele from their green parent had the genotype *Yy*.

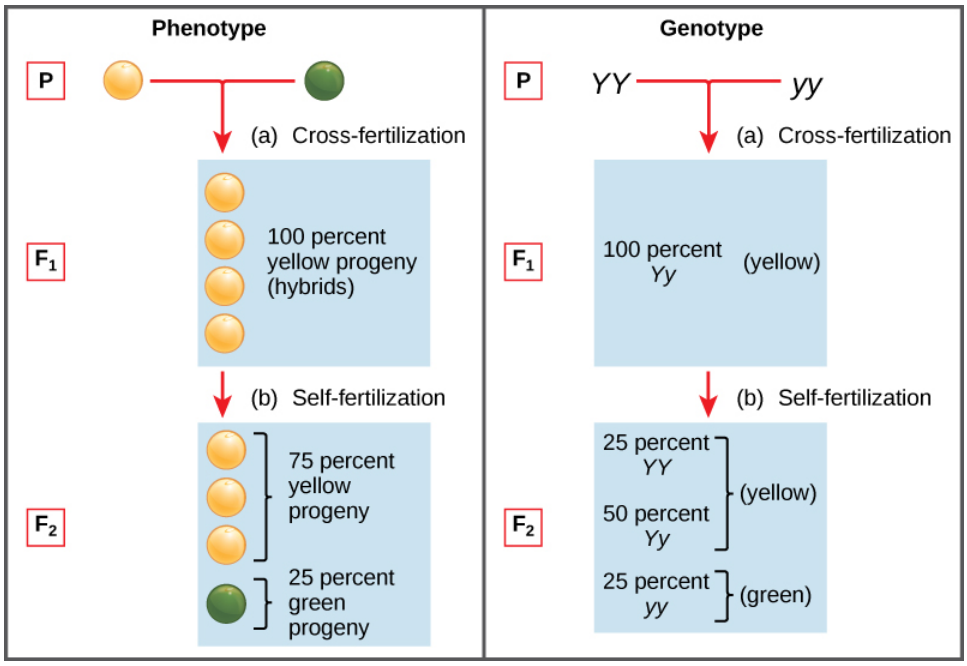


Figure 4.5.1**:** Phenotypes are physical expressions of traits that are transmitted by alleles. Capital letters represent dominant alleles and lowercase letters represent recessive alleles. The phenotypic ratios are the ratios of visible characteristics. The genotypic ratios are the ratios of gene combinations in the offspring, and these are not always distinguishable in the phenotypes.

Law of Dominance

Our discussion of homozygous and heterozygous organisms brings us to why the F_1 heterozygous offspring were identical to one of the parents, rather than expressing both alleles. In all seven pea-plant characteristics, one of the two contrasting alleles was dominant, and the other was recessive. Mendel called the dominant allele the expressed unit factor; the recessive allele was referred to as the latent unit factor. We now know that these so-called unit factors are actually genes on homologous chromosomes. For a gene that is expressed in a dominant and recessive pattern, homozygous dominant and heterozygous organisms will look identical (that is, they will have different genotypes but the same phenotype), and the recessive allele will only be observed in homozygous recessive individuals (Table 4.5.1).

Table 4.5.1***:*** Correspondence between Genotype and Phenotype for a Dominant-Recessive Characteristic.

	Homozygous	Heterozygous	Homozygous
Genotype	*YY*	*Yy*	*yy*
Phenotype	yellow	yellow	green

Mendel's law of dominance states that in a heterozygote, one trait will conceal the presence of another trait for the same characteristic. For example, when crossing true-breeding violet-flowered plants with true-breeding white-flowered plants, all of the offspring were violet-flowered, even though they all had one allele for violet and one allele for white. Rather than both alleles contributing to a phenotype, the dominant allele will be expressed exclusively. The recessive allele will remain latent, but will be transmitted to offspring in the same manner as that by which the dominant allele is transmitted. The recessive trait will only be expressed by offspring that have two copies of this allele (Figure 4.5.2), and these offspring will breed true when self-crossed.



Figure 4.5.2**:** The allele for albinism, expressed here in humans, is recessive. Both of this child's parents carried the recessive allele.

Monohybrid Cross and the Punnett Square

When fertilization occurs between two true-breeding parents that differ by only the characteristic being studied, the process is called a monohybrid cross, and the resulting offspring are called monohybrids. Mendel performed seven types of monohybrid crosses, each involving contrasting traits for different characteristics. Out of these crosses, all of the F_1 offspring had the phenotype of one parent, and the F_2 offspring had a 3:1 phenotypic ratio. On the basis of these results, Mendel postulated that each parent in the monohybrid cross contributed one of two paired unit factors to each offspring, and every possible combination of unit factors was equally likely.

The results of Mendel's research can be explained in terms of probabilities, which are mathematical measures of likelihood. The probability of an event is calculated by the number of times the event occurs divided by the total number of opportunities for the event to occur. A probability of one (100 percent) for some event indicates that it is guaranteed to occur, whereas a probability of zero (0 percent) indicates that it is guaranteed to not occur, and a probability of 0.5 (50 percent) means it has an equal chance of occurring or not occurring.

To demonstrate this with a monohybrid cross, consider the case of true-breeding pea plants with yellow versus green seeds. The dominant seed color is yellow; therefore, the parental genotypes were *YY* for the plants with yellow seeds and *yy* for the plants with green seeds. A Punnett square, devised by the British geneticist Reginald Punnett, is useful for determining probabilities because it is drawn to predict all possible outcomes of all possible random fertilization events and their expected frequencies. Figure 4.5.5 shows a Punnett square for a cross between a plant with yellow peas and one with green peas. To prepare a Punnett square, all possible combinations of the parental alleles (the genotypes of the gametes) are listed along the top (for one parent) and side (for the other parent) of a grid. The combinations of egg and sperm gametes are then made in the boxes in the table on the basis of which alleles are combining. Each box then represents the diploid genotype of a zygote, or fertilized egg. Because each possibility is equally likely, genotypic ratios can be determined from a Punnett square. If the pattern of inheritance (dominant and recessive) is known, the phenotypic ratios can be inferred as well. For a monohybrid cross of two true-breeding parents, each parent contributes one type of allele. In this case, only one genotype is possible in the F_1 offspring. All offspring are *Yy* and have yellow seeds.

When the F_1 offspring are crossed with each other, each has an equal probability of contributing either a *Y* or a *y* to the F_2 offspring. The result is a 1 in 4 (25 percent) probability of both parents contributing a *Y*, resulting in an offspring with a yellow phenotype; a 25 percent probability of parent A contributing a *Y* and parent B a *y*, resulting in offspring with a yellow phenotype; a 25 percent probability of parent A contributing a *y* and parent B a *Y*, also resulting in a yellow phenotype; and a (25 percent) probability of both parents contributing a *y*, resulting in a green phenotype. When counting all four possible outcomes, there is a 3 in 4 probability of offspring having the yellow phenotype and a 1 in 4 probability of offspring having the green phenotype. This explains why the results of Mendel's F_2 generation occurred in a 3:1 phenotypic ratio. Using large numbers of crosses, Mendel was able to calculate probabilities, found that they fit the model of inheritance, and use these to predict the outcomes of other crosses.

Law of Segregation

Observing that true-breeding pea plants with contrasting traits gave rise to F_1 generations that all expressed the dominant trait and F_2 generations that expressed the dominant and recessive traits in a 3:1 ratio, Mendel proposed the law of segregation. This law states that paired unit factors (genes) must segregate equally into gametes such that offspring have an equal likelihood of inheriting either factor. For the F_2 generation of a monohybrid cross, the following three possible combinations of genotypes result: homozygous dominant, heterozygous, or homozygous recessive. Because heterozygotes could arise from two different pathways (receiving one dominant and one recessive allele from either parent), and because heterozygotes and homozygous dominant individuals are phenotypically identical, the law supports Mendel's observed 3:1 phenotypic ratio. The equal segregation of alleles is the reason we can apply the Punnett square to accurately predict the offspring of parents with known genotypes. The physical basis of Mendel's law of segregation is the first division of meiosis in which the homologous chromosomes with their different versions of each gene are segregated into daughter nuclei. This process was not understood by the scientific community during Mendel's lifetime (Figure 4.5.3).

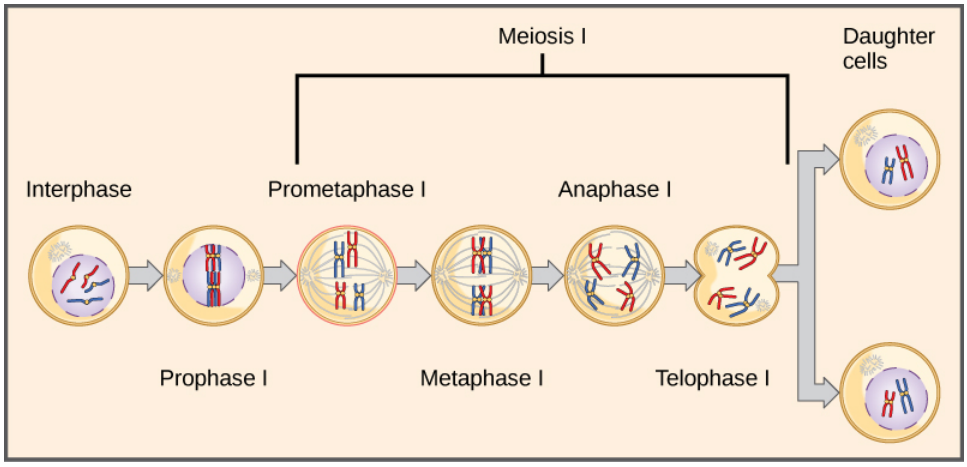


Figure 4.5.3: The first division in meiosis is shown.

Test Cross

Beyond predicting the offspring of a cross between known homozygous or heterozygous parents, Mendel also developed a way to determine whether an organism that expressed a dominant trait was a heterozygote or a homozygote. Called the test cross, this technique is still used by plant and animal breeders. In a test cross, the dominant-expressing organism is crossed with an organism that is homozygous recessive for the same characteristic. If the dominant-expressing organism is a homozygote, then all F_1 offspring will be heterozygotes expressing the dominant trait (Figure 4.5.4). Alternatively, if the dominant-expressing organism is a heterozygote, the F_1 offspring will exhibit a 1:1 ratio of heterozygotes and recessive homozygotes (Figure 4.5.4). The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.

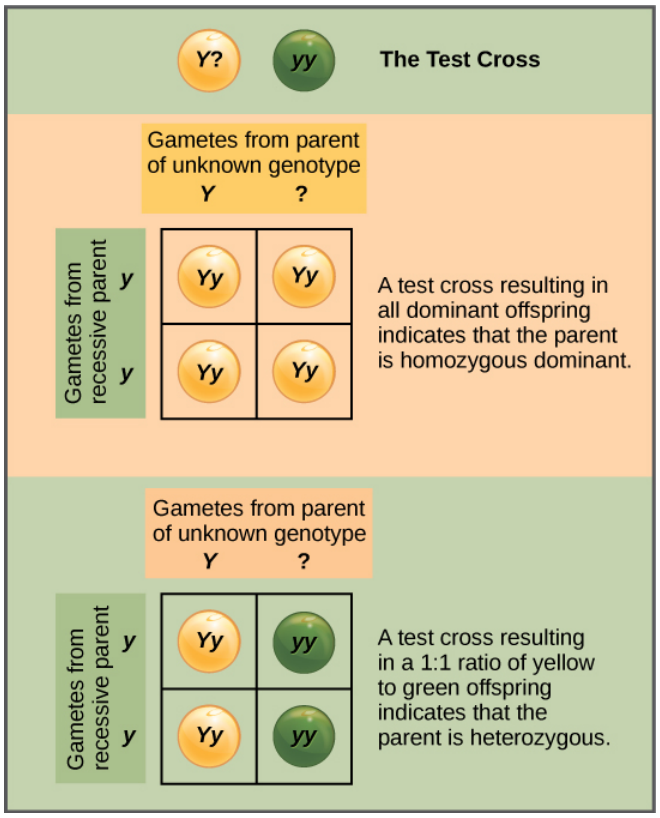


Figure 4.5.4**:** A test cross can be performed to determine whether an organism expressing a dominant trait is a homozygote or a heterozygote.

ART CONNECTION

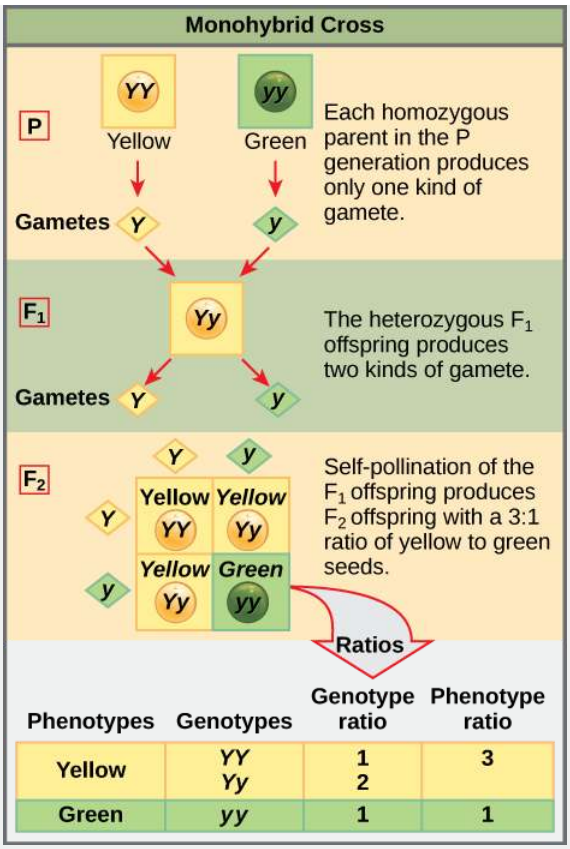


Figure 4.5.5**:** This Punnett square shows the cross between plants with yellow seeds and green seeds. The cross between the true-breeding P plants produces F_1 heterozygotes that can be self-fertilized. The self-cross of the F_1 generation can be analyzed with a Punnett square to predict the genotypes of the F_2 generation. Given an inheritance pattern of dominant–recessive, the genotypic and phenotypic ratios can then be determined.

In pea plants, round peas (*R*) are dominant to wrinkled peas (*r*). You do a test cross between a pea plant with wrinkled peas (genotype *rr*) and a plant of unknown genotype that has round peas. You end up with three plants, all which have round peas. From this data, can you tell if the parent plant is homozygous dominant or heterozygous?

Law of Independent Assortment

Mendel's law of independent assortment states that genes do not influence each other with regard to the sorting of alleles into gametes, and every possible combination of alleles for every gene is equally likely to occur. Independent assortment of genes can be illustrated by the dihybrid cross, a cross between two true-breeding parents that express different traits for two characteristics. Consider the characteristics of seed color and seed texture for two pea plants, one that has wrinkled, green seeds (*rryy*) and another that has round, yellow seeds (*RRYY*). Because each parent is homozygous, the law of segregation indicates that the gametes for the wrinkled–green plant all are *ry*, and the gametes for the round–yellow plant are all *RY*. Therefore, the F_1 generation of offspring all are *RrYy* (Figure 4.5.6).

ART CONNECTION

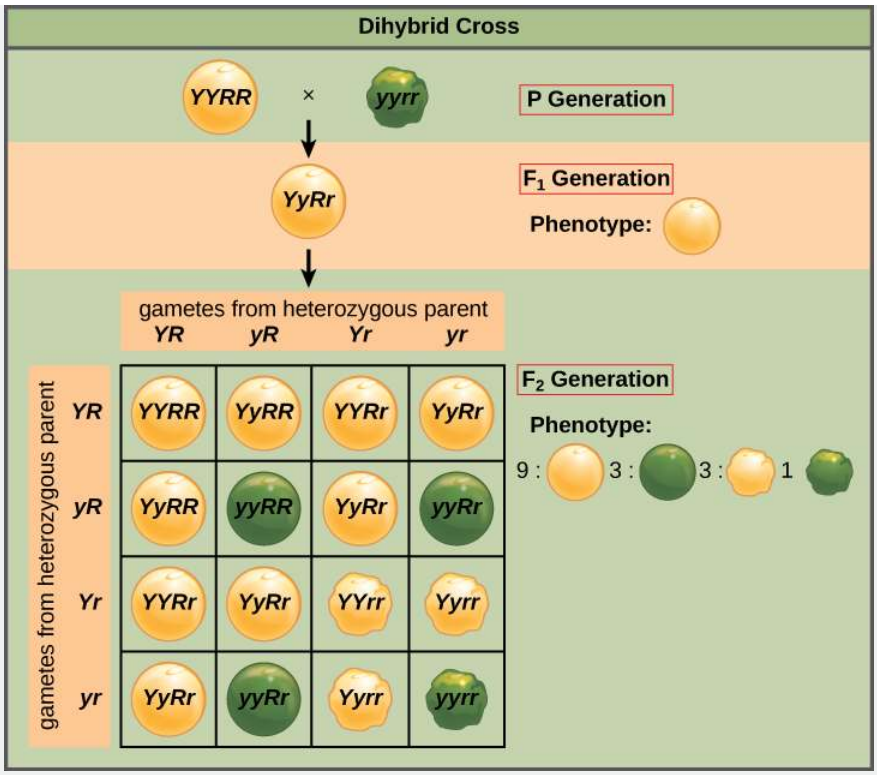


Figure 4.5.6**:** A dihybrid cross in pea plants involves the genes for seed color and texture. The P cross produces F_1 offspring that are all heterozygous for both characteristics. The resulting 9:3:3:1 F_2 phenotypic ratio is obtained using a Punnett square.

In pea plants, purple flowers (*P*) are dominant to white (*p*), and yellow peas (*Y*) are dominant to green (*y*). What are the possible genotypes and phenotypes for a cross between *PpYY* and *ppYy* pea plants? How many squares would you need to complete a Punnett square analysis of this cross?

The gametes produced by the F_1 individuals must have one allele from each of the two genes. For example, a gamete could get an *R* allele for the seed shape gene and either a *Y* or a *y* allele for the seed color gene. It cannot get both an *R* and an *r* allele; each gamete can have only one allele per gene. The law of independent assortment states that a gamete into which an *r* allele is sorted would be equally likely to contain either a *Y* or a *y* allele. Thus, there are four equally likely gametes that can be formed when the *RrYy* heterozygote is self-crossed, as follows: *RY*, *rY*, *Ry*, and *ry*. Arranging these gametes along the top and left of a 4 × 4 Punnett square (Figure 4.5.6) gives us 16 equally likely genotypic combinations. From these genotypes, we find a phenotypic ratio of 9 round–yellow:3 round–green:3 wrinkled–yellow:1 wrinkled–green (Figure 4.5.6). These are the offspring ratios we would expect, assuming we performed the crosses with a large enough sample size.

The physical basis for the law of independent assortment also lies in meiosis I, in which the different homologous pairs line up in random orientations. Each gamete can contain any combination of paternal and maternal chromosomes (and therefore the genes on them) because the orientation of tetrads on the metaphase plane is random (Figure 4.5.7).

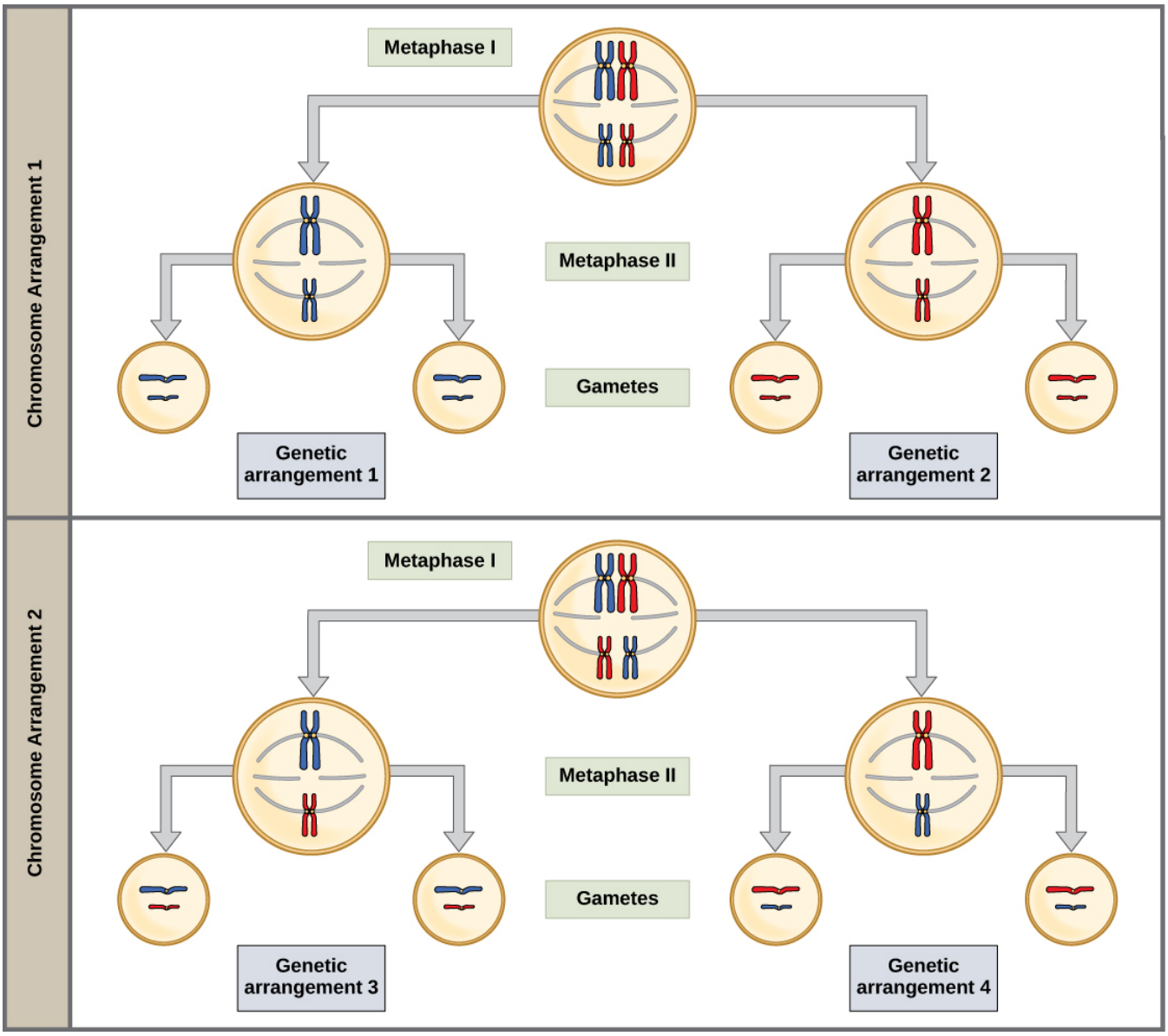


Figure 4.5.7**:** The random segregation into daughter nuclei that happens during the first division in meiosis can lead to a variety of possible genetic arrangements.

Section Summary

When true-breeding, or homozygous, individuals that differ for a certain trait are crossed, all of the offspring will be heterozygous for that trait. If the traits are inherited as dominant and recessive, the F_1 offspring will all exhibit the same phenotype as the parent homozygous for the dominant trait. If these heterozygous offspring are self-crossed, the resulting F_2 offspring will be equally likely to inherit gametes carrying the dominant or recessive trait, giving rise to offspring of which one quarter are homozygous dominant, half are heterozygous, and one quarter are homozygous recessive. Because homozygous dominant and heterozygous individuals are phenotypically identical, the observed traits in the F_2 offspring will exhibit a ratio of three dominant to one recessive.

Mendel postulated that genes (characteristics) are inherited as pairs of alleles (traits) that behave in a dominant and recessive pattern. Alleles segregate into gametes such that each gamete is equally likely to receive either one of the two alleles present in a diploid individual. In addition, genes are assorted into gametes independently of one another. That is, in general, alleles are not more likely to segregate into a gamete with a particular allele of another gene.

Art Connections

Figure 4.5.5: In pea plants, round peas (*R*) are dominant to wrinkled peas (*r*). You do a test cross between a pea plant with wrinkled peas (genotype *rr*) and a plant of unknown genotype that has round peas. You end up with three plants, all which have round peas. From this data, can you tell if the parent plant is homozygous dominant or heterozygous?

Answer

You cannot be sure if the plant is homozygous or heterozygous as the data set is too small: by random chance, all three plants might have acquired only the dominant gene even if the recessive one is present.

Figure 4.5.6: In pea plants, purple flowers (*P*) are dominant to white (*p*), and yellow peas (*Y*) are dominant to green (*y*). What are the possible genotypes and phenotypes for a cross between *PpYY* and *ppYy* pea plants? How many squares would you need to complete a Punnett square analysis of this cross?

Answer

The possible genotypes are *PpYY*, *PpYy*, *ppYY*, and *ppYy*. The former two genotypes would result in plants with purple flowers and yellow peas, while the latter two genotypes would result in plants with white flowers with yellow peas, for a 1:1 ratio of each phenotype. You only need a 2 × 2 Punnett square (four squares total) to do this analysis because two of the alleles are homozygous.

Glossary

allele
one of two or more variants of a gene that determines a particular trait for a characteristic

dihybrid
the result of a cross between two true-breeding parents that express different traits for two characteristics

genotype
the underlying genetic makeup, consisting of both physically visible and non-expressed alleles, of an organism

heterozygous
having two different alleles for a given gene on the homologous chromosomes

homozygous
having two identical alleles for a given gene on the homologous chromosomes

law of dominance
in a heterozygote, one trait will conceal the presence of another trait for the same characteristic

law of independent assortment
genes do not influence each other with regard to sorting of alleles into gametes; every possible combination of alleles is equally likely to occur

law of segregation
paired unit factors (i.e., genes) segregate equally into gametes such that offspring have an equal likelihood of inheriting any combination of factors

monohybrid
the result of a cross between two true-breeding parents that express different traits for only one characteristic

phenotype
the observable traits expressed by an organism

Punnett square

a visual representation of a cross between two individuals in which the gametes of each individual are denoted along the top and side of a grid, respectively, and the possible zygotic genotypes are recombined at each box in the grid

test cross

a cross between a dominant expressing individual with an unknown genotype and a homozygous recessive individual; the offspring phenotypes indicate whether the unknown parent is heterozygous or homozygous for the dominant trait

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).

This page titled 4.5: Laws of Inheritance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **8.2: Laws of Inheritance** by OpenStax is licensed CC BY 4.0.

4.6: Mendel's First Law

Character Traits Exist in Pairs that Segregate at Meiosis

Through careful study of patterns of inheritance, Mendel recognized that a single trait could exist in different versions, or **alleles**, even within an individual plant or animal. For example, he found two allelic forms of a gene for seed color: one allele gave green seeds, and the other gave yellow seeds. Mendel also observed that although different alleles could influence a single trait, they remained indivisible and could be inherited separately. This is the basis of **Mendel's First Law**, also called **The Law of Equal Segregation**, which states: during gamete formation, the two alleles at a gene locus segregate from each other; each gamete has an equal probability of containing either allele.

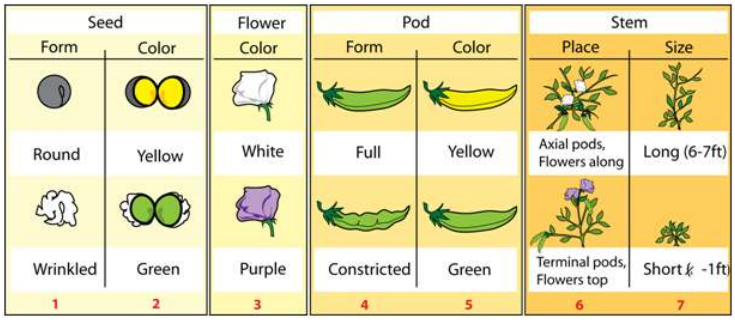


Figure 4.6.2: Seven traits Mendel studied in peas. (Wikipedia-Mariana Ruiz-PD)

Hetero-, Homo-, Hemizygosity

Mendel's First Law is especially remarkable because he made his observations and conclusions (1865) without knowing about the relationships between genes, chromosomes, and DNA. We now know the reason why more than one allele of a gene can be present in an individual: most eukaryotic organisms have at least two sets of homologous chromosomes. For organisms that are predominantly diploid, such as humans or Mendel's peas, chromosomes exist as pairs, with one homolog inherited from each parent. Diploid cells therefore contain two different alleles of each gene, with one allele on each member of a pair of homologous chromosomes. If both alleles of a particular gene are identical, the individual is said to be **homozygous** for that gene. On the other hand, if the alleles are different from each other, the genotype is **heterozygous**. In cases where there is only one copy of a gene present, for example if there is a deletion on the homologous chromosome, we use the term **hemizygous**.

Although a typical diploid individual can have at most two different alleles of a particular gene, many more than two different alleles can exist in a population of individuals. In a natural population the most common allelic form is usually called the **wild-type** allele. However, in many populations there can be multiple **variants** at the DNA sequence level that are visibly indistinguishable as all exhibit a normal, wild type appearance. There can also be various **mutant** alleles (in wild populations and in lab strains) that vary from wild type in their appearance, each with a different change at the DNA sequence level. Such collections of mutations are known as an **allelic series**.

This page titled 4.6: Mendel's First Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.1: Mendel's First Law** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.7: Relationships Between Genes, Genotypes and Phenotypes

Terminology

A specific position along a chromosome is called a **locus**. Each gene occupies a specific locus (so the terms locus and gene are often used interchangeably). Each locus will have an allelic form (allele). The complete set of alleles (at all loci of interest) in an individual is its **genotype**. Typically, when writing out a genotype, only the alleles at the locus (loci) of interest are considered – all the others are present and assumed to be wild type. The visible or detectable effect of these alleles on the structure or function of that individual is called its **phenotype** – what it looks like. The phenotype studied in any particular genetic experiment may range from simple, visible traits such as hair color, to more complex phenotypes including disease susceptibility or behavior. If two alleles are present in an individual, then various interactions between them may influence their expression in the phenotype.

Complete Dominance

Let us return to an example of a simple phenotype: flower color in Mendel's peas. We have already said that one allele as a homozygote produces purple flowers, while the other allele as a homozygote produces white flowers (see Figures 1.8 and 3.3). But what about an individual that has one purple allele and one white allele; what is the phenotype of an individual whose genotype is heterozygous? This can only be determined by experimental observation. We know from observation that individuals heterozygous for the purple and white alleles of the flower color gene have purple flowers. Thus, the allele associated with purple color is therefore said to be **dominant** to the allele that produces the white color. The white allele, whose phenotype is masked by the purple allele in a heterozygote, is **recessive** to the purple allele.

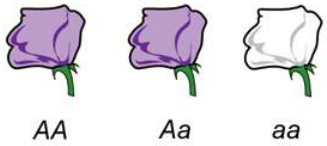


Figure 4.7.3: Relationship between genotype and phenotype for an allele that is completely dominant to another allele. Original-Deholos (Fireworks)-CC:AN)

To represent this relationship, often, a dominant allele will be represented by a capital letter (e.g. *A*) while a recessive allele will be represented in lower case (e.g. *a*). However, many different systems of genetic symbols are in use. The most common are shown in Table 4.7.1. Also note that genes and alleles are usually written in *italics* and chromosomes and proteins are not. For example, the *white* gene in *Drosophila melanogaster* on the X chromosome encodes a protein called WHITE.

Table 4.7.1: Examples of symbols used to represent genes and alleles.

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used because not all alleles show complete dominance and many genes have more than two alleles.
a^+ and a^1	Superscripts or subscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene, but on homologous chromosomes.

Incomplete Dominance

Besides dominance and recessivity, other relationships can exist between alleles. In **incomplete dominance** (also called **semi-dominance,** Figure 4.7.4), both alleles affect the trait additively, and the phenotype of the heterozygote is intermediate between either of the homozygotes. For example, alleles for color in carnation flowers (and many other species) exhibit incomplete dominance. Plants with an allele for red petals (A_1) and an allele for white petals (A_2) have pink petals. We say that the A_1 and the A_2 alleles show incomplete dominance because neither allele is completely dominant over the other.

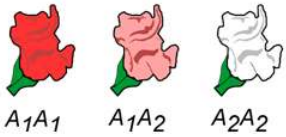


Figure 4.7.4: Relationship between genotype and phenotype for incompletely dominant alleles affecting petal colour in carnations. (Original-Deholos-CC:AN)

Co-Dominance

Co-dominance is another type of allelic relationship, in which a heterozygous individual expresses the phenotype of both alleles simultaneously. An example of co-dominance is found within the **ABO blood group** of humans. The *ABO* gene has three common alleles which were named (for historical reasons) I^A, I^B, and *i*. People homozygous for I^A or I^B display only A or B type antigens, respectively, on the surface of their blood cells, and therefore have either type A or type B blood (Figure 4.7.5). Heterozygous I^AI^B individuals have both A and B antigens on their cells, and so have type AB blood. Notice that the heterozygote expresses both alleles simultaneously, and is not some kind of novel intermediate between A and B. Co-dominance is therefore distinct from incomplete dominance, although they are sometimes confused.

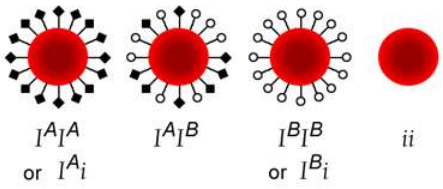


Figure 4.7.5***:*** *Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The I^A allele is completely dominant to the i allele. The I^B allele is completely dominant to the i allele. (Original-Deholos -CC:AN)*

It is also important to note that the third allele, *i*, does not make either antigen and is recessive to the other alleles. *I^A/i or I^B/i* individuals display A or B antigens respectively. People homozygous for the *i* allele have type O blood. This is a useful reminder that different types of dominance relationships can exist, even for alleles of the same gene. Many types of molecular markers, which we will discuss in a later chapter, display a co-dominant relationship among alleles.

This page titled 4.7: Relationships Between Genes, Genotypes and Phenotypes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.2: Relationships Between Genes, Genotypes and Phenotypes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

SECTION OVERVIEW

4.8: Pedigree Analysis

4.8.1: Symbols used in Pedigree Charts

4.8.2: Modes of Inheritance

4.8.3: Sporadic and Non-Heritable Diseases

4.8.4: Calculating Probabilities Using Pedigree Charts

4.8.5: Chapter 4 Summary

4.8.6: Key Terms in Chapter 4

4.8.7: Chapter 4 Study Questions

4.8.8: Chapter 4 Answers

This page titled 4.8: Pedigree Analysis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

4.8.1: Symbols used in Pedigree Charts

4.2 Symbols used in Pedigree Charts

Pedigreecharts are diagrams that show the phenotypes and/or genotypes for a particular organism, its ancestors, and descendants.

In order to glean useful information from a Pedigree Chart, the signs and symbols used to construct the chart must be properly recognized and interpreted. Sufficient information must be given via the Chart, and sometimes, supplementary information is also required (e.g., the frequency at which the particular trait is found in the population from which the family is derived). Geneticists use a standardized set of symbols to represent an individual's sex, family relationships, and phenotype. These diagrams are used to determine the **mode of inheritance** of a particular disease or trait, and to predict the probability of its appearance among offspring. Pedigree analysis is therefore an important tool in basic research, agriculture, and **genetic counselling**.

Each pedigree chart represents all the available information about the inheritance of a single trait (most often a disease) within a family. The pedigree chart is therefore drawn using phenotypic information, but there is always some possibility of errors in this information, especially when relying on family members' recollections or even clinical diagnoses. In real pedigrees, further complications can arise due to **incomplete penetrance** (including age of onset) and **variable expressivity** of disease alleles, but for the examples presented in this book, we will presume complete accuracy of the pedigrees — that is, the phenotype accurately reflects the genotype. A pedigree may be drawn when trying to determine the nature of a newly discovered disease, or when an individual with a family history of a disease wants to know the probability of passing the disease on to their children. In either case, a tree is drawn, as shown in **Figure 4.2.1**, with circles to represent females, and squares to represent males. Matings are drawn as a line joining a male and female, while a consanguineous mating (closely related) is two lines.

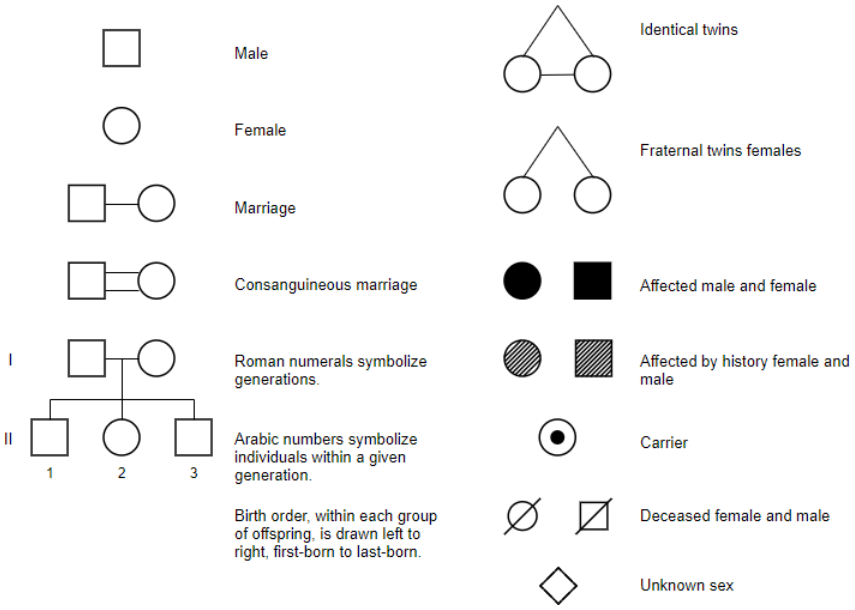


Figure 4.2.1 Symbols Used in Constructing a Pedigree Chart [Long description]

The affected individual that brings the family to the attention of a geneticist is called the **proband** (or propositus). If the individual is unaffected, they are called the consultand. If an individual is known to have symptoms of the disease (**affected**), the symbol is filled in. Sometimes, a half filled-in symbol is used to indicate a known **carrier** of a disease; this is someone who does not have any symptoms of the disease, but who passed the disease on to subsequent generations because they are a heterozygote. Female carriers of X-linked traits are indicated by a circle with a dot in the centre. Note, that when a pedigree is constructed, it is often unknown whether a particular individual is a carrier or not, so not all carriers are always explicitly indicated in a pedigree. For simplicity, in this chapter we will assume that the pedigrees presented are accurate, and represent fully penetrant traits. If possible, the male partner should be left of female partner on relationship line. Siblings should be listed from left to right in birth order, oldest to youngest.

Media Attributions

- **Figure 4.2.1** Symbols used in constructing a pedigree chart by John Locke (2019), CC BY-NC-SA 3.0

Long Description

- **Figure 4.2.1** Symbols are used to construct pedigree charts. The major ones are as follows: male – square; female – circle; marriage – square and circle linked by a horizontal line; affected male – shaded square; affected female – shaded circle; carrier – circle or square with a dot inside; deceased – circle or square with a line running through it; unknown sex – a diamond; Roman numerals symbolize generations, and Arabic numbers symbolize individuals within a given generation; birth order within each group of offspring is drawn left to right, from first-born to last-born. [Back to Figure 4.2.1]

This page titled 4.8.1: Symbols used in Pedigree Charts is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.2: Symbols used in Pedigree Charts** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

4.8.2: Modes of Inheritance

4.3 Modes of Inheritance

Usually, we are presented with a pedigree of an uncharacterized disease or trait, and one of the first tasks is to determine which modes of inheritance are possible, and then, which mode of inheritance is most likely. This information is essential in calculating the probability that the trait will be inherited in any future offspring. We will mostly consider five major types of inheritance: autosomal dominant (AD), autosomal recessive (AR), X-linked dominant (XD), X-linked recessive (XR), and Y-linked (Y) inheritance.

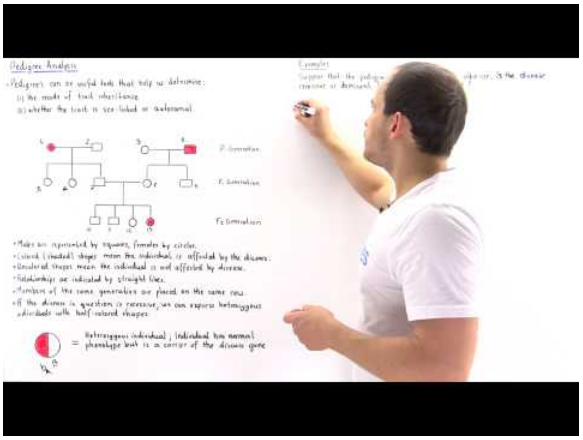
We generally make two assumptions in analyzing Pedigree Charts. These are as follows:

1. **Complete Penetrance** — an individual in the pedigree will be affected (express the phenotype associated with a trait) when the individual carries at least one dominant allele of a dominant trait, or two recessive alleles of a recessive a trait.
2. **Rare-in-Population** — generally, the trait in question is rare in the general population.

The following are some hints and clues to help us interpret Pedigree Charts:

1. An unaffected individual cannot have any alleles of a dominant trait (because a single allele of a dominant trait causes an individual to be affected).
2. Individuals marrying into the family are assumed to have no disease alleles – they will never be affected and can never be carriers of a recessive trait (because the trait is rare in the population).
3. An unaffected individual can be a carrier (have one allele) of a recessive trait (because two alleles of a recessive trait are required for an individual to be affected).
4. When a trait is X-linked, a single recessive allele is sufficient for a male to be affected (because the male is hemizygous – he only has one allele of an X-linked trait).
5. A father transmits his allele of X-linked genes to his daughters, but not his sons. A mother transmits an allele of X-linked genes to both her daughters and her sons.

Take a look at the following video, *Pedigree Analysis*, by AK Lecture Series (2015) on YouTube, which discusses Pedigree Charts and how to analyze them.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=547

Let us now take a look at the various modes of inheritance and typical pedigree charts which are characteristic of each mode.

Autosomal Dominant (AD)

When a disease is caused by a dominant allele of a gene, every person with that allele will show symptoms of the disease (assuming complete penetrance), and only one disease allele needs to be inherited for an individual to be affected. Thus, every affected individual must have an affected parent. A pedigree with affected individuals in every generation is typical of AD diseases.

However, beware that other modes of inheritance can also show the disease in every generation, as described below. It is also possible for an affected individual with an AD disease to have a family without any affected children, if the affected parent is a heterozygote. This is particularly true in small families, where the probability of every child inheriting the normal, rather than disease allele is not extremely small. Note that AD diseases are usually rare in populations, therefore affected individuals with AD diseases tend to be heterozygotes (otherwise, both parents would have had to been affected with the same rare disease). Huntington Disease, Achondroplastic dwarfism, and Polydactyly are all examples of human conditions that may follow an AD mode of inheritance.

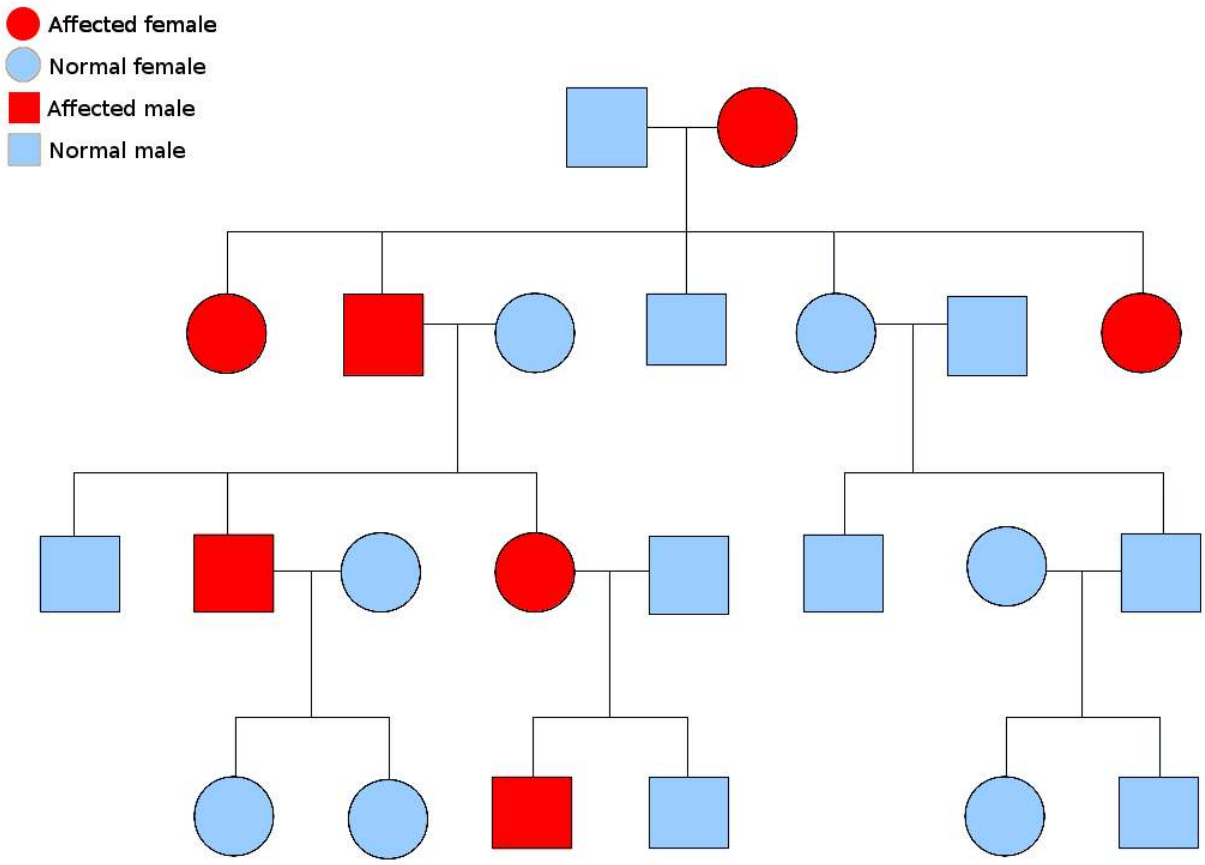


Figure 4.3.1 A Pedigree Chart Showing Autosomal Dominant Inheritance [Long description]

Example: Achondroplasia is a common form of dwarfism. *FGFR3* gene at 4p16 (chromosome 4, p arm, region 1, band 6) encodes a receptor protein that negatively regulates bone development. A specific base pair substitution in the gene makes an over-active protein and this results in shortened bones. Achondroplasia is considered autosomal dominant because the defective proteins made in A / a embryos halt bone growth prematurely. A / A embryos do not make enough limb bones to survive. Most, but not all dominant mutations are also recessive lethal. In achondroplasia, the A allele shows dominant visible phenotype (shortness) and recessive lethal phenotype.

X-Linked Dominant (XD)

In X-linked dominant inheritance, the gene responsible for the disease is located on the X-chromosome, and the allele that causes the disease is dominant to the normal allele in females. Because females have twice as many X-chromosomes as males, females tend to be more frequently affected than males in the population. However, not all pedigrees provide sufficient information to distinguish XD and AD. One definitive indication that a trait is inherited as AD, and not XD, is that an affected father passes the disease to a son; this type of transmission is not possible with XD, since males inherit their X chromosome from their mothers.

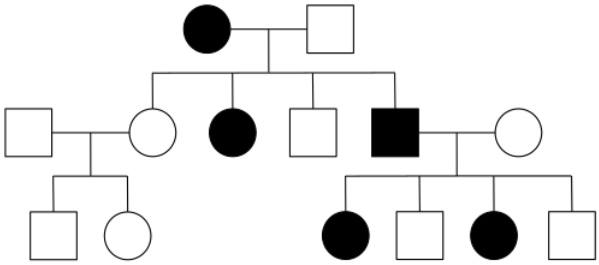


Figure 4.3.2 A Pedigree Consistent with X-Linked Dominant Inheritance [Long description]

Example: fragile x syndrome — The *FMR1* gene at Xq21 (X chromosome, q arm, region 2, band 1) encodes a protein needed for neuron development. There is a (CGG)n repeat array in the 5'UTR (untranslated region). If there is expansion of the repeat in the germline cell the child will inherit a non-functional allele. X^A / Y males have fragile X mental retardation (IQ < 50) because none of their neurons can make FMR1 proteins. Fragile X syndrome is considered X-linked dominant because only some neurons in X^A / X^a females can make FMR1 proteins. The severity (IQ 50 – 70) in these females depends upon the number and location of these cells within in the brain.

Autosomal Recessive (AR)

Diseases that are inherited in an autosomal recessive pattern require that both parents of an affected individual carry at least one copy of the disease allele. With AR traits, many individuals in a pedigree can be carriers, probably without knowing it. Compared to pedigrees of dominant traits, AR pedigrees tend to show fewer affected individuals and are more likely than AD or XD to "skip a generation". Thus, the major feature that distinguishes AR from AD or XD is that unaffected individuals can have affected offspring. Attached earlobes is a human condition that may follow an AR mode of inheritance.

Blue and red pedigree chart showing three generations and inheritance of an autosomal recessive trait

Figure 4.3.3 A Pedigree Consistent with Autosomal Recessive Inheritance [Long description]

AR example: phenylketonuria (PKU) – Individuals with phenylketonuria (PKU) have a mutation in the *PAH* gene at 12q24 (chromosome 12, q arm, region 2, band 4), which encodes an enzyme that breaks down phenylalanine into tyrosine called phenylalanine hydrolase (PAH). Without PAH, the accumulation of phenylalanine and other metabolites, such as phenylpyruvic acid, disrupts brain development, typically within a year after birth, and can lead to intellectual disability. Fortunately, this condition is both easy to diagnose and can be successfully treated with a low phenylalanine diet. There are over 450 different

mutant alleles of the PAH gene, so most people with PKU are **compound heterozygotes** . Compound heterozygotes have two different mutant alleles (different base pair changes) at a given locus, in this case the *PAH* gene.

X-Linked Recessive (XR)

Because males have only one X-chromosome, any male that inherits an X-linked recessive disease allele will be affected by it (assuming complete penetrance). Therefore, in XR modes of inheritance, males tend to be affected more frequently than females in a population. This contrasts with AR and AD, where both sexes tend to be affected equally, and XD, in which females are affected more frequently. Note, however, in the small sample sizes typical of human families, it is usually not possible to accurately determine whether one sex is affected more frequently than others. On the other hand, one feature of a pedigree that can be used to definitively establish that an inheritance pattern is not XR is the presence of an affected daughter from unaffected parents; because she would have had to inherit one X-chromosome from her father, he would also have been affected in XR.

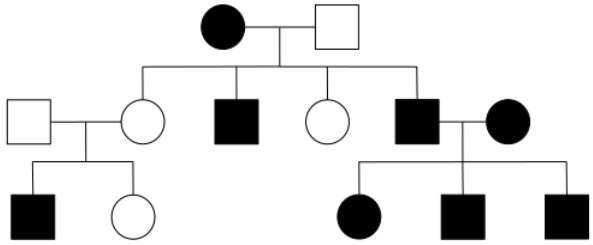


Figure 4.3.4 A Pedigree Consistent with X-Linked Recessive Inheritance [Long description]

XR example: hemophilia A- *F8* gene at Xq28 (X chromosome, q arm, region 2, band 8) encodes blood clotting factor VIIIc. Without Factor VIIIc, internal and external bleeding can't be stopped. Back in the 1900s, Xa / Y male's average life expectancy was 1.4 years, but in the 2000s it has increased to 65 years with the advent of Recombinant Human Factor VIIIc. Hemophilia A is recessive because XA / Xa females have normal blood coagulation, while Xa / Xa females have hemophilia.

Y-Linked

Only males are affected in human Y-linked inheritance (and other species with the X/Y sex determining system). There is only father-to-son transmission. This is the easiest mode of inheritance to identify, but it is one of the rarest because there are so few genes located only on the Y-chromosome.

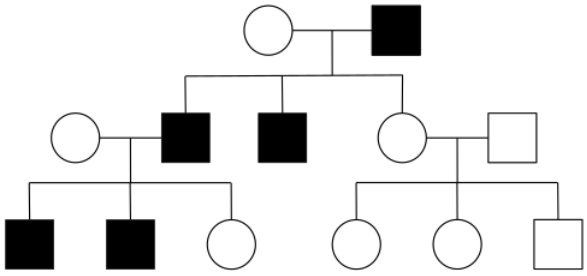


Figure 4.3.5 A Pedigree Consistent with Y-Linked Inheritance [Long description]

A common, but incorrect, example of Y-linked inheritance is the **hairy-ear-rim** phenotype seen in some Indian families. A better example are the **Y-chromosome DNA polymorphisms** that have been used to follow the male lineage in large families or through ancient ancestral lineages. For example, the Y-chromosome of Mongolian ruler Genghis Khan (1162-1227 CE), and his male relatives, accounts for ~8% of the Y-chromosome lineage of men in Asia (0.5% world wide).

Media Attributions

- **Figure 4.3.1** Autosomal dominant by Simon Caulton, CC BY-SA 3.0, via Wikimedia Commons
- **Figure 4.3.2** Wiki Drawing – X-Linked Dominant (1) by Madibc68, CC BY-SA 4.0, via Wikimedia Commons
- **Figure 4.3.3** Autosomal Recessive Pedigree Chart by Jerome Walker, CC BY-SA 3.0, via Wikimedia Commons
- **Figure 4.3.4** Wiki Drawing – X-Linked Recessive (1) by Madibc68, CC BY-SA 4.0, via Wikimedia Commons

- **Figure 4.3.5** Wiki Drawing – Y-Linked (1) by Madibc68, CC BY-SA 4.0, via Wikimedia Commons

Reference

AK Lecture Series. (2015, January 13). Pedigree Analysis (video file). YouTube. https://youtu.be/Wgmgt_Ph6Ko

Long Descriptions

- **Figure 4.3.1** Pedigree chart showing inheritance of an autosomal dominant trait over four generations. Affected females are shown as red-coloured circles; normal females are blue-coloured circles; affected males are red-coloured squares; normal males are blue-coloured squares. Generation I begins with a normal male and an affected female, mating to produce five offspring. Their offspring are: two affected and three unaffected. Two of these Generation II individuals mate, and their progeny is shown, along with a final Generation IV, with the characteristic pattern for autosomal dominant traits depicted. [Back to Figure 4.3.1]
- **Figure 4.3.2** Pedigree chart showing the inheritance of a X-linked dominant trait over three generations. An affected female mates with a normal male in Generation I, to produce four offspring – one normal male, one affected male, one normal female and one affected female. The normal female mates with a normal male, and they produce two unaffected children. The affected male of Generation II mates with a normal female, and they produce four children – two affected females and two unaffected males. This pattern is characteristic for X-linked dominant inheritance. [Back to Figure 4.3.2]
- **Figure 4.3.3** Pedigree chart showing the inheritance of an autosomal recessive trait, over three generations. Red and blue colours depict differences in the male and female. Both colours mean heterozygous, solid red colour means homozygous recessive, and solid blue color means wild type or homozygous dominant. Generation I begins with two carrier parents, who are therefore heterozygous. They produce four children, two are normal, one is a carrier and one is affected. This pattern is consistent with the inheritance of autosomal recessive traits. [Back to Figure 4.3.3]
- **Figure 4.3.4** Pedigree chart showing the inheritance of a X-linked recessive trait over three generations. Generation I outlines one affected female and one affected male, mating to produce four children, two unaffected females and two affected males. One of the unaffected females mates with a normal man, and they produce two children: one affected male and an unaffected female. This indicates the unaffected mother was a carrier. This pattern of inheritance is typical of an X-linked recessive trait. [Back to Figure 4.3.4]
- **Figure 4.3.5** Pedigree chart showing the inheritance of a Y-linked trait. Three generations are shown, starting with a normal female and an affected male, who produce three offspring. Offspring are: one normal female and two affected males. One affected male mates with a normal female, and their offspring comprises one normal female and two affected males. This pattern of inheritance is typical of Y-linked traits. [Back to Figure 4.3.5]

This page titled 4.8.2: Modes of Inheritance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.3: Modes of Inheritance** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

4.8.3: Sporadic and Non-Heritable Diseases

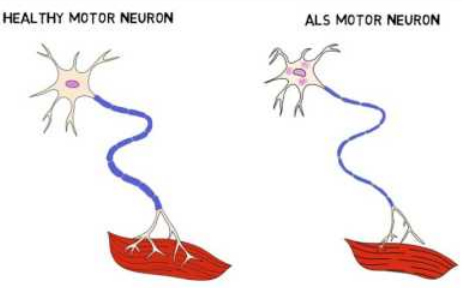
4.4 Sporadic and Non-Heritable Diseases

Not all the characterized human traits and diseases are attributed to mutant alleles at a single gene locus. Many diseases that have a heritable component, have more complex inheritance patterns due to (1) the involvement of multiple genes, and/or (2) environmental factors.

On the other hand, some non-genetic diseases may appear to be heritable because they affect multiple members of the same family, but this is due to the family members being exposed to the same toxins or other environmental factors (e.g., in their homes).

Finally, diseases with similar symptoms may have different causes, some of which may be genetic while others are not. One example of this is Amyotrophic lateral sclerosis (ALS); approximately 5–10% of cases are inherited in an AD pattern, while most of the remaining cases appear to be **sporadic**, in other words, not caused by a mutation inherited from a parent. We now know that different genes or proteins are affected in the inherited and sporadic forms of ALS. The physicist Stephen Hawking and baseball player Lou Gehrig both suffered from sporadic ALS.

Take a look at the following *2-Minute Neuroscience: Amyotrophic Lateral Sclerosis (ALS)* video, by Neuroscientifically Challenged (2017), which describes how ALS arises in humans.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=569

Reference

Mayo Clinic Staff (n.d.). Amyotrophic lateral sclerosis (ALS). MayoClinic.org. www.mayoclinic.org/diseases-conditions/amyotrophic-lateral-sclerosis/symptoms-causes/syc-20354022

Neuroscientifically Challenged. (2017). 2-minute neuroscience: Amyotrophic lateral sclerosis (ALS) [Video file]. YouTube. https://www.youtube.com/watch?v=kOnk9Hh20eg

This page titled 4.8.3: Sporadic and Non-Heritable Diseases is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.4: Sporadic and Non-Heritable Diseases** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

4.8.4: Calculating Probabilities Using Pedigree Charts

4.5 Calculating Probabilities Using Pedigree Charts

Once the mode of inheritance of a disease or trait is identified, some inferences about the genotype of individuals in a pedigree can be made, based on their phenotypes and where they appear in the family tree. Given these genotypes, it is possible to calculate the probability of a particular genotype being inherited in subsequent generations. This can be useful in genetic counselling, for example when prospective parents wish to know the likelihood of their offspring inheriting a disease for which they have a family history.

Probabilities in pedigrees are calculated using knowledge of Mendelian inheritance and the same basic methods as are used in other fields.

The first formula is the **product rule**: the joint probability of two independent events is the product of their individual probabilities; this is the probability of one event AND another event occurring.

For example:

The probability of a rolling a “five” with a single throw of a single six-sided die is , and the probability of rolling “five” in each of three successive rolls is .

The second useful formula is the **sum rule**, which states that the combined probability of two independent events is the sum of their individual probabilities. This is the probability of one event OR another event occurring.

For example:

The probability of rolling a five or six in a single throw of a dice is .

With these rules in mind, we can calculate the probability that two carriers (i.e., heterozygotes) of an AR disease will have a child affected with the disease as , since for each parent, the probability of any gametes carrying the disease allele is ½. This is consistent with what we already know from calculating probabilities using a Punnett square (e.g., in a monohybrid cross *Aa* × *Aa*, ¼ of the offspring are *aa*).

We can likewise calculate probabilities in the more complex pedigree shown in **Figure 4.5.1**.

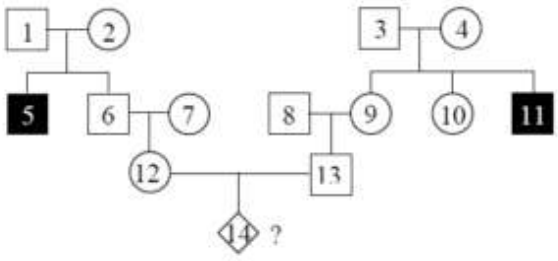


Figure 4.5.1 Individuals in this Pedigree are Labelled with Numbers to Make Discussion Easier [Long description]

Assuming the disease has an AR pattern of inheritance, what is the probability that individual #14 will be affected? We can assume that individuals #1, #2, #3 and #4 are heterozygotes (*Aa*), because they each had at least one affected (*aa*) child, but they are not affected themselves. This means that there is a chance that individual #6 is also *Aa* . This is because according to Mendelian inheritance, when two heterozygotes mate, there is a 1:2:1 distribution of genotypes *AA* : *Aa* : *aa* . However, because #6 is unaffected, he can’t be *aa* , so he is either *Aa* or *AA* , but the probability of him being *Aa* is twice as likely as *AA* . By the same reasoning, there is likewise a chance that #9 is a heterozygous carrier of the disease allele.

If individual #6 is a heterozygous for the disease allele, then there is a ½ chance that #12 will also be a heterozygote (i.e., if the mating of #6 and #7 is *Aa* × *AA*, half of the progeny will be *Aa*; we are also assuming that #7, who is unrelated, does not carry any disease alleles). Therefore, the combined probability that #12 is also a heterozygote is . This reasoning also applies to individual #13, i.e., there is a probability that he is a heterozygote for the disease. Thus, the overall probability that both individual #12 and #13 are heterozygous, and that a particular offspring of theirs will be homozygous for the disease alleles is .

Media Attribution

- **Figure 4.5.1**Original attributed to Unknown (2017), CC BY-NC 3.0, Open Genetics Lectures

Reference

Unknown. (2017). Figure 15. Individuals in this pedigree are labeled with…[digital image]. In Locke, J., Harrington, M., Canham, L. and Min Ku Kang (Eds.), *Open Genetics Lectures, Fall 2017* (Chapter 23, p. 7). Dataverse/ BCcampus. http://solr.bccampus.ca:8001/bcc/file/7a7b00f9-fb56-4c49-81a9-cfa3ad80e6d8/1/OpenGeneticsLectures_Fall2017.pdf

Long Description

- **Figure 4.5.1** Complex pedigree chart showing four generations. The purpose of this chart is simply to demonstrate that we are able to use the rules of probability to calculate the percentage chance of inheritance of a particular trait within a given family tree. The individuals in the chart are labelled from 1 to 14, throughout successive generations. All are shown to be unaffected, except for individual 5, who is part of Generation II and individual 11, also part of generation II. The text directly proceeding this image describes in detail, how this chart can be interpreted using the rules of probability. [Back to Figure 4.5.1]

This page titled 4.8.4: Calculating Probabilities Using Pedigree Charts is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

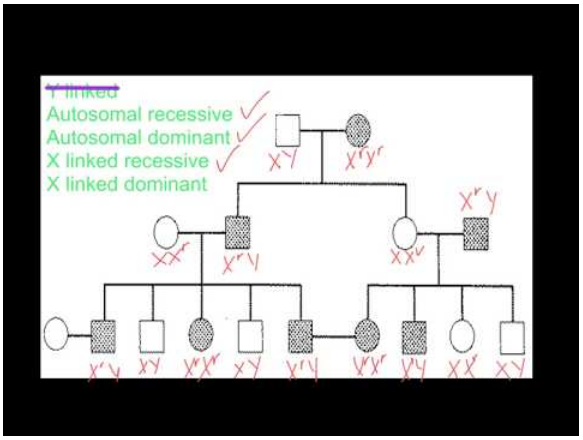
4.8.5: Chapter 4 Summary

Chapter 4 Summary

The topics covered in this chapter can be summarized as follows:

- Pedigree analysis can be used to determine the mode of inheritance of specific traits such as diseases.
- Loci can be X- or Y-linked or autosomal in location and alleles either dominant or recessive with respect to wild type.
- If the mode of inheritance is known, a pedigree can be used to calculate the probability of inheritance of a particular genotype by an individual.

Take a look at the video below, *Pedigree Analysis*, by Nicole Lantz (2016) on YouTube, which summarizes the various pedigree patters observed based on the different modes of inheritance.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=577

Reference

Nicole Lance. (2016). Pedigree analysis [Video file]. YouTube. https://www.youtube.com/watch?v=6VGcidNwQEo

This page titled 4.8.5: Chapter 4 Summary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.6: Chapter 4 Summary** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

4.8.6: Key Terms in Chapter 4

Key Terms in Chapter 4

Key Terms – Pedigree Analysis

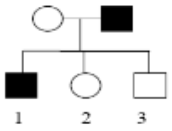
- Achondroplasia
- affected
- autosomal dominant
- autosomal recessive
- carrier
- genetic counselling
- hairy-ear-rim
- Hemophilia A
- incomplete penetrance
- mode of inheritance
- Pedigree charts
- (PKU)
- proband
- product rule
- sporadic
- sum rule
- variable expressivity
- X-linked dominant
- X-linked recessive
- X-linked recessive
- Y-linked

This page titled 4.8.6: Key Terms in Chapter 4 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.7: Key Terms in Chapter 4** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

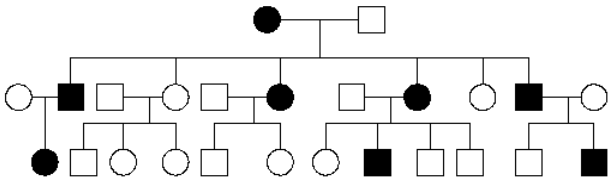
4.8.7: Chapter 4 Study Questions

Chapter 4 Study Questions

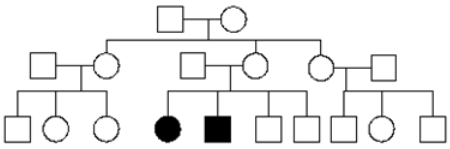


1. What are some of the modes of inheritance that are consistent with this pedigree?
2. In the pedigree in question 1, the mode of inheritance cannot be determined unambiguously. What are some examples of data (e.g., from other generations) that, if added to the pedigree, would help determine the mode of inheritance?
3. For each of the following pedigrees, name the most likely mode of inheritance (AR=autosomal recessive, AD=autosomal dominant, XR=X-linked recessive, XD=X-linked dominant). *These pedigrees were obtained from various external sources.

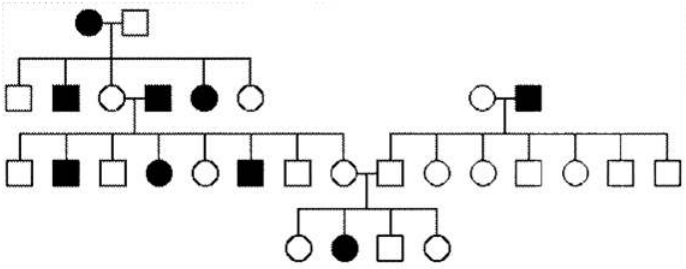
a.



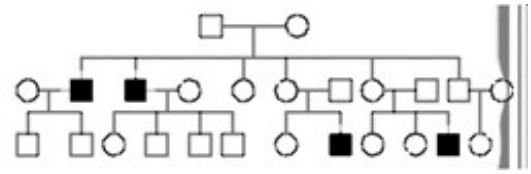
b.



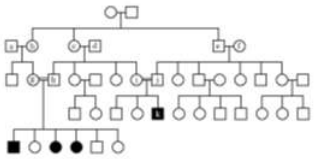
c.



d.

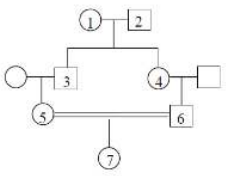


4. The following pedigree represents a rare, autosomal recessive disease. What are the genotypes of the individuals who are



indicated by letters?

5. If individual #1 in the following pedigree is a heterozygote for a rare, AR disease, what is the probability that individual #7 will



be affected by the disease? Assume that #2 and the spouses of #3 and #4 are not carriers.

This page titled 4.8.7: Chapter 4 Study Questions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.8: Chapter 4 Study Questions** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

4.8.8: Chapter 4 Answers

Chapter 4 Answers

1. The pedigree could show an AD, AR or XR mode of inheritance. It is most likely AD. It could be AR if the mother was a carrier, and the father was a homozygote. It could be XR if the mother was a carrier, and the father was a hemizygote. It cannot be XD, since the daughter (#2) would have necessarily inherited the disease allele on the X chromosome she received from her father.
2. There are many possible answers. Here are some possibilities: If neither of the parents of the father were affected (i.e., the paternal grandparents of children 1, 2, 3), then the disease could not be dominant. If only the paternal grandfather was affected, then the disease could only be X-linked recessive if the paternal grandmother was a heterozygote (which would be unlikely given that this is a rare disease allele).
3. a. The mode of inheritance is most likely AD, since every affected individual has an affected parent, and the disease is inherited even in four different matings to unrelated, unaffected individuals. It is very unlikely that it is XD or XR, in part because affected father had an affected son.
   b. The mode of inheritance cannot be AD or XD, because affected individuals must have an affected parent when a disease allele is dominant. Neither can it be XR, because there is an affected daughter of a normal father. Therefore, it must be AR, and this is consistent with the pedigree.
   c. The mode of inheritance cannot be AD or XD, because, again, there are affected individuals with unaffected parents. It is not XR, because there are unaffected sons of an affected mother. It is therefore, likely, AR — but note that the recessive alleles for this condition appear to be relatively common in the population (note that two of the marriages were to unrelated, affected individuals).
   d. The mode of inheritance cannot be AD or XD, because, again, there are affected individuals with unaffected parents. It could be either XR or AR, but because all the affected individuals are male, and no affected males pass the disease to their sons, it is likely XR.
4. If a represents the disease allele, individuals a, d, f (who all married into this unusual family) are AA, while b, c, e, g, h, i, j are all Aa, and k is aa.
5. There is a ½ chance that an offspring of any mating Aa x AA, will be a carrier (Aa). So, there is a ½ chance that #3 will be Aa, and likewise for #4. If #3 is a carrier, there is again a ½ chance that #5 will be a carrier, and likewise for #6. If #5 and #6 are both Aa, then there is a ¼ chance that this monohybrid cross will result in #7 having the genotype aa and, therefore, affected by the disease. Thus, the joint probability is .

This page titled 4.8.8: Chapter 4 Answers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

4.9: Pedigrees and Punnett Squares

Pedigrees

Inheritance of a trait through generations can be shown visually using a pedigree, such as is pictured in **Figure** 4.9.1. Square shapes represent males; circles represent females. Filled-in shapes are individuals that have whatever trait is being shown in the pedigree. Two individuals connected together with a horizontal line between them are the parents of the individuals that are connected by vertical lines below them. Siblings are typically shown in birth order with the oldest sibling to the left.

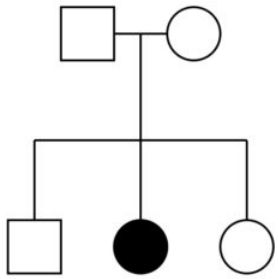


Figure 4.9.1**:** A simple pedigree. In this pedigree, the parents (at the top) have produced three children: a male and two females. The first female has the condition being shown in the pedigree.

Punnett Squares

As discussed above, diploid individuals have two copies of each chromosome: one from their male parent, one from their female parent. This means they have two copies of each gene. They can have two of the same alleles (homozygous) or two different alleles (heterozygous). Regardless of their genotype, they will randomly pass only one copy of each chromosome to their offspring. This is because meiosis produces haploid gametes that contain one copy of each chromosome, and those chromosomes are assorted into gametes randomly. Since genes are present on chromosomes, this means they will pass one copy of each gene to their offspring. That means that an offspring inherits one allele of each gene from each of its two parents. This is illustrated in Figure 4.9.2. This concept is called **Mendel's Law of Segregation**.

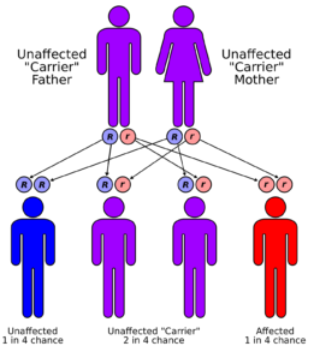


Figure 4.9.2**:** Two parents who are heterozygous each pass one chromosome / gene / allele to each offspring. Each resulting offspring has two of each chromosome / gene. The individual can have two of the same or two different alleles. (CC BY-SA 3.0; Cburnett via Wikimedia Commons)

An easy, organized way of illustrating the offspring that can result from two specific parents is to use a Punnett square. The gametes that can be generated by each parent are represented above the rows and next to the columns of the square. Each gamete is haploid for the "A gene", meaning it only contains one copy of that gene. In the Punnett square seen in Figure 4.9.3, haploid eggs are above each column and haploid sperm are next to each row. When a haploid sperm and a haploid egg (each with 1 copy of the "A gene") combine during the process of fertilization, a diploid offspring (with 2 copies of the A gene) is the result.

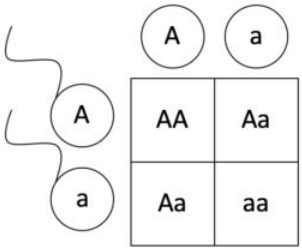


Figure 4.9.3**:** A Punnett square showing a cross between two individuals who are both heterozygous for A.

A Punnett square shows the probability of an offspring with a given genotype resulting from a cross. It does not show actual offspring. For example, the Punnett square in Figure 4.9.3 shows that there is a 25% chance that a homozygous recessive offspring will result from the cross Aa x Aa. It does *not* mean that these parents must have 4 offspring and that they will have the ratio 1 AA : 2 Aa : 1 aa. It's just like flipping a coin: you expect 50% heads, but you wouldn't be too surprised to see 7 heads out of 10 coin flips. Additionally, the probability does not change for successive offspring. The probability that the first offspring will have the genotype "aa" is 25% and the probability of the second offspring having the genotype "aa" is still 25%. Again, it's just like flipping a coin: if you flip heads the first time, that doesn't change the probability of getting heads on the next flip.

Organisms don't just inherit one trait at a time, though. They inherit all their traits at once. Sometimes, we want to determine the probability of an individual inheriting two different traits. The easiest way to do this is to determine the probability of the individual inheriting each trait separately, then multiply those probabilities together. An example of this can be seen in **Figure** 4.9.4. In order for this to work, we must assume that genes do not influence each other with regard to the sorting of alleles into gametes, and every possible combination of alleles for every gene is equally likely to occur. This is called **Mendel's Law of Independent Assortment**.

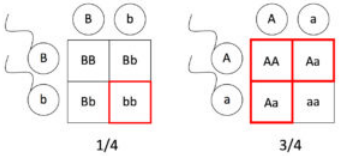


Figure 4.9.4**:** These two Punnett squares show the cross between two individuals who are both heterozygous for two different genes: BbAa x BbAa. We can determine the probability of an offspring having the recessive trait for "B" and the dominant trait for "A". The probability of the offspring having the recessive phenotype for "B" is 1/4. The probability of the offspring having the dominant phenotype for "A" is 3/4. 1/4 x 3/4 = 3/16.

Another way of determining the probability of getting two different traits is to use a dihybrid Punnett square. **Figure** 4.9.5 shows three generations of the inheritance of pea seed color and shape. Peas can be either yellow or green, and they can be either round or wrinkled. These are two of the traits that Mendel studied in his work with peas. In the first generation (the "P" generation), two true-breeding (homozygous) individuals are crossed. Their offspring will get one allele of the *Y* gene and one allele of the *R* gene from each parent. This means that all their offspring (the "F_1" generation) will be heterozygous for both genes. The results (the "F_2" generation) from crossing two heterozygous individuals can be seen in the 4×4 Punnett square in **Figure** 4.9.5.

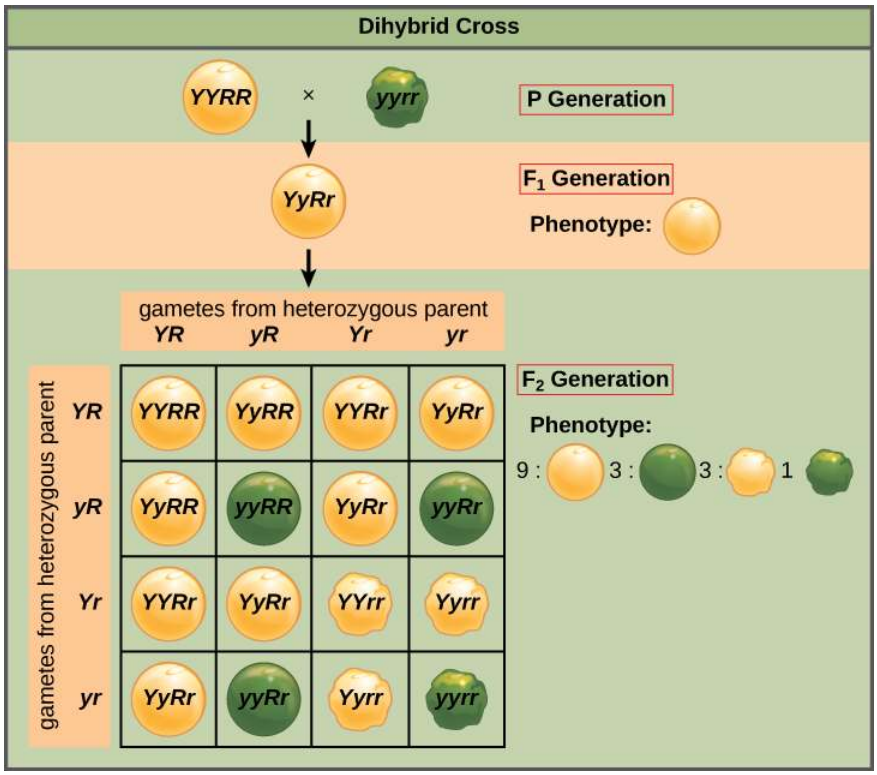


Figure 4.9.5**:** This dihybrid cross shows the expected offspring from the F_2 generation after crossing *YYRR* x *yyrr*. Compare the results from this Punnett square to the results seen in the previous figure. They match! Photo Credit: OpenStax Biology.

The gametes produced by the F_1 individuals must have one allele from each of the two genes. For example, a gamete could get an *R* allele for the seed shape gene and either a *Y* or a *y* allele for the seed color gene. It cannot get both an *R* and an *r* allele; each gamete can have only one allele per gene. The law of independent assortment states that a gamete into which an *r* allele is sorted would be equally likely to contain either a *Y* or a *y* allele. Thus, there are four equally likely gametes that can be formed when the *RrYy* heterozygote is self-crossed, as follows: *RY*, *rY*, *Ry*, and *ry*. Arranging these gametes along the top and left of a 4 × 4 Punnett square (**Figure** 4.9.5) gives us 16 equally likely genotypic combinations. From these genotypes, we find a phenotypic ratio of 9 round–yellow:3 round–green:3 wrinkled–yellow:1 wrinkled–green (**Figure** 4.9.5). These are the offspring ratios we would expect, assuming we performed the crosses with a large enough sample size.

We can look for individuals who have the recessive phenotype for *Y* and the dominant phenotype for *R*. These individuals must have two little *y*'s and at least one big *R*. The possible genotypes are *yyRR* or *yyRr*. Examining the Punnett square in **Figure** 4.9.5, we can find 3 individuals with these genotypes (they are round and green). If you compare the results from **Figure** 4.9.4 and **Figure** 4.9.5, you'll see that we have arrived at the same value: 3/16!

References

Unless otherwise noted, text and images by Lisa Bartee, 2016.

This page titled 4.9: Pedigrees and Punnett Squares is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

4.10: Biochemical Basis of Dominance

Given that a heterozygote's phenotype cannot simply be predicted from the phenotype of homozygotes, what does the type of dominance tell us about the biochemical nature of the gene product? How does dominance work at the biochemical level? There are several different biochemical mechanisms that may make one allele dominant to another.

For the majority of genes studied, the normal (i.e. wild-type) alleles are **haplosufficient**. So in diploids, even with a mutation that causes a complete loss of function in one allele, the other allele, a wild-type allele, will provide sufficient normal biochemical activity to yield a wild type phenotype and thus be dominant and dictate the heterozygote phenotype.

On the other hand, in some biochemical pathways, a single wild-type allele is not enough protein and may be **haploinsufficient** to produce enough biochemical activity to result in a normal phenotype, when heterozygous with a non-functioning mutant allele. In this case, the non-functional mutant allele will be dominant (or semi-dominant) to a wild-type allele.

Mutant alleles may also encode products that have new and/or different biochemical activities instead of, or in addition to, the normal ones. These **novel activities** could cause a new phenotype that would be dominantly expressed.

This page titled 4.10: Biochemical Basis of Dominance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.3: Biochemical Basis of Dominance** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.11: Black fur color - a dominant trait



Figure 4.11.1**:** This chocolate lab has two recessive alleles of the TYRP1 gene. (Credit: Rob Hanson; photo from Wikimedia.)

Most of us are familiar with the labrador retriever dog breed, such as the chocolate lab seen in **Figure** 4.11.1. But have you ever thought about what makes this dog brown? The difference between brown and black coat color in dogs is caused by a mutation in the TYRP1 gene. The TYRP1 gene provides instructions for making an enzyme called tyrosinase-related protein 1. This enzyme is required to produce a pigment called eumelanin. Eumelanin is a dark colored pigment. The TYRP1 gene is located on chromosome 11 in dogs (Parker, 2001).

A group of scientists who were interested in determining what caused the difference between black and brown coats sequenced the DNA within the protein-coding region of the TYRP1 gene (Schmutz, 2002). They identified three variations in the DNA making up the TYRP1 gene between brown dogs and black dogs. These variations in DNA sequence are examples of different **alleles** of the TYRP1 gene.

Table 4.11.1**: Variations in the TYRP1 allele that lead to brown color in dogs. Data from Schmutz, 2002.**

Location	Black DNA sequence	Brown DNA sequence	Effect on protein
exon 2	TGT	CGT	changes a cysteine amino acid to a serine
exon 5	CAG	TAG	introduces a premature stop codon which results in 330 amino acids instead of 512 amino acids in the protein
exon 5	CCT	— (deleted)	deletion of a proline amino acid

All of these variations in the DNA sequence are predicted to cause a change in the amino acid sequence of the TYRP1 protein. These changes affect the production of eumelanin pigment, which is black in color. When eumelanin is not being produced correctly, the dog appears brown instead of black.

Like other diploid organisms, dogs all have two copies of the TYRP1 gene (one from their male parent, one from their female parent). Dogs that are homozygous for the black allele (dogs that have two copies of the black allele) are obviously going to be black in color. Dogs that are homozygous for the brown allele are obviously going to be brown. Dogs that are heterozygous (dogs that have one black allele and one brown allele) appear black. The black and brown colors do not blend together: the black allele covers up the brown allele. This means that the black allele is **dominant** over the brown allele. Remember that dominant alleles cover up **recessive** alleles. If there is one dominant allele present, the dog will appear black. The brown allele is recessive to the black allele. There must be two copies of the recessive brown allele present in order for the dog to appear brown.



Figure 4.11.2**:** Black and brown phenotypes in labrador retrievers. (Credit: demealiffe; from Wikimedia)

Remember that genotypes can be abbreviated with a single letter and that the letter which is chosen is typically the first letter of the dominant trait. In this case, the letter “B” is used to represent the dominant black allele, while “b” represents a recessive brown allele.

The reason that the black allele is dominant over the brown allele in this specific situation is because the black allele produces functional TYRP1 protein, while the brown allele does not. The presence of one functional allele produces enough TYRP1 protein allows the cells to produce eumelanin and appear black.

Remember: dominant does not mean “better” or “more normal”. Black color does not confer any special advantages on dogs compared to brown color. It’s just a difference.



Figure 4.11.3**:** What alleles of TYRP1 does this black lab puppy have? We can’t tell by looking at it. The puppy could be homozygous (BB) or heterozygous (Bb). Since black is completely dominant over brown, both options would be black. (Credit: Alice Birkin)

Let’s visualize the inheritance of black and brown using a pedigree. The pedigree in **Figure** 4.11.4 shows a litter of puppies. The shaded symbol shows a brown puppy, while open symbols are black individuals.

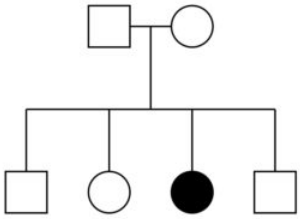


Figure 4.11.4**:** An example litter of puppies. The filled-in symbol shows a brown individual.

To interpret this pedigree, let’s start with information that we already know:

- Brown is recessive, which means brown individuals must have the genotype bb. In this pedigree, brown individuals are filled in.
- Black is dominant, which means black individuals must have at least one B allele. Their genotype could be either BB or Bb. In this pedigree, black individuals are not filled in.

Figure 4.11.5 shows the same pedigree, but with information about the individual’s genotypes filled in.

1. The shaded individual, who is a brown female puppy, must have the genotype bb. If she had any B alleles, she would be black because the black allele is dominant over the brown allele.
2. In order for the brown puppy to have the genotype bb, she must have gotten two “b” alleles: one from each of her parents. We know that her parents are both black (because they are unshaded), which means they must have a least one “B” allele. This means that both parents must be heterozygous: Bb.
3. The three black puppies must have at least one “B” allele in order for them to be black in color. However, we can’t tell whether they are homozygous dominant (BB) or heterozygous (Bb) since both of those genotypes would result in black color. One way to represent this on a pedigree is B-, meaning that the second allele could be either B or b.

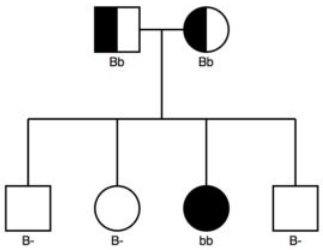


Figure 4.11.5**:** Genotypes of the individuals in this pedigree.

We can also show the cross between these parents as a Punnett square (**Figure** 4.11.6). We would expect 1/4 of the offspring to have the genotype bb, and that is what we see in the pedigree above.

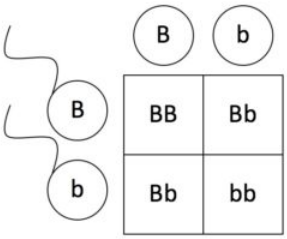


Figure 4.11.6**:** The information from the pedigree shown in Figure 4.11.5 can also be shown as a Punnett square.

Human Connection

A small number of mutations in the TYRP1 gene have been found to cause oculocutaneous albinism type 3. This condition includes a form of albinism called rufous oculocutaneous albinism, which has been described primarily in dark-skinned people from southern Africa. Affected individuals have reddish-brown skin, ginger or red hair, and hazel or brown irises. Two TYRP1 mutations are known to cause this form of albinism in individuals from Africa. One mutation replaces a protein building block (amino acid) in tyrosine-related protein 1 with a signal that prematurely stops protein production. This mutation, written as Ser166Ter or S166X, affects the amino acid serine at protein position 166. The other mutation, written as 368delA, deletes a single DNA building block from the TYRP1gene. Other alterations in this gene have been reported in a few affected people of non-African heritage. Most TYRP1 mutations lead to the production of an abnormally short, nonfunctional version of tyrosinase-related protein 1. Because this enzyme plays a role in normal pigmentation, its loss leads to the changes in skin, hair, and eye coloration that are characteristic of oculocutaneous albinism.



Photo credit: Muntuwandi; from Wikipedia.

References

Unless otherwise noted, text and images by Lisa Bartee, 2016.

Parker HG, Yuhua X, Mellersh CS, Khan S, Shibuya H, Johnson GS, Ostrander EA. Sept 2001. Meiotic linkage mapping of 52 genes onto the canine map does not identify significant levels of microrearrangement. Mamm Genome. 12(9):713-8.

Schmutz SM, Berryere TG, Goldfinch AD. 2002. TYRP1 and MC1R genotypes and their effects on coat color in dogs. Mammalian Genome 13, 380-387.

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.1...ls-Experiments

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.1...of-Inheritance

Information about TYRP1 and oculocutaneous albinism type 3: "Tyrp1" by Genetics Home Reference: Your Guide to Understanding Genetic Conditions, National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

This page titled 4.11: Black fur color - a dominant trait is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

4.12: Yellow fur color - a recessive trait

Yellow color in dogs

Labrador retrievers don't only come in brown and black, they also come in yellow. Yellow color in labs is caused by variations in a different gene: MC1R. This gene controls the production of the melanocortin 1 receptor protein. MC1R is located on chromosome 5 in dogs (Schmutz, 2001).



Figure 4.12.1**:** This yellow lab is producing light-colored pheomelanin instead of dark-colored eumelanin. (Credit: Djmirko; from Wikimedia)

Melanocytes make two forms of melanin, eumelanin and pheomelanin. The relative amounts of these two pigments help determine the color of an individual's hair and skin. Individuals who produce mostly eumelanin tend to have brown or black hair and dark skin that tans easily (in humans). Eumelanin also protects skin from damage caused by ultraviolet (UV) radiation in sunlight. Individuals who produce mostly pheomelanin tend to have red or blond hair, freckles, and light-colored skin that tans poorly. Because pheomelanin does not protect skin from UV radiation, people with more pheomelanin have an increased risk of skin damage caused by sun exposure.

The melanocortin 1 receptor controls which type of melanin is produced by melanocytes. When the receptor is activated, it triggers a series of chemical reactions inside melanocytes that stimulate these cells to make eumelanin. If the receptor is not activated or is blocked, melanocytes make pheomelanin instead of eumelanin. This means that if the receptor is working correctly and is turned on, dark pigment will be produced. If the receptor is not functional or is not turned on, light pigment will be produced.

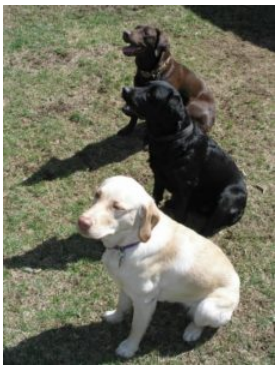


Figure 4.12.2**:** The three recognized colors of labs are due to black eumelanin, brown eumelanin, or pheomelanin. (Credit: Erikeltic, from Wikimedia)

Schmutz et. al. (2002) determined the DNA sequence for the MC1R gene from dogs of various colors. They determined that black and brown dogs all have one allele of MC1R, while yellow and red dogs have a different allele. The allele that leads to yellow or red color has a premature stop codon which results in a shorter-than-normal protein. This protein would be predicted to not function correctly. Remember that when the melanocortin 1 receptor is not functioning correctly, light pheomelanin pigment is produced and not dark eumelanin.

Dogs that are homozygous for the functioning allele of MC1R (which would cause eumelanin to be produced) are dark in color. Dogs that are homozygous for the non-functioning allele (which would cause pheomelanin to be produced) are light in color. Dogs that are heterozygous are dark in color. What does this tell you about which allele is dominant? If you said "the dark allele is dominant because it covers up the light allele", you're correct. We will use "E" to represent the genotype at MC1R because the

dominant phenotype in this case is the production of eumelanin. Dogs that have the genotype EE or Ee will produce eumelanin and be dark. Dogs that have the genotype “ee” will produce pheomelanin and be light.

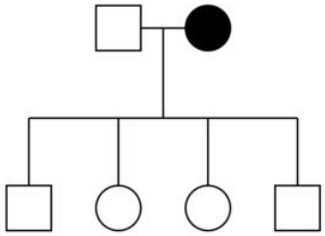


Figure 4.12.3**:** In this pedigree, the shaded individual is yellow. She therefore has the genotype ee and produces pheomelanin. We can’t tell the genotype of her mate by looking (he could be Ee or EE), but since all of their puppies were dark in color, we would predict that his genotype was EE. In this cross: EE x ee, 100% of the puppies would have the genotype Ee, so 100% of the puppies would produce eumelanin instead of pheomelanin.

The cross shown in **Figure** 4.12.3 can also be shown as a Punnett square (Figure 4.12.4). Since we are unsure whether the male dog has the genotype “EE” or “Ee”, we have to make two Punnett squares. Since all of the puppies resulting from this cross were black, we would predict that the first Punnett square shows the cross. However, it is possible that the second Punnett square is correct. There are only 4 puppies, so it’s not hard to imagine that they could all be black even though the Punnett square predicts only 50% black. It would be comparable to flipping a coin 4 times and getting 4 heads in a row. Getting 4 heads in a row is less likely, but definitely possible.

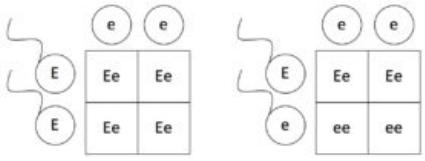


Figure 4.12.4**:** Cross from Figure 4.12.3 shown as Punnett squares.

It is very important to note here that yellow dogs still have the TYRP1 gene, even though they are not black or brown!

Human Connection

Common variations (polymorphisms) in the MC1R gene are associated with normal differences in skin and hair color. Certain genetic variations are most common in people with red hair, fair skin, freckles, and an increased sensitivity to sun exposure. These MC1R polymorphisms reduce the ability of the melanocortin 1 receptor to stimulate eumelanin production, causing melanocytes to make mostly pheomelanin. Although MC1R is a key gene in normal human pigmentation, researchers believe that the effects of other genes also contribute to a person’s hair and skin coloring.

The melanocortin 1 receptor is also active in cells other than melanocytes, including cells involved in the body’s immune and inflammatory responses. The receptor’s function in these cells is unknown.



Photo credit: dusdin on flickr; from Wikipedia.

Resources

Unless otherwise noted, text and images by Lisa Bartee, 2016.

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:FtsD6vMd@3/Mendels-Experiments

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:sbdXt0s3@4/Laws-of-Inheritance

OpenStax, Biology. OpenStax CNX. May 27, 2016 http://cnx.org/contents/s8Hh0oOc@9.10:zLLYW2hj@5/Extensions-of-the-Laws-of-Inhe

Human Connection – information about MC1R: "MC1R" by Genetics Home Reference: Your Guide to Understanding Genetic Conditions, National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

Schmutz SM, Moker JS, Berryere TG, Christison KM, Dolf G. 2001. An SNP is used to map MC1R to dog chromosome 5. Anim Genet. 32(1):43-4.

Schmutz SM, Berryere TG, Goldfinch AD. 2002. TYRP1 and MC1R genotypes and their effects on coat color in dogs. Mammalian Genome 13, 380-387.

This page titled 4.12: Yellow fur color - a recessive trait is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

4.13: Sex-Linkage- An Exception to Mendel's First Law

In the previous chapter we introduced sex chromosomes and autosomes. For loci on autosomes, the alleles follow the normal Mendelian pattern of inheritance. However, for loci on the sex chromosomes this is mostly not true, because most of the loci on the typical X-chromosome are absent from the Y-chromosome, even though they act as a homologous pair during meiosis. Instead, they will follow a **sex-linked** pattern of inheritance.

X-Linked Genes: the *white* gene in *Drosophila melanogaster*

A well-studied sex-linked gene is the *white* gene on the X chromosome of *Drosophila melanogaster*. Normally flies have red eyes but flies with a mutant allele of this gene called *white*$^-$ (w^-) have white eyes because the red pigments are absent. Because this mutation is recessive to the wild type w^+ allele females that are heterozygous have normal red eyes. Female flies that are homozygous for the mutant allele have white eyes. Because there is no *white* gene on the Y chromosome, male flies can only be hemizygous for the wild type allele or the mutant allele.

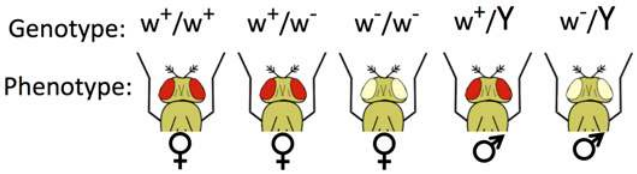


Figure 4.13.9: Relationship between genotype and phenotype for a the white gene on the X-linked gene in Drosophila melanogaster. The Y chromosome is indicated with a capital Y because it does not have a copy of the white gene. (Original-Harrington/Locke-CC:AN)

A researcher may not know beforehand whether a novel mutation is sex-linked. The definitive method to test for sex-linkage is **reciprocal crosses** (Figure 4.13.10). This means to cross a male and a female that have different phenotypes, and then conduct a second set of crosses, in which the phenotypes are reversed relative to the sex of the parents in the first cross. For example, if you were to set up reciprocal crosses with flies from pure-breeding w^+ and w^- strains the results would be as shown in Figure 4.13.10. Whenever reciprocal crosses give different results in the F1 and F2 and whenever the male and female offspring have different phenotypes the usual explanation is sex-linkage. Remember, if the locus were autosomal the F1 and F2 progeny would be different from either of these crosses.

A similar pattern of sex-linked inheritance is seen for X-chromosome loci in other species with an XX-XY sex chromosome system, including mammals and humans. The ZZ-ZW system is similar, but reversed (see below).

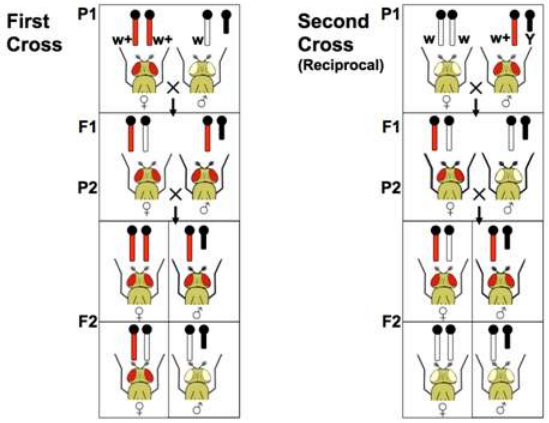


Figure 4.13.10**:** Reciprocal crosses involving an X-linked gene in Drosophila melanogaster. In the first cross (left) all of the offspring have red eyes. In second (reciprocal) cross (right) all of the female offspring have red eyes and the male offspring all have white eyes. If the F1 progeny are crossed (to make the P2), the F2 progeny will be different in each cross. The first cross has all red-eyed females and half red-eyed males. The reciprocal cross has half red-eyed males and females. Thomas Morgan won the Nobel Prize for using these crosses to demonstrate that genes (such as white) were on chromosomes (in this case the X-chromosome). (Wikipedia-PAR-PD)

Sex Determination in animals.

There are various mechanisms for sex determination in animals. These include sex chromosomes, chromosome dosage, and environment.

For example in humans and other mammals XY embryos develop as males while XX embryos become females. This difference in development is due to the presence of only a **single gene**, the TDF-Y gene, on the Y-chromosome. Its presence and expression dictates that the sex of the individual will be male. Its absence results in a female phenotype.

Although *Drosophila melanogaster* also has an XX-XY sex chromosomes, its sex determination system uses a different method, that of **X:Autosome (X:A) ratio**. In this system it is the ratio of autosome chromosome sets (A) relative to the number of X-chromosomes (X) that determines the sex. Individuals with two autosome sets and two X-chromosomes (2A:2X) will develop as females, while those with only one X-chromosome (2A:1X) will develop as males. The presence/absence of the Y-chromosome and its genes are not significant.

In other species of animals the number of chromosome sets can determine sex. For example the **haploid-diploid system** is used in bees, ants, and wasps. Typically haploids are male and diploids are female.

In other species, the environment can determine an individuals sex. In alligators (and some other reptiles) the temperature of development dictates the sex, while in many reef fish, the population sex ratio can cause some individuals to change sex.

Dosage Compensation for Loci on Sex Chromosomes.

Mammals and *Drosophila* both have XX - XY sex determination systems. However, because these systems evolved independently they work differently with regard to compensating for the difference in gene dosage (and sex determination – see above). Remember, in most cases the sex chromosomes act as a homologous pair even though the Y-chromosome has lost most of the loci when compared to the X-chromosome. Typically, the X and the Y chromosomes were once similar but, for unclear reasons, the Y chromosomes have degenerated, slowly mutating and loosing its loci. In modern day mammals the Y chromosomes have very few genes left while the X chromosomes remain as they were. This is a general feature of all organisms that use chromosome based sex determination systems. Chromosomes found in both sexes (the X or the Z) have retained their genes while the chromosome found in only one sex (the Y or the W) have lost most of their genes. In either case there is a gene dosage difference between the sexes: e.g. XX females have two doses of X-chromosome genes while XY males only have one. This gene dosage needs to be compensated in a process called **dosage compensation**. There are two major mechanisms.

In Drosophila and many other insects, to make up for the males only having a single X chromosome the genes on it are expressed at twice the normal rate. This mechanism of dosage compensation restores a balance between proteins encoded by **X-linked genes** and those made by **autosomal genes**.

In mammals a different mechanism is used, called **X-chromosome inactivation**.

X-chromosome Inactivation in Mammals

In mammals the dosage compensation system operates in females, not males. In XX embryos one X in each cell is randomly chosen and marked for inactivation. From this point forward this chromosome will be inactive, hence its name $X_{inactive}$ (X_i). The other X chromosome, the X_{active} (X_a), is unaffected. The X_i is replicated during S phase and transmitted during mitosis the same as any other chromosome but most of its genes are never allowed to turn on. The chromosome appears as a condensed mass within interphase nuclei called the Barr body. With the inactivation of genes on one X-chromosome, females have the same number of functioning X-linked genes as males.

This random inactivation of one X-chromosome leads to a commonly observe phenomenon in cats. A familiar X-linked gene is the *Orange* gene (*O*) in cats. The O^O allele encodes an enzyme that results in orange pigment for the hair. The O^B allele causes the hairs to be black. The phenotypes of various genotypes of cats are shown in Figure 4.13.11. Note that the heterozygous females have an orange and black mottled phenotype known as tortoiseshell. This is due to patches of skin cells having different X-chromosomes inactivated. In each orange hair the X_i chromosome carrying the O^B allele is inactivated. The O^O allele on the X_a is functional and orange pigments are made. In black hairs the reverse is true, the X_i chromosome with the O^O allele is inactive and the X_a chromosome with the O^B allele is active. Because the inactivation decision happens early during embryogenesis, the cells continue to divide to make large patches on the adult cat skin where one or the other X is inactivated.

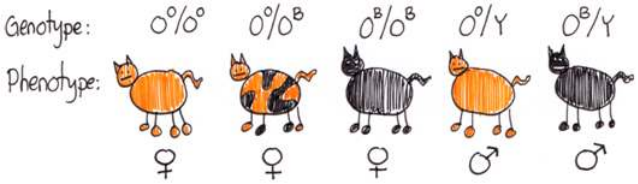


Figure 4.13.11: Relationship between genotype and phenotype for an X-linked gene in cats. The O^O allele = orange while the O^B allele = black. *(Original-Harrington-CC:AN)*

The *Orange* gene in cats is a good demonstration of how the mammalian dosage compensation system affects gene expression. However, most X-linked genes do not produce such dramatic mosaic phenotypes in heterozygous females. A more typical example is the *F8* gene in humans. It makes Factor VIII blood clotting proteins in liver cells. If a male is hemizygous for a mutant allele the result is hemophilia type A. Females homozygous for mutant alleles will also have hemophilia. Heterozygous females, those people who are $F8^+/F8^-$, do not have hemophilia because even though half of their liver cells do not make Factor VIII (because the X with the $F8^+$ allele is inactive) the other 50% can (Figure 4.13.12). Because some of their liver cells are exporting Factor VIII proteins into the blood stream they have the ability to form blood clots throughout their bodies. The genetic mosaicism in the cells of their bodies does not produce a visible mosaic phenotype.

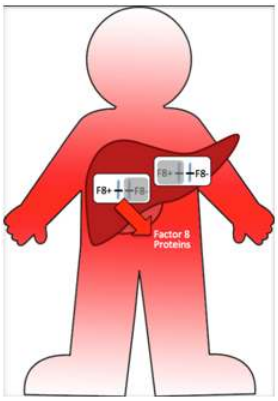


Figure 4.13.12: This figure shows the two types of liver cells in females heterozygous for an F8 mutation. Because people with the $F8^+/F8^-$ genotype have the same phenotype, normal blood clotting, as $F8^+/F8^+$ people the $F8^-$ mutation is classified as recessive. (Original-Harrington/Locke-CC:AN)

Other Sex-Linked Genes – Z-linked genes

One last example is a **Z-linked gene** that influences feather colour in turkeys. Turkeys are birds, which use the ZZ-ZW sex chromosome system. The *E* allele makes the feathers bronze and the *e* allele makes the feathers brown (Figure 4.13.13). Only male turkeys can be heterozygous for this locus, because they have two Z chromosomes. They are also uniformly bronze because the *E* allele is completely dominant to the *e* allele and birds use a dosage compensation system similar to *Drosophila* and not mammals. Reciprocal crosses between turkeys from pure-breeding bronze and brown breeds would reveal that this gene is in fact Z-linked.

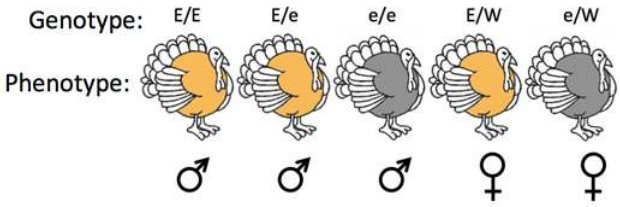


Figure 4.13.13: Relationship between genotype and phenotype for a Z-linked gene in turkeys. The W chromosome does not have an E/e-gene so it is just indicated with a capital W. (Original-Harrington/Locke-CC:AN)

Mechanisms of Sex Determination Systems

Sex is a phenotype. Typically, in most species, there are multiple characteristics, in addition to sex organs, that distinguish male from female individuals (although some species are normally **hermaphrodites** where both sex organs are present in the same

individual). The morphology and physiology of male and females is a phenotype just like hair or eye colour or wing shape. The sex of an organism is part of its phenotype and can be genetically (or environmentally) determined.

For each species, the genetic determination relies on one of several gene or chromosome based mechanisms. See Figure 4.13.14 for a summary. There are, for other species, also a variety of environmental mechanisms, too (rearing temperature, social interactions, **parthenogenesis**). Whatever the sex choice mechanism, however, there are two different means by which the cells of an organism carry out this decision: **hormonal** or **cell-autonomous**

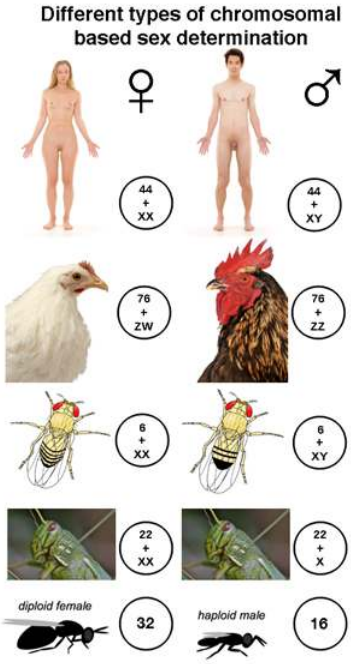


Figure 4.13.14: – Different types of chromosomal (or gene) based sex determination. From top to bottom, there is the archetypal XX/XY system found in humans (and most mammals) with the TDF-Y gene leading to a male phenotype; the ZW/ZZ system found in chickens (birds, moths, and butterflies); the same XX/XY system in Drosophila (sex is determined by the X-chromosome:autosome ratio); the XX/XO system as found in grasshoppers; and the diploid/haploid system as found in bees (and ants, and wasps). Also, the hormonal mechanism is used in humans, while all the other examples use the cell-autonomous mechanism for development of the male or female sex phenotype. (Wikipedia+original - CFCF with additions and corrections by J. Locke- CC:AN)

Hormonal mechanism: With this system, used by mammals for example, including humans, the zygote initially develops into a sexually undifferentiated embryo that can become either sex. Then, depending on the sex choice of the genital ridge cells, they will grow and differentiate into male (testis) or female (ovary) gonads, which will then produce the appropriate hormones (e.g. testosterone or estrogen). This hormone will circulate throughout the body and cause all the other tissues to develop and differentiate accordingly, into a male or female phenotype for that individual. Thus, the circulating hormone "tells" all the cells and tissues what sex to be and which sexual phenotype to be.

A **freemartin** is a type of **chimera** found in cattle (and some other mammals). Externally it appears as a female but is infertile, and has masculinized behavior and non-functioning ovaries. The animal originates as a female (XX), but acquires male (XY) cells or tissues *in utero* by exchange of some cellular material from a male twin. The female reproductive development is altered by anti-Müllerian hormone from the male twin, acquired via vascular connections between placentas.

Cell-autonomous mechanism: With this system, used by many animals, including birds and insects, the zygote cell initially has a sex phenotype set at the cell level. All cells intrinsically know, individually, which sex they are and develop accordingly, giving the appropriate sexual characteristics and phenotype. Each cell is autonomous with respect to its sex; there are no sex hormone cues to determine the sex expressed. This autonomy can lead to **sexual gynandromorphs**, which are **mosaics** that display both male and female characteristics in a mosaic fasion, typically split down the midline of the organism. These rare individuals are thought to be the result of an improper sex chromosome segregation that occurs in a cell very early in development so that one half of the individual has cells with a male chromosome set while the other half has cells with a female set. If the species is **sexually**

dimorphic (external morphology easily distinguishs males from females) they are easily visible and are even sometimes seen in the wild. See Figure 4.13.15 for a local example. A search on the internet will bring up many more examples.



Figure 4.13.15: – Drosophila sexual gynadromorph. The left side is female with the most distal abdominal segments being not heavily pigmented, while the right is male and the two most distal segments are heavily pigmeted (arrow). This example was found by a U. of Alberta student in the GENET 375 lab course, Introductions to Molecular Genetic Techniques. (Original – Locke – CC:AN)

While gynandromorphs are seen in cell-autonomous species, such as insects and birds, they are not seen in hormonally determined species, such as mammals, because all the cells display the same sex phenotype caused by the circulating sex hormones. Sexual gynandromorphs appear to be absent in reptiles, amphibians, and fish indicating that they don't use a cell-autonomous mechanism. Nevertheless, there are genetic mosaic individuals in these groups but they do not appear to involve sex determined traits, which is required for a true gynandromorph. They often involve mosaicism of alleles at a single gene locus that affect external morphology (e.g. color).

- A **gynandromorph** is an organism that made up of mosaic tissues of male and female genotypes and displays both male and female characteristics.
- A **mosaic** is an organism or a tissue that contains two or more types of genetically different cells derived from the same zygote.
- A **chimera** is a single organism composed of genetically distinct cells derived from different zygotes.

This page titled 4.13: Sex-Linkage- An Exception to Mendel's First Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.5: Sex-Linkage- An Exception to Mendel's First Law** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.14: Phenotypes May Not Be As Expected from the Genotype

Environmental Factors

The phenotypes described thus far have a nearly perfect correlation with their associated genotypes; in other words an individual with a particular genotype always has the expected phenotype. However, many phenotypes are not determined entirely by genotype alone. They are instead determined by an interaction between genotype and non-genetic, environmental factors and can be conceptualized in the following relationship:

$$\text{Genotype} + \text{Environment} \Rightarrow \text{Phenotype} \quad (\text{G} + \text{E} \Rightarrow \text{P}) \tag{4.14.1}$$

Or:

$$\text{Genotype} + \text{Environment} + \overset{\text{Genetics and Environment}}{\text{Interaction}_{\text{GE}}} \Rightarrow \text{Phenotype} \quad (\text{G} + \text{E} + \text{I}_{\text{GE}} \Rightarrow \text{P}) \tag{4.14.2}$$

This interaction is especially relevant in the study of economically important phenotypes, such as human diseases or agricultural productivity. For example, a particular genotype may pre-dispose an individual to cancer, but cancer may only develop if the individual is exposed to certain DNA-damaging chemicals. Therefore, not all individuals with the particular genotype will develop the cancer phenotype.

Penetrance and Expressivity

The terms penetrance and expressivity are also useful to describe the relationship between certain genotypes and their phenotypes.

- **Penetrance** is the proportion of individuals (usually expressed as a percentage) with a particular genotype that display a corresponding phenotype (Figure 4.14.16). Because all pea plants that are homozygous for the allele for white flowers (e.g. *aa* in Figure 4.14.3) actually have white flowers, this genotype is completely penetrant. In contrast, many human genetic diseases are incompletely penetrant, since not all individuals with the disease genotype actually develop symptoms associated with the disease.
- **Expressivity** describes the variability in mutant phenotypes observed in individuals with a particular phenotype (Figure 4.14.16). Many human genetic diseases provide examples of broad expressivity, since individuals with the same genotypes may vary greatly in the severity of their symptoms. Incomplete penetrance and broad expressivity are due to random chance, non-genetic (environmental), and genetic factors (mutations in other genes).

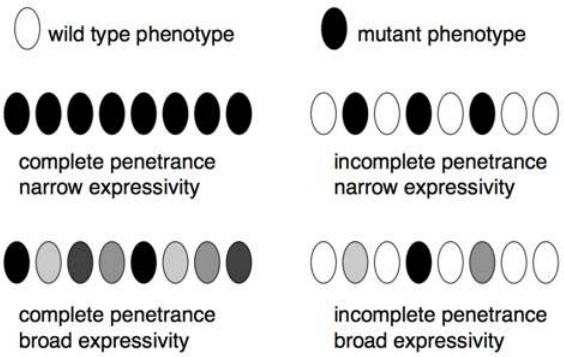


Figure 4.14.16: Relationship between penetrance and expressivity in eight individuals that all have a mutant genotype. Penetrance can be complete (all eight have the mutant phenotype) or incomplete (only some have the mutant phenotype). Amongst those individuals with the mutant phenotype the expressivity can be narrow (very little variation) to broad (lots of variation). (Original-Locke-CC:AN)

This page titled 4.14: Phenotypes May Not Be As Expected from the Genotype is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.6: Phenotypes May Not Be As Expected from the Genotype** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.15: Phenotypic Ratios May Not Be As Expected

Explanation

For a variety of reasons, the phenotypic ratios observed from real crosses rarely match the exact ratios expected based on a Punnett Square or other prediction techniques. There are many possible explanations for deviations from expected ratios. Sometimes these deviations are due to **sampling effects**, in other words, the random selection of a non-representative subset of individuals for observation. On the other hand, it may be because certain genotypes have a less than 100% survival rate. For example, *Drosophila* crosses sometimes give unexpected results because the more mutant alleles a zygote has the less likely it is to survive to become an adult. Genotypes that cause death for embryos or larvae are underrepresented when adult flies are counted.

The χTest For Goodness-of-fit

A statistical procedure called the **chi-square** (χ^2) test can be used to help a geneticist decide whether the deviation between observed and expected ratios is due to sampling effects, or whether the difference is so large that some other explanation must be sought by re-examining the assumptions used to calculate the expected ratio. The procedure for performing a chi-square test is covered in the labs.

This page titled 4.15: Phenotypic Ratios May Not Be As Expected is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.7: Phenotypic Ratios May Not Be As Expected** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.16: Extensions of the Laws of Inheritance

Mendel studied traits with only one mode of inheritance in pea plants. The inheritance of the traits he studied all followed the relatively simple pattern of dominant and recessive alleles for a single characteristic. There are several important modes of inheritance, discovered after Mendel's work, that do not follow the dominant and recessive, single-gene model.

Alternatives to Dominance and Recessiveness

Mendel's experiments with pea plants suggested that: 1) two types of "units" or alleles exist for every gene; 2) alleles maintain their integrity in each generation (no blending); and 3) in the presence of the dominant allele, the recessive allele is hidden, with no contribution to the phenotype. Therefore, recessive alleles can be "carried" and not expressed by individuals. Such heterozygous individuals are sometimes referred to as "carriers." Since then, genetic studies in other organisms have shown that much more complexity exists, but that the fundamental principles of Mendelian genetics still hold true. In the sections to follow, we consider some of the extensions of Mendelism.

Incomplete Dominance

Mendel's results, demonstrating that traits are inherited as dominant and recessive pairs, contradicted the view at that time that offspring exhibited a blend of their parents' traits. However, the heterozygote phenotype occasionally does appear to be intermediate between the two parents. For example, in the snapdragon, *Antirrhinum majus* (Figure 4.16.1), a cross between a homozygous parent with white flowers (C^WC^W) and a homozygous parent with red flowers (C^RC^R) will produce offspring with pink flowers (C^RC^W). (Note that different genotypic abbreviations are used for Mendelian extensions to distinguish these patterns from simple dominance and recessiveness.) This pattern of inheritance is described as incomplete dominance, meaning that one of the alleles appears in the phenotype in the heterozygote, but not to the exclusion of the other, which can also be seen. The allele for red flowers is incompletely dominant over the allele for white flowers. However, the results of a heterozygote self-cross can still be predicted, just as with Mendelian dominant and recessive crosses. In this case, the genotypic ratio would be 1 C^RC^R:2 C^RC^W:1 C^WC^W, and the phenotypic ratio would be 1:2:1 for red:pink:white. The basis for the intermediate color in the heterozygote is simply that the pigment produced by the red allele (anthocyanin) is diluted in the heterozygote and therefore appears pink because of the white background of the flower petals.



Figure 4.16.1**:** These pink flowers of a heterozygote snapdragon result from incomplete dominance. (credit: "storebukkebruse"/Flickr)

Codominance

A variation on incomplete dominance is codominance, in which both alleles for the same characteristic are simultaneously expressed in the heterozygote. An example of codominance occurs in the ABO blood groups of humans. The A and B alleles are expressed in the form of A or B molecules present on the surface of red blood cells. Homozygotes (I^AI^A and I^BI^B) express either the A or the B phenotype, and heterozygotes (I^AI^B) express both phenotypes equally. The I^AI^B individual has blood type AB. In a self-cross between heterozygotes expressing a codominant trait, the three possible offspring genotypes are phenotypically distinct. However, the 1:2:1 genotypic ratio characteristic of a Mendelian monohybrid cross still applies (Figure 4.16.2).

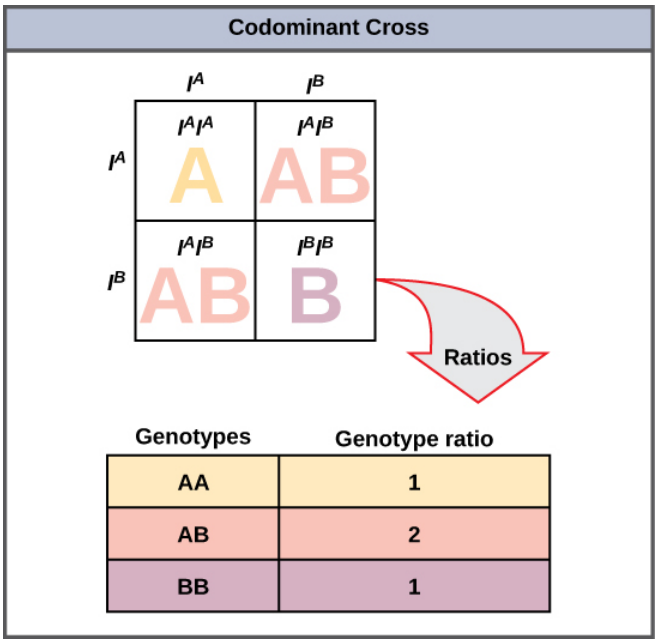


Figure 4.16.2**:** This Punnet square shows an AB/AB blood type cross.

Multiple Alleles

Mendel implied that only two alleles, one dominant and one recessive, could exist for a given gene. We now know that this is an oversimplification. Although individual humans (and all diploid organisms) can only have two alleles for a given gene, multiple alleles may exist at the population level, such that many combinations of two alleles are observed. Note that when many alleles exist for the same gene, the convention is to denote the most common phenotype or genotype in the natural population as the wild type (often abbreviated "+"). All other phenotypes or genotypes are considered variants (mutants) of this typical form, meaning they deviate from the wild type. The variant may be recessive or dominant to the wild-type allele.

An example of multiple alleles is the ABO blood-type system in humans. In this case, there are three alleles circulating in the population. The I^A allele codes for A molecules on the red blood cells, the I^B allele codes for B molecules on the surface of red blood cells, and the *i* allele codes for no molecules on the red blood cells. In this case, the I^A and I^B alleles are codominant with each other and are both dominant over the *i* allele. Although there are three alleles present in a population, each individual only gets two of the alleles from their parents. This produces the genotypes and phenotypes shown in Figure 4.16.3. Notice that instead of three genotypes, there are six different genotypes when there are three alleles. The number of possible phenotypes depends on the dominance relationships between the three alleles.

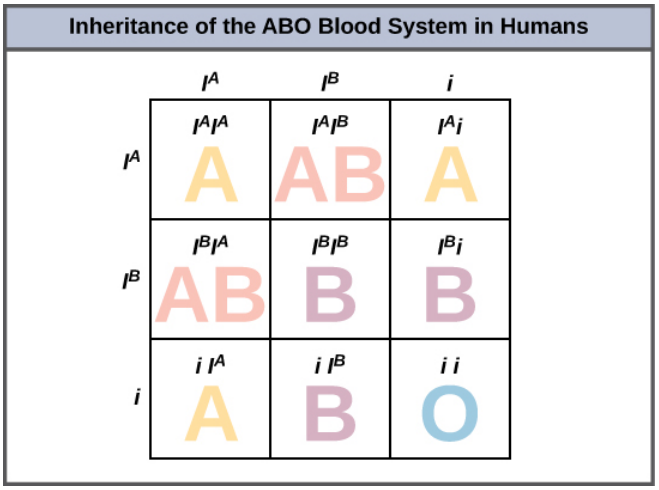


Figure 4.16.3**:** Inheritance of the ABO blood system in humans is shown.

EVOLUTION IN ACTION: Multiple Alleles Confer Drug Resistance in the Malaria Parasite

Malaria is a parasitic disease in humans that is transmitted by infected female mosquitoes, including *Anopheles gambiae*, and is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* and *P. vivax* are the most common causative agents of malaria, and *P. falciparum* is the most deadly. When promptly and correctly treated, *P. falciparum* malaria has a mortality rate of 0.1 percent. However, in some parts of the world, the parasite has evolved resistance to commonly used malaria treatments, so the most effective malarial treatments can vary by geographic region.

In Southeast Asia, Africa, and South America, *P. falciparum* has developed resistance to the anti-malarial drugs chloroquine, mefloquine, and sulfadoxine-pyrimethamine. *P. falciparum*, which is haploid during the life stage in which it is infective to humans, has evolved multiple drug-resistant mutant alleles of the *dhps* gene. Varying degrees of sulfadoxine resistance are associated with each of these alleles. Being haploid, *P. falciparum* needs only one drug-resistant allele to express this trait.

In Southeast Asia, different sulfadoxine-resistant alleles of the *dhps* gene are localized to different geographic regions. This is a common evolutionary phenomenon that comes about because drug-resistant mutants arise in a population and interbreed with other *P. falciparum* isolates in close proximity. Sulfadoxine-resistant parasites cause considerable human hardship in regions in which this drug is widely used as an over-the-counter malaria remedy. As is common with pathogens that multiply to large numbers within an infection cycle, *P. falciparum* evolves relatively rapidly (over a decade or so) in response to the selective pressure of commonly used anti-malarial drugs. For this reason, scientists must constantly work to develop new drugs or drug combinations to combat the worldwide malaria burden.[1]

Sex-Linked Traits

In humans, as well as in many other animals and some plants, the sex of the individual is determined by sex chromosomes—one pair of non-homologous chromosomes. Until now, we have only considered inheritance patterns among non-sex chromosomes, or autosomes. In addition to 22 homologous pairs of autosomes, human females have a homologous pair of X chromosomes, whereas human males have an XY chromosome pair. Although the Y chromosome contains a small region of similarity to the X chromosome so that they can pair during meiosis, the Y chromosome is much shorter and contains fewer genes. When a gene being examined is present on the X, but not the Y, chromosome, it is X-linked.

Eye color in *Drosophila*, the common fruit fly, was the first X-linked trait to be identified. Thomas Hunt Morgan mapped this trait to the X chromosome in 1910. Like humans, *Drosophila* males have an XY chromosome pair, and females are XX. In flies the wild-type eye color is red (X^W) and is dominant to white eye color (X^w) (Figure 4.16.4). Because of the location of the eye-color gene, reciprocal crosses do not produce the same offspring ratios. Males are said to be hemizygous, in that they have only one allele for any X-linked characteristic. Hemizygosity makes descriptions of dominance and recessiveness irrelevant for XY males. *Drosophila* males lack the white gene on the Y chromosome; that is, their genotype can only be X^WY or X^wY. In contrast, females have two allele copies of this gene and can be X^WX^W, X^WX^w, or X^wX^w.



Figure 4.16.4**:** In *Drosophila*, the gene for eye color is located on the X chromosome. Red eye color is wild-type and is dominant to white eye color.

In an X-linked cross, the genotypes of F_1 and F_2 offspring depend on whether the recessive trait was expressed by the male or the female in the P generation. With respect to *Drosophila* eye color, when the P male expresses the white-eye phenotype and the female is homozygously red-eyed, all members of the F_1 generation exhibit red eyes (Figure 4.16.5). The F_1 females are heterozygous (X^WX^w), and the males are all X^WY, having received their X chromosome from the homozygous dominant P female and their Y chromosome from the P male. A subsequent cross between the X^WX^w female and the X^WY male would produce only red-eyed females (with X^WX^W or X^WX^w genotypes) and both red- and white-eyed males (with X^WY or X^wY genotypes). Now, consider a cross between a homozygous white-eyed female and a male with red eyes. The F_1 generation would exhibit only heterozygous red-eyed females (X^WX^w) and only white-eyed males (X^wY). Half of the F_2 females would be red-eyed (X^WX^w) and half would be white-eyed (X^wX^w). Similarly, half of the F_2 males would be red-eyed (X^WY) and half would be white-eyed (X^wY).

ART CONNECTION

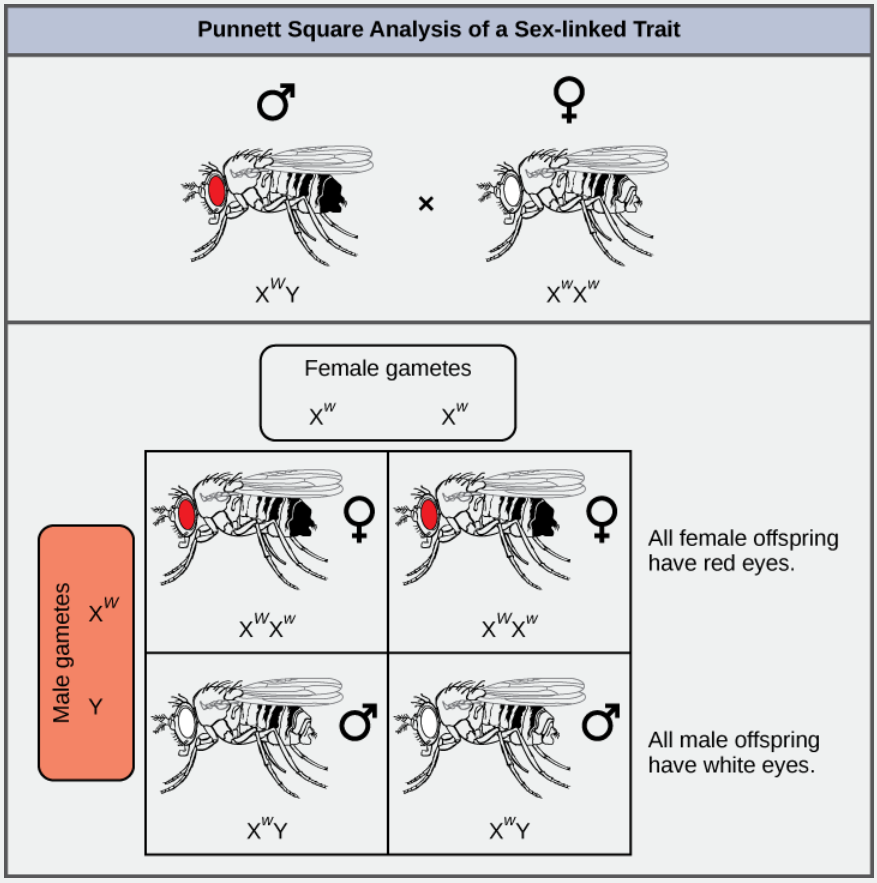


Figure 4.16.5: Crosses involving sex-linked traits often give rise to different phenotypes for the different sexes of offspring, as is the case for this cross involving red and white eye color in *Drosophila*. In the diagram, *w* is the white-eye mutant allele and *W* is the wild-type, red-eye allele.

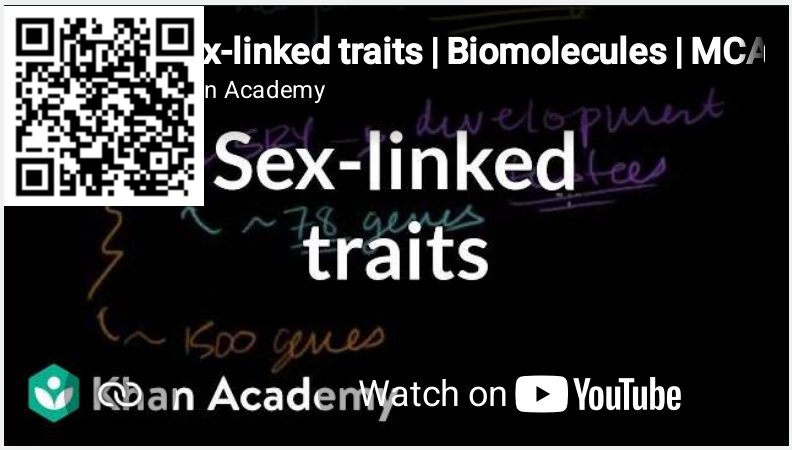
What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

Discoveries in fruit fly genetics can be applied to human genetics. When a female parent is homozygous for a recessive X-linked trait, she will pass the trait on to 100 percent of her male offspring, because the males will receive the Y chromosome from the male parent. In humans, the alleles for certain conditions (some color-blindness, hemophilia, and muscular dystrophy) are X-linked. Females who are heterozygous for these diseases are said to be carriers and may not exhibit any phenotypic effects. These females will pass the disease to half of their sons and will pass carrier status to half of their daughters; therefore, X-linked traits appear more frequently in males than females.

In some groups of organisms with sex chromosomes, the sex with the non-homologous sex chromosomes is the female rather than the male. This is the case for all birds. In this case, sex-linked traits will be more likely to appear in the female, in whom they are

hemizygous.

CONCEPT IN ACTION



Watch this video to learn more about sex-linked traits.

Linked Genes Violate the Law of Independent Assortment

Although all of Mendel's pea plant characteristics behaved according to the law of independent assortment, we now know that some allele combinations are not inherited independently of each other. Genes that are located on separate, non-homologous chromosomes will always sort independently. However, each chromosome contains hundreds or thousands of genes, organized linearly on chromosomes like beads on a string. The segregation of alleles into gametes can be influenced by linkage, in which genes that are located physically close to each other on the same chromosome are more likely to be inherited as a pair. However, because of the process of recombination, or "crossover," it is possible for two genes on the same chromosome to behave independently, or as if they are not linked. To understand this, let us consider the biological basis of gene linkage and recombination.

Homologous chromosomes possess the same genes in the same order, though the specific alleles of the gene can be different on each of the two chromosomes. Recall that during interphase and prophase I of meiosis, homologous chromosomes first replicate and then synapse, with like genes on the homologs aligning with each other. At this stage, segments of homologous chromosomes exchange linear segments of genetic material (Figure 4.16.6). This process is called recombination, or crossover, and it is a common genetic process. Because the genes are aligned during recombination, the gene order is not altered. Instead, the result of recombination is that maternal and paternal alleles are combined onto the same chromosome. Across a given chromosome, several recombination events may occur, causing extensive shuffling of alleles.

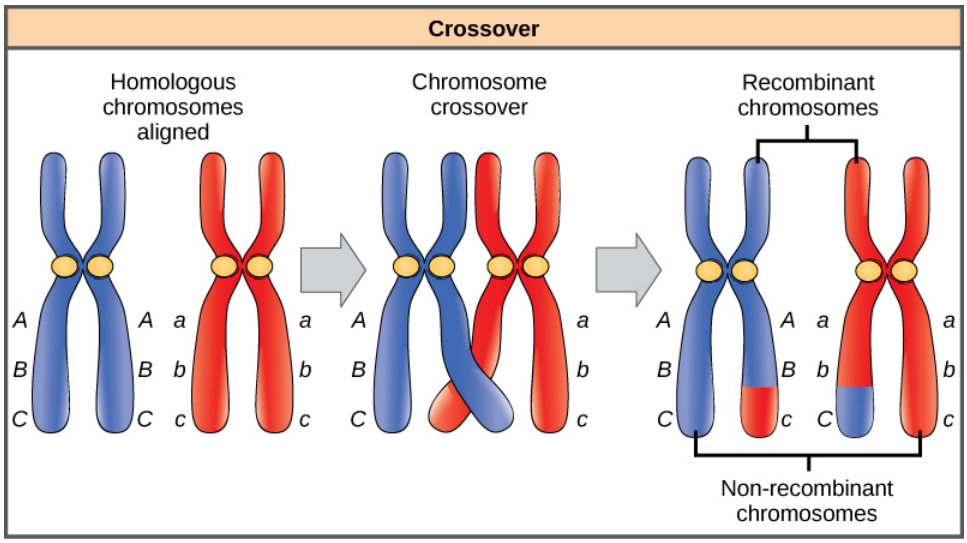


Figure 4.16.6**:** The process of crossover, or recombination, occurs when two homologous chromosomes align and exchange a segment of genetic material.

When two genes are located on the same chromosome, they are considered linked, and their alleles tend to be transmitted through meiosis together. To exemplify this, imagine a dihybrid cross involving flower color and plant height in which the genes are next to each other on the chromosome. If one homologous chromosome has alleles for tall plants and red flowers, and the other chromosome has genes for short plants and yellow flowers, then when the gametes are formed, the tall and red alleles will tend to go together into a gamete and the short and yellow alleles will go into other gametes. These are called the parental genotypes because they have been inherited intact from the parents of the individual producing gametes. But unlike if the genes were on different chromosomes, there will be no gametes with tall and yellow alleles and no gametes with short and red alleles. If you create a Punnett square with these gametes, you will see that the classical Mendelian prediction of a 9:3:3:1 outcome of a dihybrid cross would not apply. As the distance between two genes increases, the probability of one or more crossovers between them increases and the genes behave more like they are on separate chromosomes. Geneticists have used the proportion of recombinant gametes (the ones not like the parents) as a measure of how far apart genes are on a chromosome. Using this information, they have constructed linkage maps of genes on chromosomes for well-studied organisms, including humans.

Mendel's seminal publication makes no mention of linkage, and many researchers have questioned whether he encountered linkage but chose not to publish those crosses out of concern that they would invalidate his independent assortment postulate. The garden pea has seven chromosomes, and some have suggested that his choice of seven characteristics was not a coincidence. However, even if the genes he examined were not located on separate chromosomes, it is possible that he simply did not observe linkage because of the extensive shuffling effects of recombination.

Epistasis

Mendel's studies in pea plants implied that the sum of an individual's phenotype was controlled by genes (or as he called them, unit factors), such that every characteristic was distinctly and completely controlled by a single gene. In fact, single observable characteristics are almost always under the influence of multiple genes (each with two or more alleles) acting in unison. For example, at least eight genes contribute to eye color in humans.

CONCEPT IN ACTION

Eye color in humans is determined by multiple alleles. Use the Eye Color Calculator to predict the eye color of children from parental eye color.

In some cases, several genes can contribute to aspects of a common phenotype without their gene products ever directly interacting. In the case of organ development, for instance, genes may be expressed sequentially, with each gene adding to the complexity and specificity of the organ. Genes may function in complementary or synergistic fashions, such that two or more genes expressed simultaneously affect a phenotype. An apparent example of this occurs with human skin color, which appears to involve the action of at least three (and probably more) genes. Cases in which inheritance for a characteristic like skin color or human height depend on the combined effects of numerous genes are called polygenic inheritance.

Genes may also oppose each other, with one gene suppressing the expression of another. In epistasis, the interaction between genes is antagonistic, such that one gene masks or interferes with the expression of another. “Epistasis” is a word composed of Greek roots meaning “standing upon.” The alleles that are being masked or silenced are said to be hypostatic to the epistatic alleles that are doing the masking. Often the biochemical basis of epistasis is a gene pathway in which expression of one gene is dependent on the function of a gene that precedes or follows it in the pathway.

An example of epistasis is pigmentation in mice. The wild-type coat color, agouti (*AA*) is dominant to solid-colored fur (*aa*). However, a separate gene *C*, when present as the recessive homozygote (*cc*), negates any expression of pigment from the *A* gene and results in an albino mouse (Figure 4.16.7). Therefore, the genotypes *AAcc*, *Aacc*, and *aacc* all produce the same albino phenotype. A cross between heterozygotes for both genes (*AaCc* x *AaCc*) would generate offspring with a phenotypic ratio of 9 agouti:3 black:4 albino (Figure 4.16.7). In this case, the *C* gene is epistatic to the *A* gene.

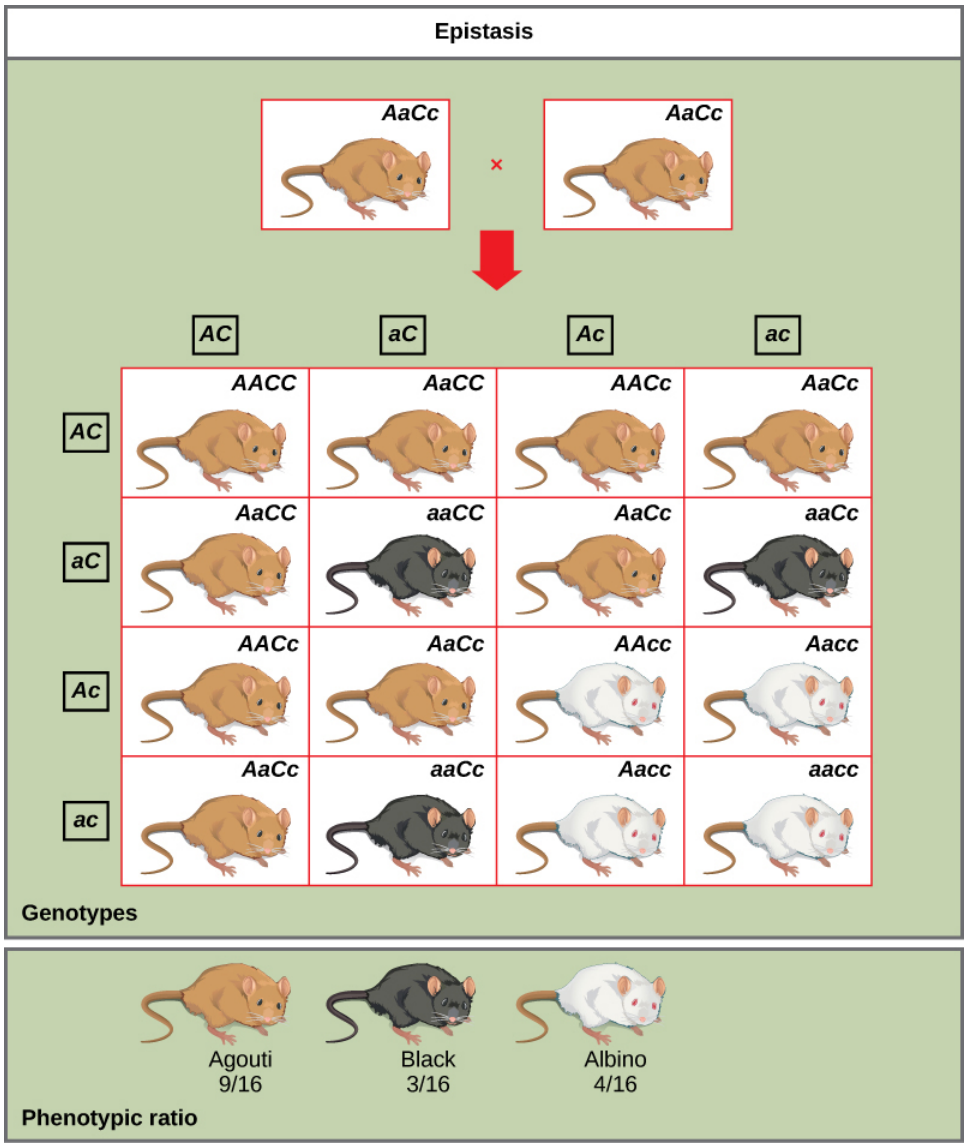


Figure 4.16.7**:** In this example of epistasis, one gene (*C*) masks the expression of another (*A*) for coat color. When the *C* allele is present, coat color is expressed; when it is absent (*cc*), no coat color is expressed. Coat color depends on the *A* gene, which shows dominance, with the recessive homozygote showing a different phenotype than the heterozygote or dominant homozygote.

Section Summary

Alleles do not always behave in dominant and recessive patterns. Incomplete dominance describes situations in which the heterozygote exhibits a phenotype that is intermediate between the homozygous phenotypes. Codominance describes the simultaneous expression of both of the alleles in the heterozygote. Although diploid organisms can only have two alleles for any given gene, it is common for more than two alleles for a gene to exist in a population. In humans, as in many animals and some plants, females have two X chromosomes and males have one X and one Y chromosome. Genes that are present on the X but not the Y chromosome are said to be X-linked, such that males only inherit one allele for the gene, and females inherit two.

According to Mendel's law of independent assortment, genes sort independently of each other into gametes during meiosis. This occurs because chromosomes, on which the genes reside, assort independently during meiosis and crossovers cause most genes on the same chromosomes to also behave independently. When genes are located in close proximity on the same chromosome, their alleles tend to be inherited together. This results in offspring ratios that violate Mendel's law of independent assortment. However, recombination serves to exchange genetic material on homologous chromosomes such that maternal and paternal alleles may be recombined on the same chromosome. This is why alleles on a given chromosome are not always inherited together. Recombination is a random event occurring anywhere on a chromosome. Therefore, genes that are far apart on the same

chromosome are likely to still assort independently because of recombination events that occurred in the intervening chromosomal space.

Whether or not they are sorting independently, genes may interact at the level of gene products, such that the expression of an allele for one gene masks or modifies the expression of an allele for a different gene. This is called epistasis.

Art Connections

Figure 4.16.5: What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

Answer

Half of the female offspring would be heterozygous (X^WX^w) with red eyes, and half would be homozygous recessive (X^wX^w) with white eyes. Half of the male offspring would be hemizygous dominant (X^WY) with red eyes, and half would be hemizygous recessive (X^wY) with white eyes.

Footnotes

1. 1 Sumiti Vinayak et al., "Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*," *PLoS Pathogens* 6 (2010): e1000830.

Glossary

codominance
in a heterozygote, complete and simultaneous expression of both alleles for the same characteristic

epistasis
an interaction between genes such that one gene masks or interferes with the expression of another

hemizygous
the presence of only one allele for a characteristic, as in X-linkage; hemizygosity makes descriptions of dominance and recessiveness irrelevant

incomplete dominance
in a heterozygote, expression of two contrasting alleles such that the individual displays an intermediate phenotype

linkage
a phenomenon in which alleles that are located in close proximity to each other on the same chromosome are more likely to be inherited together

recombination
the process during meiosis in which homologous chromosomes exchange linear segments of genetic material, thereby dramatically increasing genetic variation in the offspring and separating linked genes

wild type
the most commonly occurring genotype or phenotype for a given characteristic found in a population

X-linked
a gene present on the X chromosome, but not the Y chromosome

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).

This page titled 4.16: Extensions of the Laws of Inheritance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **8.3: Extensions of the Laws of Inheritance** by OpenStax is licensed CC BY 4.0.

4.E: Genetic Analysis of Single Genes (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Q3.1

hat is the maximum number of alleles for a given locus in a normal gamete of a diploid species?

Q3.2

Wirey hair (*W*) is dominant to smooth hair (*w*) in dogs.

a. If you cross a homozygous, wirey-haired dog with a smooth-haired dog, what will be the genotype and phenotype of the F1 generation?
b. If two dogs from the F1 generation mated, what would be the most likely ratio of hair phenotypes among their progeny?
c. When two wirey-haired *Ww* dogs actually mated, they had alitter of three puppies, which all had smooth hair. How do you explain this observation?
d. Someone left a wirey-haired dog on your doorstep. Without extracting DNA, what would be the easiest way to determine the genotype of this dog?
e. Based on the information provided in question 1, can you tell which, if either, of the alleles is wild-type?

Q3.3

An important part of Mendel's experiments was the use of homozygous lines as parents for his crosses. How did he know they were homozygous, and why was the use of the lines important?

Q3.4

Does equal segregation of alleles into daughter cells happen during mitosis, meiosis, or both?

Q3.5

If your blood type is B, what are the possible genotypes of your parents at the locus that controls the ABO blood types?

Q3.6

In the table below, match the mouse hair color phenotypes with the term from the list that best explains the observed phenotype, given the genotypes shown. In this case, the allele symbols do not imply anything about the dominance relationships between the alleles. List of terms: haplosufficiency, haploinsufficiency, pleiotropy, incomplete dominance, co-dominance, incomplete penetrance, broad (variable) expressivity.

Table for Question 3.6

	A_1A_1	A_1A_2	A_2A_2
1	all hairs black	on the same individual: 50% of hairs are all black and 50% of hairs are all white	all hairs wh
2	all hairs black	all hairs are the same shade of grey	all hairs wh
3	all hairs black	all hairs black	50% of individuals have all white hairs and hairs
4	all hairs black	all hairs black	mice have no
5	all hairs black	all hairs white	all hairs wh
6	all hairs black	all hairs black	all hairs wh
7	all hairs black	all hairs black	hairs are a wide range of

Q3.7

A rare dominant mutation causes a neurological disease that appears late in life in all people that carry the mutation. If a father has this disease, what is the probability that his daughter will also have the disease?

Q3.8

Make Punnett Squares to accompany the crosses shown in Figure 3.10.

Q3.9

Another cat hair colour gene is called *White Spotting*. This gene is autosomal. Cats that have the dominant *S* allele have white spots. What are the possible genotypes of cats that are:

a. entirely black
b. entirely orange
c. black and white
d. orange and white
e. orange and black (tortoiseshell)
f. orange, black, and white (calico)

Q3.10

Draw reciprocal crosses that would demonstrate that the turkey E-gene is on the Z chromosome.

Q3.11

Mendel's First Law (as stated in class) does not apply to alleles of most genes located on sex chromosomes. Does the law apply to the chromosomes themselves?

Q3.12

What is the relationship between the O^0 and O^B alleles of the *Orange* gene in cats?

Q3.13

Make a diagram similar to those in Figures 3.9, 3.11, and 3.13 that shows the relationship between genotype and phenotype for the *F8* gene in humans.

Answers

3.1 There is a maximum of two alleles for a normal autosomal locus in a diploid species.

3.2 a) In the F_1 generation, the genotype of all individuals will be *Ww* and all of the dogs will have wirey hair.

b) In the F_2 generation, there would be an expected 3:1 ratio of wirey-haired to smooth-haired dogs.

c) Although it is expected that only one out of every four dogs in the F2 generation would have smooth hair, large deviations from this ratio are possible, especially with small sample sizes. These deviations are due to the random nature in which gametes combine to produce offspring. Another example of this would be the fairly common observation that in some human families, all of the offspring are either girls, or boys, even though the expected ratio of the sexes is essentially 1:1.

d) You could do a test cross, i.e. cross the wirey-haired dog to a homozygous recessive dog (*ww*). Based on the phenotypes among the offspring, you might be able to infer the genotype of the wirey-haired parent.

e) From the information provided, we cannot be certain which, if either, allele is wild-type. Generally, dominant alleles are wild-type, and abnormal or mutant alleles are recessive.

3.3 Even before the idea of a homozygous genotype had really been formulated, Mendel was still able to assume that he was working with parental lines that contained the genetic material for only one variant of a trait (e.g. EITHER green seeds of yellow seeds), because these lines were pure-breeding. Pure-breeding means that the phenotype doesn't change over several generations of self-pollination. If the parental lines had not been pure-breeding, it would have been very hard to make certain key inferences, such as that the F_1 generation could contain the genetic information for two variants of a trait, although only one variant was expressed. This inference led eventually to Mendel's First Law.

3.4 Equal segregation of alleles occurs only in meiosis. Although mitosis does produce daughter cells that are genetically equal, there is no segregation (i.e. separation) of alleles during mitosis; each daughter cell contains both of the alleles that were originally present.

3.5 If your blood type is B, then your genotype is either I^Bi or I^BI^B. If your genotype is I^Bi, then your parents could be any combination of genotypes, as long as one parent had at least one *i* allele, and the other parent had at least one I^B allele. If your genotype was $I^B\ I^B$, then both parents would have to have at least one I^B allele.

3.6 case 1 co-dominance

case 2 incomplete-dominance

case 3 incomplete penetrance

case 4 pleiotropy

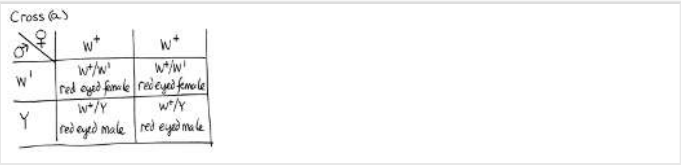
case 5 haplosufficiency

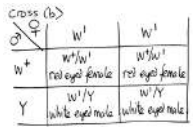
case 6 haploinsufficiency

case 7 broad (variable) expressivity

3.7 If the gene is autosomal, the probability is 50%. If it is sex-linked, 100%. In both situations the probability would decrease if the penetrance was less than 100%.

3.8

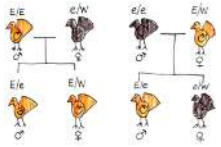




3.9 Note that a semicolon is used to separate genes on different chromosomes.

	Phenotype	Genotype(s)
a	entirely black	O^B / O^B ; *s* / *s* O^B / Y ; *s* / *s*
b	entirely orange	O^0 / O^0 ; *s* / *s* O^0 / Y ; *s* / *s*
c	black and white	O^B / O^B ; S / _ O^B / Y ; S / _
d	orange and white	O^0 / O^0 ; S / _ O^0 / Y ; S / _
e	orange and black (tortoiseshell)	O^0 / O^B ; *s* / *s*
f	orange, black, and white (calico)	O^0 / O^B ; S / _

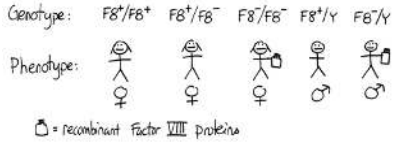
3.10



3.11 Because each egg or sperm cell receives exactly one sex chromosome (even though this can be either an X or Y, in the case of sperm), it could be argued that the sex chromosomes themselves do obey the law of equal segregation, even though the alleles they carry may not always segregate equally. However, this answer depends on how broadly you are willing to stretch Mendel's First Law.

3.12 Co-dominance

3.13 People with hemophilia A use injections of recombinant Factor VIII proteins on demand (to control bleeding) or regularly (to limit damage to joints).



https://bio.libretexts.org/@go/page/27241

This page titled 4.E: Genetic Analysis of Single Genes (Exercises) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

4.S: Genetic Analysis of Single Genes (Summary)

- A diploid can have up to two different alleles at a single locus. The alleles segregate equally between gametes during meiosis.
- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- Classical geneticists make use of true breeding lines, monohybrid crosses, Punnett squares, test crosses, reciprocal crosses, and the chi-square test.
- Sex-linked genes are an exception to standard Mendelian inheritance. Their phenotypes are influenced by the type of sex chromosome system and the type of dosage compensation system found in the species.
- The male/female phenotype (sex) can be determined by chromosomes, genes, or the environment.

Key Terms

allele	co-dominance	hermaphrodites
Mendel's First Law	ABO blood group	parthenogenesis
Law of Equal Segregation	haplosufficiency	hormonal
homozygous	haploinsufficiency	cell-autonomous
heterozygous	classical genetics	sexual gynandro-morphs
hemizygous	molecular genetics	sexually dimorphic
wild-type	true breeding lines	gynandromorphy
variant	monohybrid cross	mosaic
locus	Punnett Square	chimera
genotype	test cross	G + E = P
phenotype	tester	penetrance
dominant	sex-linked	expressivity
recessive	dosage compensation	sampling effects
complete dominance	X-linked genes	chi-square χ^2 test
incomplete (semi) dominance	autosomal genes	
	reciprocal cross	
	Z-linked genes	

This page titled 4.S: Genetic Analysis of Single Genes (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **3.S: Genetic Analysis of Single Genes (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

5: Genetic Analysis of Multiple Genes

The principles of genetic analysis that we have described for a single locus can be extended to the study of alleles at two loci simultaneously. Analysis of two loci in parallel is required for genetic mapping and can also reveal gene interactions. These techniques are very useful for both basic and applied research. Before discussing these techniques, we will first revisit Mendel's classical experiments.

5.1: Dihybrid Crosses
5.2: Epistasis and Other Gene Interactions
5.3: Example of Multiple Genes Affecting One Character
5.4: Alleles at a Single Locus
5.4.1: Introduction
5.4.2: Somatic vs. Germline Mutations
5.4.3: Alleles- Hetero-, Homo-, Hemizygosity
5.4.4: Pleiotropy vs. Polygenic Inheritance
5.4.5: Types of Dominance
5.4.6: Biochemical Basis of Dominance
5.4.7: Classification of Mutants
5.4.8: Muller's Morphs
5.4.9: Chapter 6 Summary
5.4.10: Key Terms in Chapter 6
5.4.11: Chapter 6 Study Questions
5.4.12: Chapter 6 Answers
5.E: Genetic Analysis of Multiple Genes (Exercises)
5.S: Genetic Analysis of Multiple Genes (Summary)

Thumbnail: Coat color in animals is an example of a trait that is controlled by more than one locus. (Flickr-Gossamer1013-CC:AND).

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 5: Genetic Analysis of Multiple Genes is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Dihybrid Crosses

Mendel's Second Law

Before Mendel, it had not yet been established that heritable traits were controlled by discrete factors. Therefor an important question was therefore whether distinct traits were controlled by discrete factors that were inherited independently of each other? To answer this, Mendel took two apparently unrelated traits, such as seed shape and seed color, and studied their inheritance together in one individual. He studied two variants of each trait: seed color was either green or yellow, and seed shape was either round or wrinkled. (He studied seven traits in all.) When either of these traits was studied individually, the phenotypes segregated in the classical 3:1 ratio among the progeny of a monohybrid cross (Figure 5.1.2), with ¾ of the seeds green and ¼ yellow in one cross, and ¾ round and ¼ wrinkled in the other cross. Would this be true when both were in the same individual?

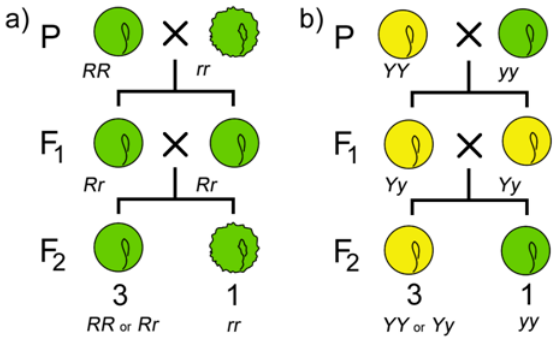


Figure 5.1.2: Monohybrid crosses involving two distinct traits in peas. a) is R/r and b) is Y/y. *(Original-Deyholos-CC:AN)*

To analyze the segregation of both traits at the same time in the same individual, he crossed a pure breeding line of green, wrinkled peas with a pure breeding line of yellow, round peas to produce F_1 progeny that were all green and round, and which were also **dihybrids**; they carried two alleles at each of two loci (Figure 5.1.3).

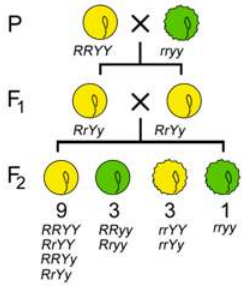


Figure 5.1.3: Pure-breeding lines are crossed to produce dihybrids in the F_1 generation. The cross of these particular dihybrids produces four phenotypic classes. (Original-Deyholos-CC:AN)

If the inheritance of seed color was truly independent of seed shape, then when the F_1 dihybrids were crossed to each other, a 3:1 ratio of one trait should be observed within each phenotypic class of the other trait (Figure 5.1.3). Using the product law, we would therefore predict that if ¾ of the progeny were green, and ¾ of the progeny were round, then ¾ × ¾ = 9/16 of the progeny would be both round and green (Table 5.1.1). Likewise, ¾ × ¼ = 3/16 of the progeny would be both round and yellow, and so on. By applying the product rule to all of these combinations of phenotypes, we can predict a **9:3:3:1** phenotypic ratio among the progeny of a dihybrid cross, if certain conditions are met, including the independent segregation of the alleles at each locus. Indeed, 9:3:3:1 is very close to the ratio Mendel observed in his studies of dihybrid crosses, leading him to state his Second Law, the **Law of Independent Assortment**, which we now express as follows: two loci assort independently of each other during gamete formation.

Definition: Mendel's Second Law

Two loci assort *independently* of each other during gamete formation.

https://bio.libretexts.org/@go/page/27263

Table 5.1.1: Phenotypic classes expected in monohybrid and dihybrid crosses for two seed traits in pea.

						Frequency of phenotypic crosses within separate monohybrid crosses: seed shape: ¾ round ¼ wrinkled seed color: ¾ yellow ¼ green **Frequency of phenotypic crosses within a dihybrid cross:** ¾ round × ¾ yellow = 9/16 round & yellow ¾ round × ¼ green = 3/16 round & green ¼ wrinkled × ¾ yellow = 3/16 wrinkled & yellow ¼ wrinkled × ¼ green = 1/16 wrinkled & green

The 9:3:3:1 phenotypic ratio that we calculated using the product rule can also be obtained using Punnett Square (Figure 5.1.4). First, we list the genotypes of the possible gametes along each axis of the Punnett Square. In a diploid with two heterozygous genes of interest, there are up to four combinations of alleles in the gametes of each parent. The gametes from the respective rows and column are then combined in the each cell of the array. When working with two loci, genotypes are written with the symbols for both alleles of one locus, followed by both alleles of the next locus (e.g. *AaBb*, not *ABab*). Note that the order in which the loci are written does not imply anything about the actual position of the loci on the chromosomes.

	RY	***Ry***	***rY***	***ry***
RY	*RRYY*	*RRYy*	*RrYY*	*RrYy*
Ry	*RRYy*	*RRyy*	*RrYy*	*Rryy*
rY	*RrYY*	*RrYy*	*rrYY*	*rrYy*
ry	*RrYy*	*Rryy*	*rrYy*	*rryy*

Figure 5.1.3. Each of the four phenotypic classes is represented by a different color of shading. (Original-Deyholos-CC:AN)

To calculate the expected phenotypic ratios, we assign a phenotype to each of the 16 genotypes in the Punnett Square, based on our knowledge of the alleles and their dominance relationships. In the case of Mendel's seeds, any genotype with at least one *R* allele and one *Y* allele will be round and yellow; these genotypes are shown in the nine, green-shaded cells in Figure 5.1.4. We can represent all of four of the different genotypes shown in these cells with the notation (*R_Y_*), where the blank line (__), means "any

allele". The three offspring that have at least one R allele and are homozygous recessive for *y* (i.e. *R_yy*) will have a round, green phenotype. Conversely the three progeny that are homozygous recessive *r*, but have at least one *Y* allele (*rrY_*) will have wrinkled, yellow seeds. Finally, the rarest phenotypic class of wrinkled, yellow seeds is produced by the doubly homozygous recessive genotype, *rryy*, which is expected to occur in only one of the sixteen possible offspring represented in the square.

Assumptions of the 9:3:3:1 ratio

Both the product rule and the Punnett Square approaches showed that a 9:3:3:1 phenotypic ratio is expected among the progeny of a dihybrid cross such as Mendel's *RrYy* × *RrYy*. In making these calculations, we assumed that:

1. both loci assort independently;
2. one allele at each locus is completely dominant; and
3. each of four possible phenotypes can be distinguished unambiguously, with no interactions between the two genes that would alter the phenotypes.

Deviations from the 9:3:3:1 phenotypic ratio may indicate that one or more of the above conditions has not been met. Modified ratios in the progeny of a dihybrid cross can therefore reveal useful information about the genes involved.

Linkage is one of the most important reasons for distortion of the ratios expected from independent assortment. Linked genes are located close together on the same chromosome. This close proximity alters the frequency of allele combinations in the gametes. We will return to the concept of linkage in Chapter 7. Deviations from 9:3:3:1 ratios can also be due to interactions between genes. These interactions will be discussed in the remainder of this chapter. For simplicity, we will focus on examples that involve easily scored phenotypes, such as pigmentation. Nevertheless, keep in mind that the analysis of segregation ratios of any markers can provide insight into a wide range of biological processes they represent.

This page titled 5.1: Dihybrid Crosses is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.1: Dihybrid Crosses** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

5.2: Epistasis and Other Gene Interactions

Some dihybrid crosses produce a phenotypic ratio that differs from 9:3:3:1, such as 9:3:4, 12:3:1, 9:7, or 15:1. Note that each of these modified ratios can be obtained by summing one or more of the 9:3:3:1 classes expected from our original dihybrid cross. In the following sections, we will look at some modified phenotypic ratios obtained from dihybrid crosses and what they might tell us about interactions between genes.

Recessive epistasis

Epistasis (which means "standing upon") occurs when the phenotype of one locus masks, or prevents, the phenotype of another locus. Thus, following a dihybrid cross fewer than the typical four phenotypic classes will be observed with epistasis. As we have already discussed, in the absence of epistasis, there are four phenotypic classes among the progeny of a dihybrid cross. The four phenotypic classes correspond to the genotypes: *A_B_*, *A_bb*, *aaB_*, and *aabb*. If either of the singly homozygous recessive genotypes (i.e. *A_bb* or *aaB_*) has the same phenotype as the double homozygous recessive (*aabb*), then a **9:3:4** phenotypic ratio will be obtained. For example, in the Labrador Retriever breed of dogs (Figure 5.2.5), the B locus encodes a gene for an important step in the production of melanin. The dominant allele, *B* is more efficient at pigment production than the recessive *b* allele, thus *B_* hair appears black, and *bb* hair appears brown. A second locus, which we will call *E*, controls the deposition of melanin in the hairs. At least one functional *E* allele is required to deposit any pigment, whether it is black or brown. Thus, all retrievers that are *ee* fail to deposit any melanin (and so appear pale yellow), regardless of the genotype at the *B* locus (Figure 5.2.6).



Figure 5.2.5: Labrador Retrievers with different coat colors: (from left to right) black, chocolate, yellow: an example of recessive epistasis. (Flickr-John Curley/PhilRomans/Miss Chien-CC:AND)

The *ee* genotype is therefore said to be **epistatic** to both the B and b alleles, since the homozygous *ee* phenotype masks the phenotype of the B locus. The B/b locus is said to be **hypostatic** to the *ee* genotype. Because the masking allele is in this case is recessive, this is called **recessive epistasis**.

	EB	*Eb*	*eB*	*eb*
EB	*EEBB*	*EEBb*	*EeBB*	*EeBb*
Eb	*EEBb*	*EEbb*	*EeBb*	*Eebb*
eB	*EeBB*	*EeBb*	*eeBB*	*eeBb*
eb	*EeBb*	*Eebb*	*eeBb*	*eebb*

Figure 5.2.6: Genotypes and phenotypes among the progeny of a dihybrid cross of Labrador Retrievers heterozygous for two loci affecting coat color. The phenotypes of the progeny are indicated by the shading of the cells in the table: black coat (black, E_B_); chocolate coat (brown, E_bb); yellow coat (yellow, eeB_ or eebb). (Original-Locke-CC:AN)

Dominant epistasis

In some cases, a dominant allele at one locus may mask the phenotype of a second locus. This is called **dominant epistasis**, which produces a segregation ratio such as **12:3:1**, which can be viewed as a modification of the 9:3:3:1 ratio in which the *A_B_* class is combined with one of the other genotypic classes that contains a dominant allele. One of the best known examples of a 12:3:1 segregation ratio is fruit color in some types of squash (Figure 5.2.7). Alleles of a locus that we will call *B* produce either yellow (*B_*) or green (*bb*) fruit. However, in the presence of a dominant allele at a second locus that we call *A*, no pigment is produced at all, and fruit are white. The dominant *A* allele is therefore epistatic to both *B* and *bb* combinations (Figure 5.2.8). One possible biological interpretation of this segregation pattern is that the function of the *A* allele somehow blocks an early stage of pigment synthesis, before neither yellow or green pigments are produced.



Figure 5.2.7: Green, yellow, and white fruits of squash. (Flickr-Unknown-CC:AN)

	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 5.2.8***:*** *Genotypes and phenotypes among the progeny of a dihybrid cross of squash plants heterozygous for two loci affecting fruit color. (Original-Deyholos-CC:AN)*

This page titled 5.2: Epistasis and Other Gene Interactions is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.2: Epistasis and Other Gene Interactions** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

5.3: Example of Multiple Genes Affecting One Character

Cat Fur Genetics

Most aspects of the fur phenotypes of common cats can be explained by the action of just a few genes (Table 6-2). Other genes, not described here, may further modify these traits and account for the phenotypes seen in tabby cats and in more exotic breeds, such as Siamese.

For example, the X-linked ***Orange*** gene has two allelic forms. The O^O allele produces orange fur, while the O^B alleles produce non-orange (often black) fur. Note however, that because of X-chromosome inactivation the result is mosaicism in expression. In O^O / O^B female heterozygotes patches of black and orange are seen, which produces the tortoise shell pattern (Figure 6-13 A,B). This is a rare example of **co-dominance** since the phenotype of both alleles can be seen. Note that the cat in part A has short fur compared to the cat in part B; recessive alleles at an independent locus (L/l) produce long (*ll*) rather than short (*L_*) fur.

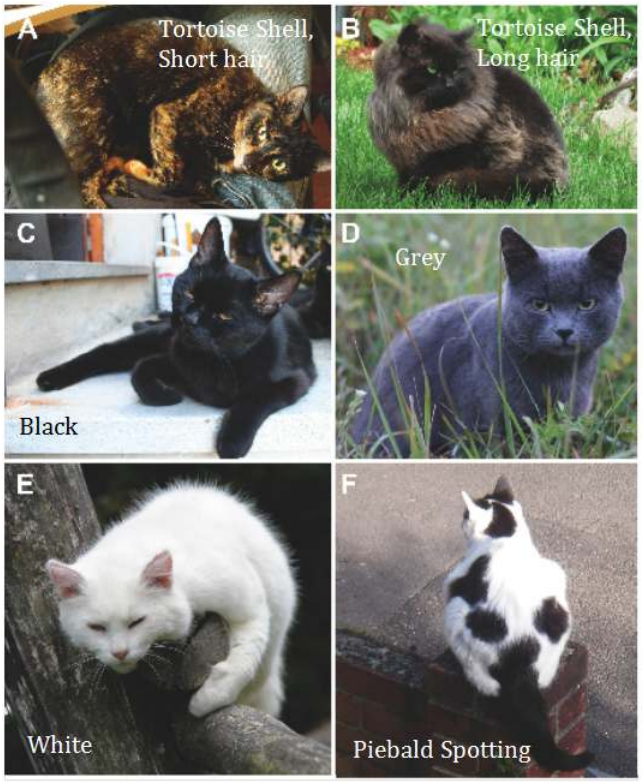


Figure 5.3.13: **Representatives of various fur phenotypes in cats:** Tortoise shell (A,B) pigmentation in cats with short (A) and long (B) fur; black (C) and grey (D) cats that differ in genotype at the dilute locus. The pure white pattern (E) is distinct from piebald spotting (F). (flickr-cwbuecheler–CC:AN, flickr-=-.0=-CC:AND, flickr-Leonardo Zanchi CC:AS, flickr-malfet_ CC:AN, flickr- valentinastorti-CC:AN, flickr- Denni Schnapp-CC:AS)

Alleles of the ***dilute*** gene affect the intensity of pigmentation, regardless of whether that pigmentation is due to black or orange pigment. Part C shows a black cat with at least one dominant allele of *dilute* (*D_*), in contrast to the cat in D, which is grey rather than black, because it has the *dd* genotype.

Epistasis is demonstrated by an allele of only one of the genes in Table 5.3.2. One dominant allele of ***white masking*** (*W*) prevents normal development of melanocytes (pigment producing cells). Therefore, cats with genotype (*W_*) will have entirely white fur regardless of the genotype at the *Orange* or *dilute* loci (part E). Although this locus produces a white colour, *W_* is not the same as albinism, which is a much rarer phenotype caused by mutations in other genes. Albino cats can be distinguished by having red eyes, while *W_* cats have eyes that are not red.

Piebald spotting is the occurrence of patches of white fur. These patches vary in size due to many reasons, including genotype. Homozygous cats with genotype *ss* do not have any patches of white, while cats of genotype *Ss* and *SS* do have patches of white,

and the homozygotes tend to have a larger proportion of white fur than heterozygotes (part F). The combination of piebald spotting and tortoise shell patterning produce a **calico cat**, which has separate patches of orange, black, and white fur.

Table 5.3.2: Summary of simplified cat fur phenotypes and genotypes.

Trait	Phenotype	Genotype	Comments
fur length	short	*LL* or *Ll*	*L* is completely dominant
	long	*ll*	
all white fur (non-albino)	100% white fur	*WW* or *Ww*	If the cat has red eyes it is albino, not *W_*. *W* is epistatic to all other fur color genes; if cat is *W_*, can't infer genotypes for any other fur color genes.
		ww	
piebald spotting	> 50% white patches (but not 100%)	*SS*	*S* is incompletely dominant and shows variable expressivity
	< 50% white patches	*Ss*	
	no white patches	*ss*	
orange fur	all orange fur	X^OX^O *or* X^OY	*O* is X-linked
	tortoise shell variegation	X^OX^B	
	no orange fur (often black)	X^BX^B *or* X^BY	
dilute pigmentation	pigmentation is intense	*Dd* or *dd*	*D* is completely dominant
	pigmentation is dilute (e.g. gray rather than black; cream rather than orange; light brown rather than brown)	*dd*	
tabby	tabby pattern	*AA* or *Aa*	This is a simplification of the tabby phenotype, which involves multiple genes
	solid coloration	*aa*	
sex	female	XX	
	male	XY	

References

1. *Adapted from Christensen (2000) Genetics 155:999-1004*)

This page titled 5.3: Example of Multiple Genes Affecting One Character is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.3: Example of Multiple Genes Affecting One Character** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

SECTION OVERVIEW

5.4: Alleles at a Single Locus

5.4.1: Introduction

5.4.2: Somatic vs. Germline Mutations

5.4.3: Alleles- Hetero-, Homo-, Hemizygosity

5.4.4: Pleiotropy vs. Polygenic Inheritance

5.4.5: Types of Dominance

5.4.6: Biochemical Basis of Dominance

5.4.7: Classification of Mutants

5.4.8: Muller's Morphs

5.4.9: Chapter 6 Summary

5.4.10: Key Terms in Chapter 6

5.4.11: Chapter 6 Study Questions

5.4.12: Chapter 6 Answers

This page titled 5.4: Alleles at a Single Locus is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

5.4.1: Introduction

6.1 Introduction

Learning Objectives

- Discuss what is meant by “Extensions and Modifications of Basic Principles of Mendelian Genetics”.
- Identify additional factors at a single locus and the effect on the results of genetic crosses.
- Describe the various types of dominance, lethal alleles and pleiotropic genes.
- State the molecular conditions required to produce “Muller’s Morphs”.

Mendelian inheritance describes inheritance patterns that obey two laws — the Law of Segregation and the Law of Independent Assortment. Simple Mendelian inheritance involves a single gene with two different alleles, and alleles which display a simple dominant/recessive relationship. We will examine traits that deviate from the simple dominant/recessive relationship — the inheritance patterns of these traits still obey Mendelian laws, however, they are more complex and interesting than Mendel had realized. In this chapter, we will investigate the interactions of alleles at a single locus. We will begin with the difference between somatic and germ line mutations, followed by the concept of Pleiotrophy. Then, we will look at the various types of dominance, followed by the Biochemical basis of dominance. Finally, we will end with more sophisticated interactions that can be described by “Muller’s Morphs”, which deal with the interrelationships of mutant and wild type alleles at a more detailed level.

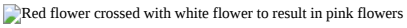


Figure 6.1.1 Incomplete Dominance [Long description]

A specific section of a chromosome is called a **locus**. Because each gene occupies a specific locus along a chromosome, the terms locus and gene are often used interchangeably. However, the term “gene” is a much more general term, while “locus” usually is limited to defining the position along a chromosome. Each locus will have an allelic form (allele); that is, a specific DNA sequence. In a population of individuals, there will be sequence variation so there will be different alleles. Some may be defined as wild type, some as variants, others as mutant. The complete set of alleles at all loci in an individual is its **genotype**. Typically, when writing out a genotype, only the alleles at the locus (or loci) of interest are considered and written down — all the others are still present and assumed to be wild type. So, usually only the alleles at the few mutant loci appear in the written genotype. All the many, many others that are wild type, are not. The visible or detectable effect of alleles on the structure or function of that individual is called its **phenotype** — what it looks like. The phenotype studied in any particular genetic experiment may range from simple, visible traits, such as hair color, to more complex phenotypes including disease susceptibility or behaviour. If two alleles are present in an individual, as is the case with diploid organisms, then various interactions between them may influence their expression in the phenotype.

Media Attribution

- **Figure 6.1.1** Incomplete dominance by Spencerbaron,CC BY-SA 3.0, via Wikimedia Commons

Long Description

- **Figure 6.1.1** Incomplete dominance: A red-coloured flower is shown to be mated with a white-coloured flower. All offspring of this cross are pink flowered, indicating the production of the “intermediate” phenotype typically exemplified in incomplete dominance. [Back to Figure 6.1.1]

This page titled 5.4.1: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.1: Introduction** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.2: Somatic vs. Germline Mutations

6.2 Somatic vs. Germline Mutations

A mutation occurs in the DNA of a single cell. In single-cell organisms, that mutation is passed on directly to its descendants, typically through the process of mitosis. In multicellular animals, there is a partitioning early in development into **somatic cells**, which form the body cells, and **germline cells**, which form the gametes for the next generation. Mutations may be passed on to somatic cells via mitosis and to gametes via meiosis. In plants, this somatic/germline separation occurs later in the cells that form the flower.

Somatic Mutations

Somatic cells form the tissues of the organism and are not passed on as gametes. Any mutations in somatic cells will only affect the individual in which they occur, not its progeny. If mutations occur in somatic cells, its mutant descendants will exist alongside other non-mutant (wild type) cells. If the mutation occurs at a very early stage of development, the mutation will be present in more cells. This gives rise to an individual composed of two or more types of cells that differ in their genetic composition. Such an individual is said to be a **mosaic**. An example is shown in **Figure 6.2.1** Cancer cells are another example of mosaicism.



Figure 6.2.1 Segmental Heterochromia in Left Eye (shown right). [Long description]

Germline Mutations

Germline cells are those that form the eggs or sperm cells (ovum or pollen in plants), and are passed on to form the next generation. Therefore, mutations in germline cells will be passed on to the next generation but won't affect the individual in which they occur.

In animals, somatic cells are segregated from germ line cells. In plants, somatic cells become germline cells; so somatic mutations can become germline mutations.

Take a look at the video below, *2H – Somatic & germline mutations*, presented by Professor Redfield of UBC (Useful Genetics, 2015) on YouTube, which explains Somatic and Germline mutations.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=670

Haploid vs. Diploid Organisms

Haploid organisms, have only one copy of a gene, thus a mutation will directly affect the organism's phenotype. Therefore, the phenotype can be used to directly infer the genotype of the organism.

However, in diploid organisms, there are two copies of each gene. The phenotype depends upon an interaction between the two alleles. Thus, any mutation may not have a direct impact on the organism's phenotype. The interaction of the two alleles can show complete dominance, incomplete dominance, co-dominance, or recessiveness. Therefore, inferring the genotype based upon its phenotype is not as simple as in diploids.

Watch the video below, *The Life Cycle of Yeast – Professor Rhona Borts*, from University of Leicester (2010) on YouTube.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=670

Media Attribution

- **Figure 6.2.1** Segmental heterochromia in left eye by Jvedral, CC BY-SA 4.0, via Wikimedia Commons

References

Redfield, R./UBC [Useful Genetics]. (2015, August 10). 2H – Somatic & germline mutations (video file). YouTube. https://www.youtube.com/watch?v=bSPBow616f8

University of Leicester. (2010, November 19). The life cycle of yeast – Professor Rhona Borts (video file). YouTube. https://youtu.be/XxDlzMARURE

Long Description

- **Figure 6.1.1** A pair of human eyes. Left eye colour is blue-grey and right eye colour is shown to be mosaic, with two colours shown – blue-grey and orange. This is an example of segmental heterochromia. [Back to Figure 6.2.1]

This page titled 5.4.2: Somatic vs. Germline Mutations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.2: Somatic vs. Germline Mutations** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

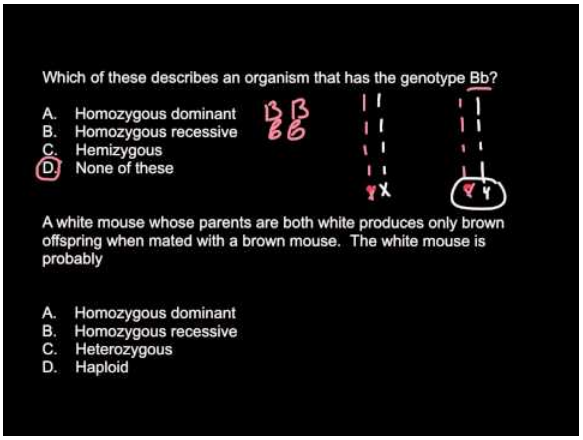
5.4.3: Alleles- Hetero-, Homo-, Hemizygosity

6.3 Alleles: Hetero-, Homo-, Hemizygosity

Mendel's First Law (segregation of alleles) is especially remarkable because he made his observations and conclusions (1865) without knowing about the relationships between genes, chromosomes, and DNA. We now know the reason why more than one allele of a gene can be present in an individual; most eukaryotic organisms are diploid and have at least two sets of homologous chromosomes. For organisms that are predominantly diploid, such as humans or Mendel's peas, chromosomes exist as pairs, with one copy inherited from each parent. Diploid cells, therefore, can contain two different alleles of each gene, with one allele part of each member of a pair of homologous chromosomes. If both alleles of a particular gene are the same (indistinguishable), the individual is said to be **homozygous** at that gene or locus. On the other hand, if the alleles are different (can be distinguished) from each other, the genotype is **heterozygous**. In cases where there is only one copy of a gene present, for example if there is a deletion of the locus on the homologous chromosome, we use the term **hemizygous**. Another example is the single X-chromosome in X/Y males, where almost all the loci on that chromosome are hemizygous.

Although a single diploid individual can have at most two different alleles of a particular gene, many more alleles can exist in a population of individuals. In a natural population the most common allelic form is usually called the **wildtype** allele. However, in many populations there can be multiple **variants** at the DNA sequence level that are visibly indistinguishable as all exhibit a normal, wild type appearance. There can also be various **mutant** alleles (in wild populations and in lab strains) that vary from wild type in their appearance, each with a different change at the DNA sequence level. The many different mutations (alleles) at the same locus are called an **allelic series** for a locus.

Take a look at the video below, *Homozygous*, *Heterozygous*, *Hemizygous*, *Haploid*, by Nikolay's Genetics Lessons (2015) on YouTube, which discusses the terms homozygous, heterozygous, hemizygous and haploid.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=676

Reference

Nikolay's Genetics Lessons. (2015, September 9). Homozygous, heterozygous, hemizygous, haploid (video file). YouTube. https://www.youtube.com/watch?v=tUV90q6pzOU

This page titled 5.4.3: Alleles- Hetero-, Homo-, Hemizygosity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.3: Alleles- Hetero-, Homo-, Hemizygosity** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.4: Pleiotropy vs. Polygenic Inheritance

6.4 Pleiotropy vs. Polygenic Inheritance

There is usually not a one-to-one correspondence between a gene and a physical characteristic. Often a gene is responsible for several phenotypic traits and it is said to be **pleiotropic**. Pleiotropy occurs when one gene influences two or more seemingly unrelated phenotypic traits. Such a gene that exhibits multiple phenotypic expression is called a pleiotropic gene. For example, mutations in the vestigial gene (*vg*) in Drosophila results in an easily visible short wing phenotype. However, mutations in this gene also affect the number of egg strings, position of the bristles on scutellum, and lifespan in Drosophila. Therefore, *vg* gene is said to be pleiotropic in that it affects many different phenotypic characteristics. During his study of inheritance in pea plants, Mendel made several interesting observations regarding the colour of various plant components. Specifically, Mendel noticed that plants with coloured seed coats always had coloured flowers and coloured leaf axils — axils are the parts of the plant that attach leaves to stems. Mendel also observed that pea plants with colourless seed coats always had white flowers and no pigmentation on their axils. In other words, in Mendel's pea plants, seed coat colour was always associated with specific flower and axil colours. Today, we know that Mendel's observations were the result of pleiotropy, or the phenomenon in which a single gene contributes to multiple phenotypic traits. In this case, the seed coat colour gene, denoted *a*, was not only responsible for seed coat colour, but also for flower and axil pigmentation.

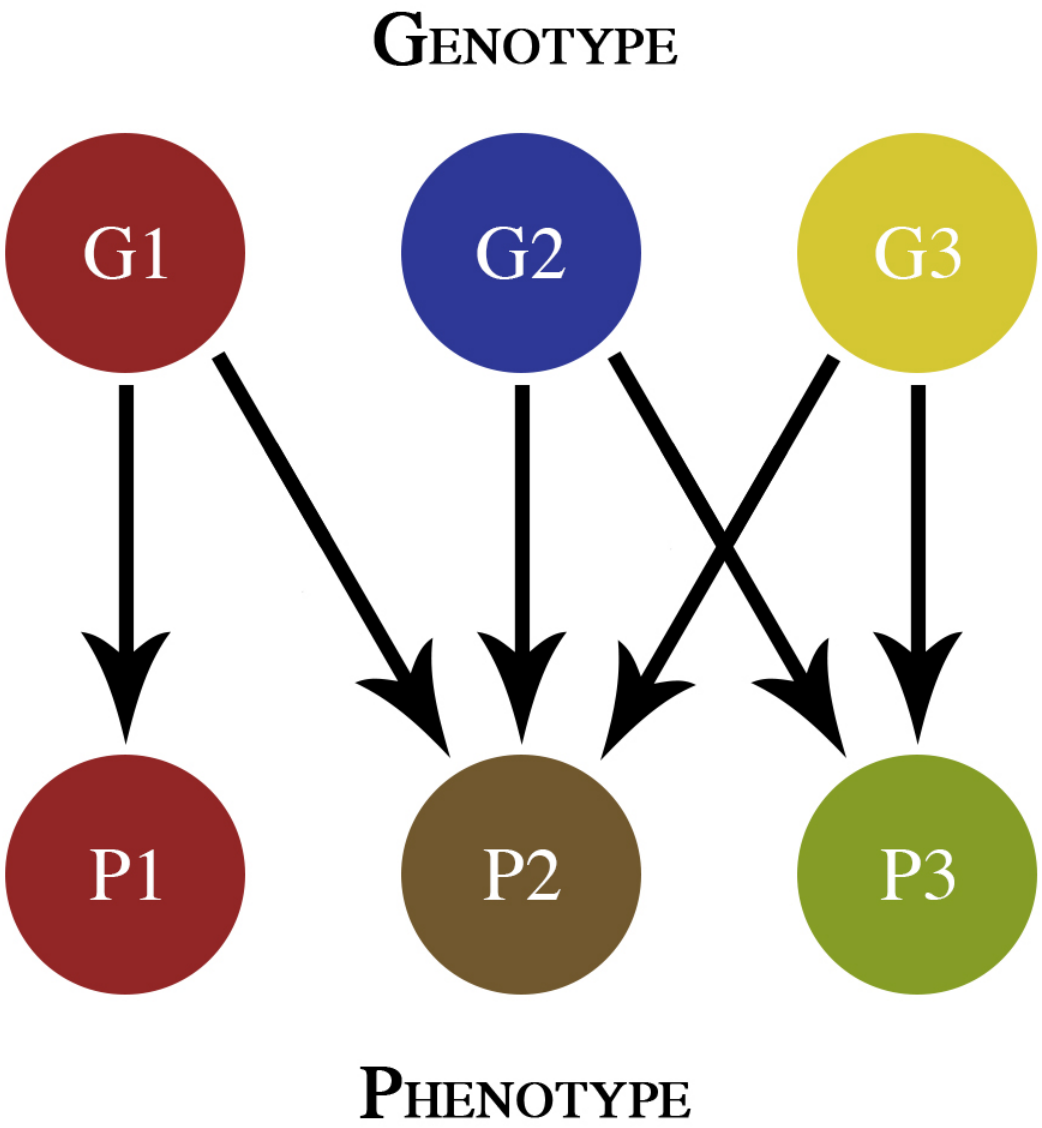


Figure 6.4.1 Possible Relationships Between Genes and Phenotypes [Long description]

On the other hand, single characteristics can be affected by mutations in multiple, different genes. This implies that many genes are needed to make each characteristic. For example, if we return to the Drosophila wing, there are dozens of genes that when mutant alter the normal shape of the wing, not just the *vg* locus. Thus there are many genes that are needed to make a normal wing; the mutation of any one causes an abnormal, mutant, phenotype. This type of arrangement is called **polygenic inheritance** .

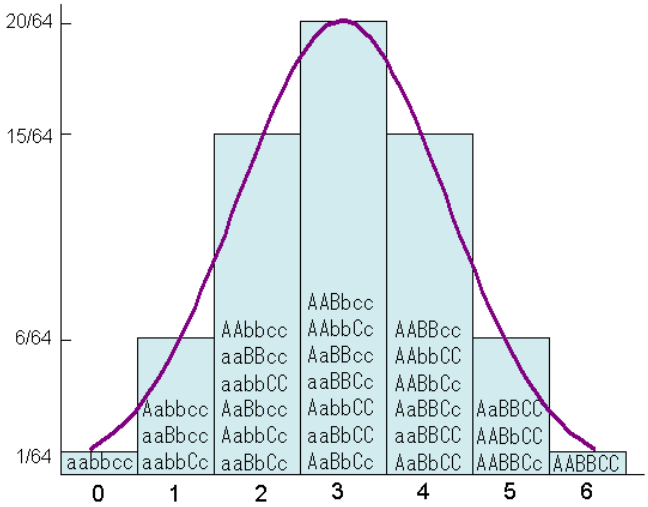


Figure 6.4.2 Typical Distribution of Phenotypes in Polygenic Inheritance. Traits that display a continuous distribution, such as height or skin color, are polygenic. [Long description]

Media Attributions

- **Figure 6.4.1** Simple Genotype Phenotype Map by Alphillips6, CC BY-SA 4.0, via Wikimedia Commons
- **Figure 6.4.2** Polygene00 by Maulits, CC BY-SA 4.0, via Wikimedia Commons

Long Descriptions

- **Figure 6.4.1** Pleiotropy occurs when one gene influences two or more seemingly unrelated phenotypic traits. This relationship between genes and phenotypes is demonstrated by mapping one genotype; e.g., G1 to multiple phenotypes; e.g.; P1 and P2. [Back to Figure 6.4.1]
- **Figure 6.4.2** A bell curve showing the typical distribution of phenotypes in Polygenic Inheritance. On the extreme left and the extreme right of the curve, a small frequency of outlier genotypes are represented. As the curve approaches the middle, the frequency of more common genotypes increases. At the very centre of the curve, a maxima is achieved, producing the typical bell shaped graph. [Back to Figure 6.4.2]

This page titled 5.4.4: Pleiotropy vs. Polygenic Inheritance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.4: Pleiotropy vs. Polygenic Inheritance** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.5: Types of Dominance

6.5 Types of Dominance

As we discussed in the previous section on polygenic traits, in humans most characteristics do not fit into two different phenotypes — complex traits, e.g., height, hair texture, skin colour etc., seemingly do not follow Mendelian analysis. As more scientists began analyzing genetic crosses using different types of plants and animals, it was found that while some traits obeyed Mendel's laws (they were determined by a single gene with 1 dominant and 1 recessive allele), many other traits did not. In such cases, there were no definite recessive or dominant traits observed, or more than two alleles identified in a particular cross. In some instances, traits seem to be determined by more than one gene (multifactorial), and the environment also seemed to play a role through interaction with genes, to produce varying phenotypes.



Figure 6.5.1 Colour, Shape, and Size of Tomatoes are Examples of Multifactorial Traits [Long description]

These examples of the behaviour of certain traits implies a more complex array of interactions are at play, as these do not generate the typical Mendelian phenotypic ratios. We are extending Mendel's Laws in order to provide explanations for the behaviour of such traits, and not necessarily challenging them.

One of the first concepts we need to understand, is that dominance is not always complete. Thus far, we have looked at the concept of dominance and recessiveness, whereby these conditions arise upon crossing two pure-breeding lines to create hybrids, and the hybrids are identical in phenotype to one parent for the particular trait in question. In this simplistic case, the allele passed down by that parent is said to be completely dominant when compared with the allele passed down by the parent whose trait is not manifested in the hybrid offspring. This type of arrangement is termed **complete dominance**.

As we will now see, there are two other types of Dominance — namely, **incomplete dominance** and **co-dominance**.

Complete Dominance

An example of a simple phenotype, is flower color in Mendel's peas. We have already said that one allele as a homozygote produces purple flowers, while the other allele as a homozygote produces white flowers. But what about a heterozygous individual that has one purple allele and one white allele? What is the phenotype of a heterozygote?

This can only be determined by experimental observation. We know from observation that individuals heterozygous for the purple and white alleles of the flower colour gene have purple flowers. Thus, the allele associated with purple colour is, therefore, said to be **dominant** to the allele that produces the

white colour. The white allele, whose phenotype is masked by the purple allele in a heterozygote, is **recessive** to the purple allele. The dominant/recessive character is a relationship between two alleles and must be determined by observation of the heterozygote phenotype.

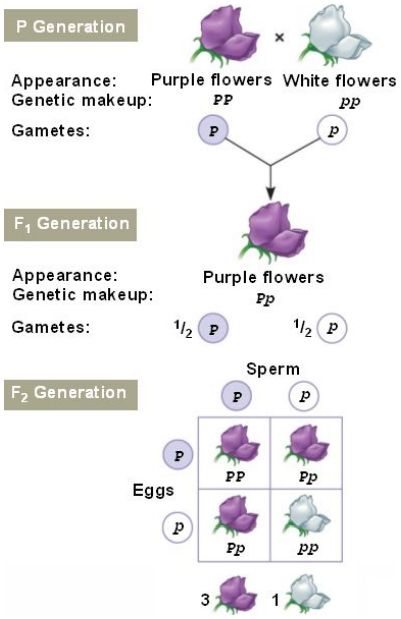


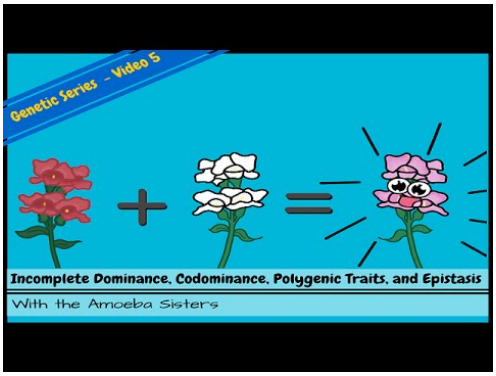
Figure 6.5.2 Allele for Purple Flowers (P) is Completely Dominant Over Allele for White Flowers (p) [Long description]

Sometimes, to represent this relationship, a dominant allele will be written as a capital letter (e.g., *A*) while a recessive allele will be written in lower case (e.g., *a*). However, this is not the only system. Many different systems of genetic symbols are in use. The most common are shown in **Table 6.5.1** Also note, genotypes (alleles) are usually written in *italics* and chromosomes and proteins are not. For example, the *white* gene in *Drosophilamelanogaster* on the X chromosome encodes a protein called WHITE, which is a pigment precursor transmembrane transporter enzyme.

Table 6.5.**1 Examples of Symbols Used to Represent Genes and Alleles**

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used because not all alleles show complete dominance and many genes have more than two alleles.
a^+ and a^1	Superscripts or subscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene locus, but on homologous chromosomes.

Take a look at the video below. *Incomplete Dominance*, *Codominance*, *Polygenic Traits*, *Epistasis*, by Amoeba Sisters (2015) on YouTube, which discusses the various types of dominance and polygenic traits.



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=684

Incomplete Dominance

Other than the complete dominant and recessive relationship, other relationships can exist between alleles. In **incomplete dominance** (also called **semi-dominance**), both alleles affect the trait additively, and the phenotype of the heterozygote shows a typically intermediate between the homozygotes, which is often referred to as blended phenotype. For example, alleles for colour in carnation flowers (and many other species) exhibit incomplete dominance. Plants with alleles for red petals (RR) when crossed with a plant with alleles for white petals (rr) have offspring which have pink petals (Rr). We say that

the R and the r alleles show incomplete dominance because neither allele is completely dominant over the other (**Figure 6.5.3**). Even though in **Figure 6.5.3,** there is the use of capital and common letters to indicate the two incompletely dominant alleles, a better way to represent such alleles would be the use of superscripts on the same letter e.g., R^1 and R^2.

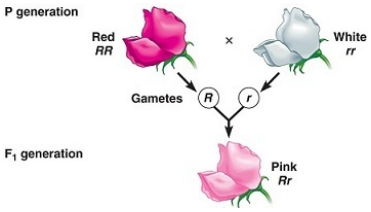


Figure 6.5.3 Incomplete Dominance (Blended Phenotype Produced) [Long description]

Co-Dominance

Co-dominance is another type of allelic relationship in which a heterozygous individual expresses the phenotype of both alleles simultaneously. An example of co-dominance is found within the **ABO blood group** of humans. The *ABO* gene has three common alleles that were named (for historical reasons) I^A, I^B, and *i*. People homozygous for I^A or I^B display only A or B type antigens, respectively, on the surface of their blood cells, and therefore, have either type A or type B blood (**Figure 6.5.4**). Heterozygous I^AI^B individuals have both A and B antigens on their cells, and so have type AB blood. Note that the heterozygote expresses both alleles simultaneously, and is not some kind of novel intermediate between A and B. Co-dominance is, therefore, distinct from incomplete dominance, although they are sometimes confused.

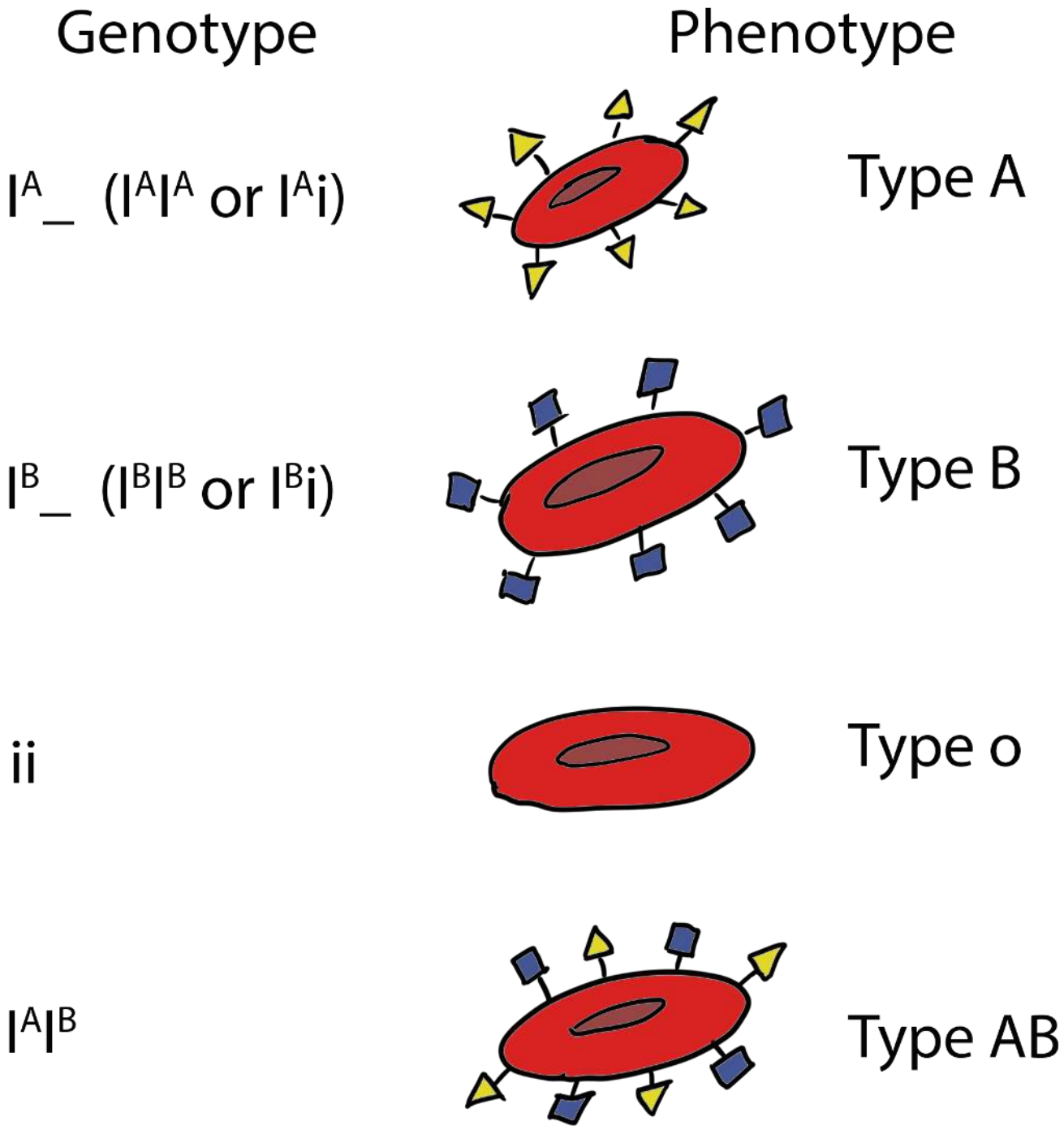


Figure 6.5.4 Co-Dominance Exhibited by Human Blood Groups [Long description]

It is also important to note that the third allele, *i* , does not make either antigen and thus is recessive to the other alleles. I^A/i *or* I^B/i individuals display only A or B antigens, respectively. People homozygous for the *i* allele have type O blood.

This is a useful reminder that different types of dominance relationships can exist, even for alleles of the same gene.

Media Attributions

- **Figure 6.5.1**2013 09 10 Tomate by Friedrich Haag, CC-BY-SA-4.0, via Wikimedia Commons
- **Figure 6.5.2** 14 05LawOfSegregation 2 L by Ashinkaaa, CC BY-SA 4.0, via Wikimedia Commons
- **Figure 6.5.3** 09 11aIncompleteDominance-L by RudLus02, CC BY-SA 4.0, via Wikimedia Commons
- **Figure 6.5.4** Blood Type Codominance by DylanAudette, CC0 1.0 Universal Public Domain, via Wikimedia Commons

References

Amoeba Sisters. (2015, May 25). Incomplete dominance, codominance, polygenic traits, and epistasis! (video file). YouTube. https://www.youtube.com/watch?v=YJHGfbW55l0

Long Descriptions

- **Figure 6.5.1** Two tomatoes: one is large and red in colour, and the other is small and orange in colour. The image seeks to demonstrate the variety of phenotypes that are possible with multifactorial traits, which are determined by more than one gene. [Back to Figure 6.5.1]
- **Figure 6.5.2** Cross to demonstrate complete dominance. P generation is pure bred purple flowers crossed with pure breeding white flowers, resulting in all purple flowers (heterozygous) in the F1 generation. The F2 generation produces flowers in a ratio of 3 purple to 1 white, which is typical of complete dominance in a monohybrid cross. [Back to Figure 6.5.2]
- **Figure 6.5.3** Cross to demonstrate incomplete dominance. P generation is pure-bred red flowers crossed with pure breeding white flowers. The F1 generation produced is comprised of pink flowers only, which is indicative of incomplete dominance between these two alleles controlling flower colour. [Back to Figure 6.5.3]
- **Figure 6.5.4** The variety of blood types in humans. Four phenotypes are shown which are A, B, O and AB. These phenotypes are the result of combinations of alleles which exemplify co-dominance (A and B) as well as alleles which exemplify complete dominance (A and B over O). The combinations of alleles result on specific antigens being expressed on red blood cells of that organism, resulting in the four typical blood group phenotypes seen in humans. [Back to Figure 6.5.4]

This page titled 5.4.5: Types of Dominance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.5: Types of Dominance** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.6: Biochemical Basis of Dominance

6.6 Biochemical Basis of Dominance

Given that a heterozygote's phenotype cannot simply be predicted from the phenotype of homozygotes, what does the type of dominance tell us about the biochemical nature of the gene product? How does dominance work at the biochemical level? There are several different biochemical mechanisms that may make one allele dominant to another.

For the majority of genes studied, the normal (i.e., wild-type) alleles are **haplo-sufficient**. So in diploids, even with a mutation that causes a complete loss of function in one allele, the other allele — a wild-type allele — will provide sufficient normal biochemical activity to yield a wild type phenotype and thus be dominant and dictate the heterozygote phenotype.

On the other hand, in some biochemical pathways, a single wild-type allele is not enough protein and may be **haplo-insufficient** to produce enough biochemical activity to result in a normal phenotype, when heterozygous with a non-functioning mutant allele. In this case, the non-functional mutant allele will be dominant (or semi-dominant) to a wild-type allele.

Mutant alleles may also encode products that have new and/or different biochemical activities instead of, or in addition to, the normal ones. These **novel activities** could cause a new phenotype that would be dominantly expressed.

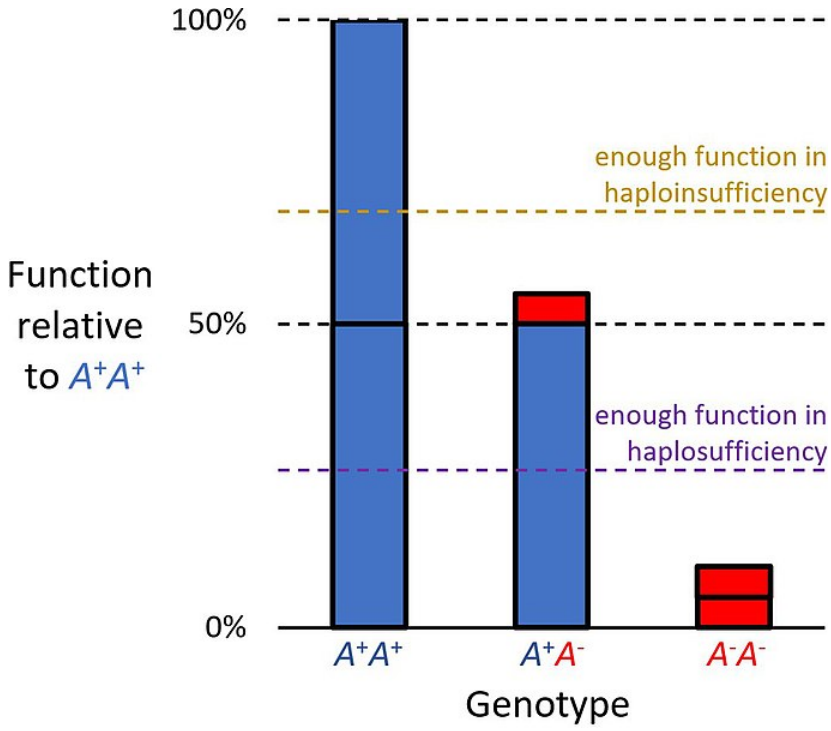


Figure 6.6.1 Graph Showing the Haploinsufficiency Model of Dominant Genetic Disorders. A+ is a normal allele. A- is a mutant allele with little or no function. In haplosufficiency (most genes), a single normal allele provides enough function, so A+A- individuals are healthy. In haploinsufficiency, a single normal allele does not provide enough function, so A+A- individuals have a genetic disorder. [Long description]

Media Attribution

- **Figure 6.6.1**Haploinsufficiency graph only by Adrian J. Hunter, CC BY-SA 4.0, via Wikimedia Commons

Long Description

- **Figure 6.6.1** The haploinsufficiency model of dominant genetic disorders: A+ is a normal allele and A- is a mutant allele with little or no function. In haplosufficiency (most genes), a single normal allele provides enough function, so A+A- individuals are healthy. In haploinsufficiency, a single normal allele does not provide enough function, so A+A- individuals have a genetic disorder. [Back to Figure 6.6.1]

This page titled 5.4.6: Biochemical Basis of Dominance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.6: Biochemical Basis of Dominance** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.7: Classification of Mutants

6.7 Classification of Mutants

Previously, we looked at complementation groups and we understood how mutations can work together (or not) to produce different phenotypes. In this section, we will look at the various types of mutations that can arise in cells: morphological, lethal, biochemical and conditional. The final type, known as Muller's Morphs, will be discussed in the next section.

Morphological Mutants

Morphological mutations cause changes in the visible form of the organism as they give rise to altered forms of a trait e.g., change in size, shape (normal wing vs. curly wing in fruit flies), colour, number etc.



Figure 6.7.1 Examples of Morphological Mutations in Dogs [Long description]

Lethal Mutants

A lethal mutation causes the premature death of an organism. For example, in Drosophila, lethal mutations can result in death during the embryonic, larval, or pupal stage. Lethal mutations are usually recessive, so both copies of a gene must be lost for premature death to occur (homozygous lethal alleles will not be viable). Heterozygotes, which have one lethal allele and one wild type allele, are typically viable. In the example shown in **Figure 6.7.2** regarding yellow coat colour in mice, the lethal allele is recessive because it causes death only in homozygotes. Unlike its effect on survival, the effect of the allele on colour is dominant. In mice, a single copy of the allele in heterozygotes produces a yellow colour. These examples illustrate the point that the type of dominance depends on the aspect of the phenotype examined.

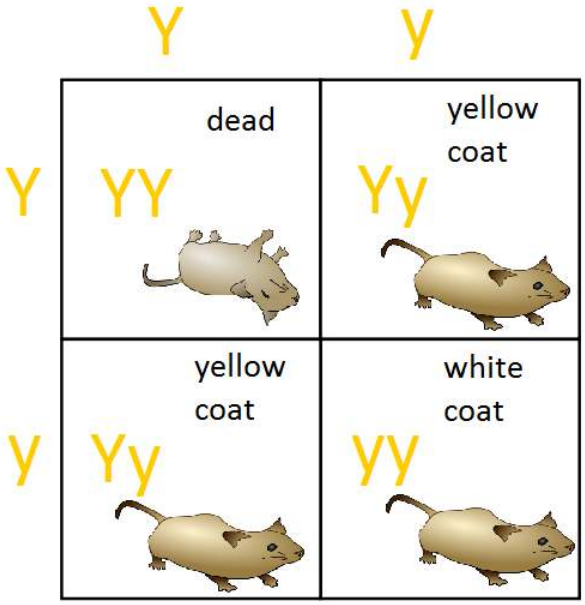


Figure 6.7.2 Punnett Square Showing Effect of a Lethal Allele in Mice [Long description]

Biochemical Mutants

Auxotrophic mutants can be derived from prototrophic parents. This type of mutation blocks a step in a biochemical pathway for the *arg-* mutants of Beadle and Tatum. Such biochemical mutations are a specific type of the conditional mutation class. Biochemical mutants result in the inability to carry out a specific biochemical pathway.

Growth Data

Mutant	Growth of mutant on minimal medium supplemented with:				
	Nothing	Ornithine	Citrulline	Arginino-succinate	Arginine
Arg^{+}	+	+	+	+	+
ArgE^{-}	-	+	+	+	+
ArgF^{-}	-	-	+	+	+
ArgG^{-}	-	-	-	+	+
ArgH^{-}	-	-	-	-	+

+ fungus grows on this medium
- fungus fails to grow on this medium

Figure 6.7.3 Data Derived by Beadle and Tatum on Biochemical Mutants [Long description]

Conditional Mutants

Conditional mutations rely on the concept of: phenotype = genotype + environment + interaction. Organisms with this mutation express a mutant phenotype, but only under specific environmental conditions. Under **restrictive conditions**, they express the

mutant phenotype, while under **permissive conditions**, they show a wild type phenotype. One example of a conditional mutation is the temperature-sensitive pigmentation of Siamese cats. Siamese cats have temperature sensitive fur colour; their fur appears unpigmented (light coloured) when grown in a, warm temperature environment. The hair appears pigmented (dark) when grown at a cooler temperature. This is seen at the peripheral regions of the feet, snout, and ears (**Figure 6.7.4**). This is because in warm temperatures, the enzyme needed for melanin pigment synthesis becomes nonfunctional. However, in cooler temperatures, the enzyme needed for melanin synthesis is functional, and the deposition of melanin makes the fur look dark.



Figure 6.7.4 Siamese Cats Showing Temperature Sensitive Fur Colourn [Long description]

Media Attributions

- **Figure 6.7.1** Dog morphological variation by Mary Bloom, American Kennel Club, CC BY-SA 4.0, via Wikimedia Commons
- **Figure 6.7.2** Lethal alleles punnett squareby Dead_mouse.svg and Mouse.svg: Madprime derivative work: Adabow, CC BY-SA 3.0, via Wikimedia Commons
- **Figure 6.7.3** Beadle and Tatum data by Allen Gathman is licensed under CC BY-NC-SA 2.0, via Flickr
- **Figure 6.7.4** Two Siamese Cats by Steve Gilham is licensed under CC BY-NC-SA 2.0, via Flickr

Long Descriptions

- **Figure 6.7.1** Six different dogs, each of different species, display a wide range of physical attributes that point to morphological mutations, such as differences in fur colour and texture, tail length, height of limbs, facial features, ear presentation, etc. [Back to Figure 6.7.1]
- **Figure 6.7.2 A** Punnett square shows the production of offspring from two parents, both of which carry the allele dominant for coat colour but recessive for lethality. Both parents are genotype Yy; both display a yellow coat colour. In the offspring, there is a 3:1 ratio of living to dead progeny, with the accumulation of the YY alleles in one-out-of-four offspring results in death. Yy in the offspring (50%) display yellow coat colour and are alive, and yy offspring (25%) display white coat colour, which is recessive to yellow, and are alive. [Back to Figure 6.7.2]
- **Figure 6.7.3** Growth data demonstrated for various mutants and corresponding minimal media alone or media supplemented with various nutrients including nothing, ornithine, citrulline, arginosuccinate and arginine. The ability of the various mutants to grow in different types of media is indicated by a plus sign, and no growth is represented by a negative sign. This information represents the data derived by Beadle and Tatum on biochemical mutants. [Back to Figure 6.7.3]
- **Figure 6.7.4** Two Siamese cats display temperature sensitive colour, which is an example of conditional mutations, whereby the environment dictates the phenotype expressed. Their fur appears unpigmented (light coloured) when grown in a warm temperature environment. The fur appears pigmented (dark) when grown at a cooler temperature. This is seen at the peripheral regions of the feet, snout, and ears. [Back to Figure 6.7.4]

This page titled 5.4.7: Classification of Mutants is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.7: Classification of Mutants** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.8: Muller's Morphs

6.8 Muller's Morphs

In this section, we will discuss the final class of mutants, called Muller's Morphs. Most DNA sequence changes (mutations) occur at essentially random locations along a chromosome. Of those mutations occurring within genes, their mutant phenotypes (often recovered through genetic screens) are caused by **loss-of-functionalleles**. These alleles are due to sequence changes in the DNA that cause a gene to produce fewer, less active, or non-active product (typically a protein), compared to the wild-type allele. Loss-of-function alleles tend to be recessive because the wildtype allele is haplo-sufficient. A loss-of-function allele that produces no active product is called an **amorphs**, or **null**, while alleles with only a partial loss-of-function are called **hypomorphs**.

More rarely, a mutant may have a **gain-of-function allele**, producing either more of the active product (**hypermorphs**) or producing an active product with a new and different function (**neomorphs**). Finally, **antimorphs** alleles have an activity that is dominant and opposite to the wild-type product's function; antimorphs are also known as **dominant negative** mutations.

Thus, mutations (changes in a gene sequence) can result in mutant alleles that no longer produce the same level or type of active product as the wild-type allele. Any mutant allele can be classified into one of five types: (1) amorph, (2) hypomorph, (3) hypermorph, (4) neomorph, and (5) antimorph.

Amorph

Amorphic alleles have a complete loss-of-function. They make no active product — zero function. They are known as a "**Null**" mutation or a "loss-of-function" mutation.

Molecular explanation: Changes in the DNA base pair sequence of an amorphic allele may cause one or more of the following:

1. Gene deletion — The DNA sequence is removed from the chromosome.
2. Gene is present, but is not transcribed because of a sequence change in the promoter or enhancer/regulatory elements.
3. Gene is present, but the transcript is aberrantly processed. There is normal transcription but base pair changes cause the mature mRNA to incorrectly splice introns, therefore the translated amino acid sequence would be altered and nonfunctional.
4. Gene is present and a transcript is produced but no translation occurs — changes in the base pair sequences would preclude the mRNA from binding to the ribosome for translation.
5. Gene is present and a transcript is produced and translated but a nonfunctional protein product is produced — the mutation alters a key amino acid in the polypeptide sequence producing a completely non-functional polypeptide.

Genetic/phenotypic explanation: Amorphic mutations of most genes usually act as recessive to wild type (Case #1). However, with some genes the amorphic mutations are dominant to wild type (Case #2).

Cases 1 & 2

case #1: *white* gene in Drosophila

Allele Combination	Result
w^+/w^+	wildtype and red eyed
w^+/w^-	wildtype and red eyed
w^-/w^-	mutant and white eyed

case #2: *Minute* locus in Drosophila

Allele Combination	Result
M^+/M^+	wildtype and long bristled
M^+/M^-	mutant and short bristled
M^-/M^-	dead, recessive lethal

For the *Minute* gene, we concluded that the organism needs both copies to have a wild type phenotype. Loss of one copy (an amorphic mutation) produces a dominant visible mutant phenotype. Deletion of the gene is an example of a classic amorphic mutation.

Hypomorph

Hypomorphic alleles show only a partial loss-of-function. These alleles are sometimes referred to as "**leaky**" mutations, because they provide some function, but not complete, normal function.

Molecular explanation: Changes in the DNA base pair sequence of the hypomorphic allele may cause one or more of the following, with the gene still present.

1. Reduced transcription — changed DNA sequence in the promoter or enhancer/regulatory elements can reduce the level of transcription.
2. Aberrant processing of the transcript — normal transcription but base pair changes cause the mature mRNA to incorrectly splice introns, therefore the translated protein sequence would be altered and function at a reduced level.
3. Reduced translation — changes in the base pair sequences would reduce the efficiency of the mRNA binding to the ribosome for translation.
4. Reduced-function protein product — normal transcription, processing, and translation, but mutation changes certain amino acid in the polypeptide sequence so its function is reduced.

Genetic/phenotypic explanation: Hypomorphic mutations of most genes usually act as recessive to wild type, though hypomorphic mutations theoretically could be dominant to wildtype.

Hypomorph

***white-apricot* allele in Drosophilia**

Allele Combination	Result
w^+/w^+	wildtype ad red eyed
w^+/w^a	wildtype and red eyed
w^a/w^a	mutant and apricot eye colour

Both amorphs and hypomorphs tend to be recessive to wild type in diploids because the wild type allele is usually able to supply sufficient product to produce a wild type phenotype (called **haplo-sufficient**). If the mutant allele is not able to produce a wild type phenotype, it is **haplo-insufficient**, and it will be dominant to the wild type allele. Here -/+ heterozygotes produce a mutant phenotype.

While the first two classes involve a **loss-of-function**, the next two involve a **gain-of-function** — quantity or quality. Gain-of-function alleles are almost always dominant to the wild type allele.

Hypermorph

Hypermorphic alleles produce quantitatively more of the same, active product.

Molecular explanation: Changes in the DNA base pair sequence of the hypermorphic allele may cause one or more of the following, with the gene still present.

1. Increased transcription — changed DNA sequence in the promoter or enhancer/regulatory elements that increase the level of transcription.
2. Increased translation — changes in the base pair sequences would increase the efficiency of the mRNA binding to the ribosome for translation.
3. Increased function protein product — normal transcription, processing, translation, but base pair changes alter certain amino acids in the polypeptide sequence, so its function is normal but increased in amount.

Genetic/phenotypic explanation: Hypermorphic mutations of most genes usually act as dominant to wild type since they are a gain of function, The classic hypermorph is a gene duplication.

Neomorph

Neomorphic alleles produce a product with a new, different function, something that the wild type allele does not do.

Molecular explanation: Changes in the DNA base pair sequence of the neomorphic allele may cause one or more of the following, with the gene still being present.

1. New transcription — changed DNA sequence in the promoter or enhancer/regulatory elements that makes new transcription either temporally or in a tissue-specific manner.
2. New function protein product — normal transcription, processing, translation but base pair changes alter certain amino acids in the polypeptide sequence so it acquires a new function (activity) that is different from the normal function (e.g., additional substrate or new binding site).

Genetic/phenotypic explanation: Most neomorphic mutations act dominant to wild type since they are a gain-of-function. The classical neomorphic mutation is a translocation that moves a new regulatory element next to a gene promoter so it is expressed in a new tissue or at a new time during development. Such mutations are often produced when chromosome breaks are rejoined and the regulatory sequences of one gene are juxtaposed next to the transcriptional unit of another, creating a novel, chimeric gene.

Antimorph

Antimorphic alleles are relatively rare, and have a new activity that is dominant and opposite to the wildtype function. These alleles usually interfere with the function from the wild type allele. (They often lose their normal function as well.) The new function works against the normal expression of the wild type allele. This can happen at the transcriptional, translational, or later level of expression. Thus, when an antimorphic allele is heterozygous with wild type, the wild type allele function is reduced or prevented. At the molecular level, there are many ways this can happen. The simplest model to explain an antimorphic effect is that the protein acts as a dimer (or any multimer) and the inclusion of a mutant subunit poisons the whole complex, thereby preventing or reducing its level of function. Antimorphs are also known as **dominant-negative** mutations because they are usually dominant and act negatively against the wild type function.

Identifying Muller's Morphs

All mutations can be sorted into one of the five morphs based on how they behave with heterozygous with three other standard alleles (**Figure 6.8.1**): (1) deletion alleles (zero function), (2) wild type alleles (normal function), and (3) duplication alleles (double normal function).

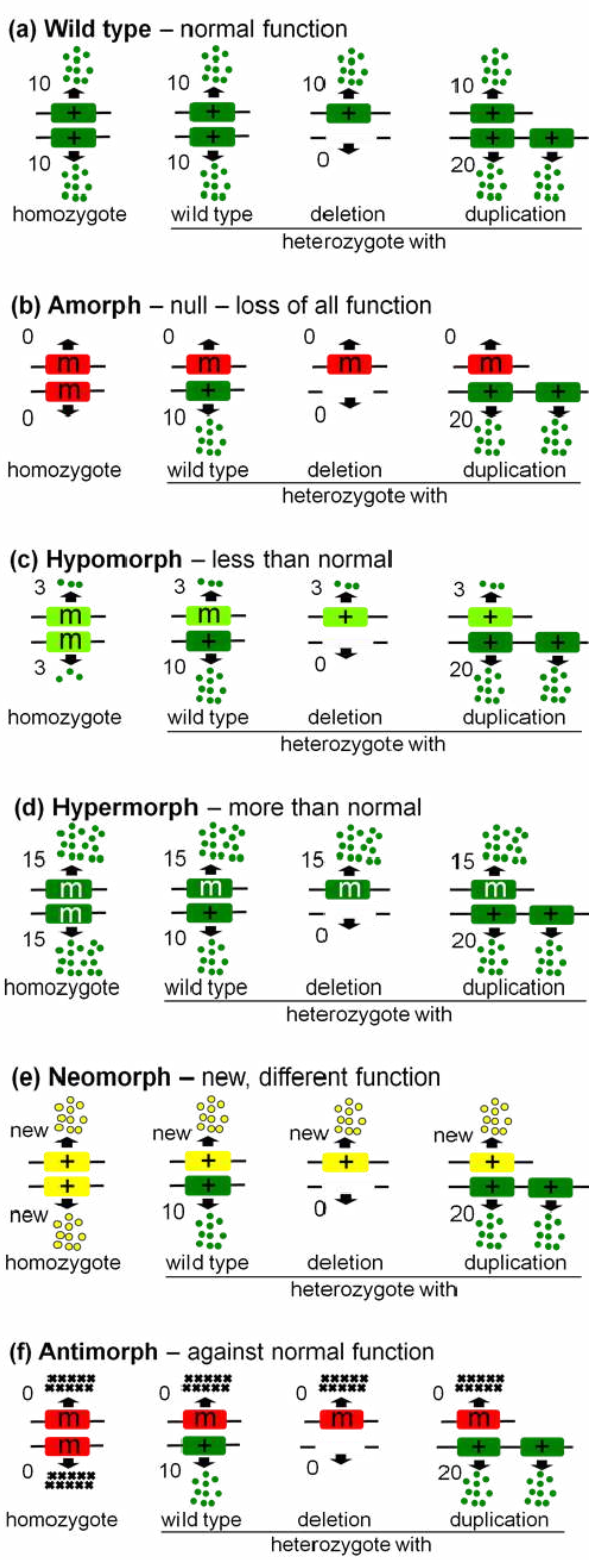


Figure 6.8.1 (a-f) Five Classes of Mutants Designated as Morphs (Forms) by a Nobel Prize Winner, H.J. Muller, Known as Muller's Morphs. [Long description]

Loss-of-Function

- Haplo-insufficient — a loss-of-function dominant mutation
- Null mutations — loss-of-function mutation that remove the activity of the gene product

- Hypomorphic mutations — decreased activity of the gene product — loss-of-function mutant phenotypes can be due to partial or complete elimination of the activity of a gene's encoded product

Gain-of-function

- Hypermorph — mutation that produces more gene activity per gene dose than wild-type
- Neomorph — mutation that produces some novel gene activity that is not characteristic of the wild-type

Media Attributions

- **Figure 6.8.1** Original by Locke (2017), CC BY-NC 3.0, Open Genetics Lectures

References

Locke, J. (2017). Figure 7. Five classes of mutants designated as morphs [5 digital images]. In Locke, J., Harrington, M., Canham, L. and Min Ku Kang (Eds.), *Open Genetics Lectures, Fall 2017* (Chapter 13, p. 8). Dataverse/ BCcampus. http://solr.bccampus.ca:8001/bcc/file/7a7b00f9-fb56-4c49-81a9-cfa3ad80e6d8/1/OpenGeneticsLectures_Fall2017.pdf

Long Description

- **Figure 6.8.1** (a-f) Comparison of the wild type expresses normal function with the five types of mutants known as Muller's Morphs. The Amorph shows a loss of all function; the hypomorph shows less than normal function; the hypermorph shows more than normal function; the neomorph shows a new different function, and the antimorph produces a product which works against normal function. [Back to Figure 6.8.1]

This page titled 5.4.8: Muller's Morphs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.8: Muller's Morphs** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.9: Chapter 6 Summary

Chapter 6 Summary

The topics covered in this chapter can be summarized as follows:

- Symbols are used to denote the alleles, or genotype, of a locus.
- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- A somatic mutation affects the individual but not the progeny, whereas a germline mutation affects the progeny in the next generation but not the individual in which they occur.
- In a diploid organism, alleles can be homozygous, heterozygous, or hemizygous.
- Allelic interactions at a locus can be described as dominant vs. recessive, incomplete dominance, or co-dominance.
- Muller's morphs classify all types of mutations including: amorph, hypomorph, hypermorph, neomorph, and antimorph.

This page titled 5.4.9: Chapter 6 Summary is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.9: Chapter 6 Summary** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.10: Key Terms in Chapter 6

Key Terms in Chapter 6

Key Terms – Alleles at a Single Locus

- ABO blood group
- amorph
- co-dominance
- complete dominance
- dominant
- gain-of-function
- genotype
- haploinsufficiency
- haplosufficiency
- hemizygous
- heterozygous
- homozygous
- hypermorph
- hypomorph
- incomplete (semi) dominance
- locus
- loss-of-function
- neomorph
- null
- phenotype
- recessive
- variant
- wild-type

This page titled 5.4.10: Key Terms in Chapter 6 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.10: Key Terms in Chapter 6** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.11: Chapter 6 Study Questions

Chapter 6 Study Questions

1. Distinguish among the following terms: (1) gene, (2) locus, (3) allele, (4) transcription unit.
2. A flower geneticist crosses a red flowered diploid plant with a white flower diploid plant and all the progeny are red. Use two different forms of symbols to show this cross and its progeny. What if all the progeny were pink?
3. If your blood type is B, what are the possible genotypes of your parents at the locus that controls the ABO blood types?
4. In the table below, match the mouse hair colour phenotypes with the term from the list that best explains the observed phenotype, given the genotypes shown. In this case, the allele symbols do not imply anything about the dominance relationships between the alleles.
   List of terms: haplo-sufficiency, haplo-insufficiency, pleiotropy, incomplete dominance, co-dominance, incomplete penetrance, broad (variable) expressivity.

Case	A1A1	A1A2	A2A2
1	all hairs black	on the same individual: 50% of hairs are all black and 50% of hairs are all white	all hairs white
2	all hairs black	all hairs are the same shade of grey	all hairs white
3	all hairs black	all hairs black	50% of individuals have all white hairs and 50% of individuals have all black hairs
4	all hairs black	all hairs black	mice have no hair
5	all hairs black	all hairs white	all hairs white
6	all hairs black	all hairs black	all hairs white
7	all hairs black	all hairs black	hairs are a wide range of shades of grey

5. In this hypothetical example of Drosophila bristle mutations, when various, true-breeding mutant strains (all at a single locus) are crossed to a wild type strain the following phenotypes are observed in the progeny:
   - Mutant#1 = bristles 20% shorter
   - Mutant#2 = bristles 30% longer
   - Mutant#3 = bristles 50% shorter
   - Mutant#4 = bristles kinked and misshapen
   - Mutant#5 = bristles are missing

   What is the best characterization, using Muller's Morphs, for each?

This page titled 5.4.11: Chapter 6 Study Questions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.11: Chapter 6 Study Questions** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.4.12: Chapter 6 Answers

Chapter 6 Answers

1. These are four common terms that are often used interchangeably by novice students, but do have distinctly different meanings and uses.
   i. gene = general term for a segment of nucleic acid that is responsible for one or more phenotypes
   ii. locus = the position of a gene along a chromosome,
   iii. allele = the form (DNA sequence) of a gene at a locus,
   iv. transcription unit = the segment of DNA that is transcribed into RNA (often mRNA in the case of a protein coding gene).
2. Form (1) RR (red) x rr (white) gives Rr (red progeny). “R” is dominant to “r”. Form (2) r+r+ (red) x r-r- (white) gives r+r- (red progeny). “r+” is dominant to “r-”. For pink progeny, the symbols are the same, only “R” or “r+” is semi-dominant to “r” or “r-”.
3. If your blood type is B, then your genotype is either I^Bi or I^BI^B. If your genotype is I^Bi, then your parents could be any combination of genotypes, as long as one parent had at least one *i* allele, and the other parent had at least one I^B If your genotype was I^B I^B, then both parents would have to have at least one I^B allele.
4. case 1 co-dominance; case 2 incomplete-dominance; case 3 incomplete penetrance; case 4 pleiotropy; case 5 haplo-sufficiency; case 6 haplo-insufficiency; case 7 broad (variable) expressivity
5. - Mutant#1 = hypomorph;
   - Mutant#2 = hypermorph;
   - Mutant#3 = amorph;
   - Mutant#4 = neomorph;
   - Mutant#5 = antimorph

This page titled 5.4.12: Chapter 6 Answers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

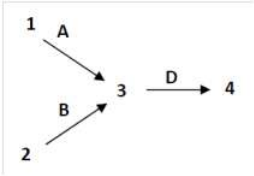
- **6.12: Chapter 6 Answers** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

5.E: Genetic Analysis of Multiple Genes (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions

Answer questions 6.1 -6.3 using the following biochemical pathway for fruit color.Assume all mutations (lower case allele symbols) are recessive, and that *either* precursor 1 or precursor 2 can be used to produce precursor 3.If the alleles for a particular gene are not listed in a genotype, you can assume that they are wild-type.



6.1 If1 and 2 and 3 are all colorless, and 4 is red, what will be the phenotypes associated with the following genotypes?

a. *aa*
b. *bb*
c. *dd*
d. *aabb*
e. *aadd*
f. *bbdd*
g. *aabbdd*
h. What will be the phenotypic ratios among the offspring of a cross *AaBb* × *AaBb*?
i. What will be the phenotypic ratios among the offspring of a cross *BbDd*× *BbDd*?
j. What will be the phenotypic ratios among the offspring of a cross *AaDd* × *AaDd*?

6.2 If1 and 2 are both colorless, and 3 is blue and 4 is red, what will be the phenotypes associated with the following genotypes?

a. *aa*
b. *bb*
c. *dd*
d. *aabb*
e. *aadd*
f. *bbdd*
g. *aabbdd*
h. What will be the phenotypic ratios among the offspring of a cross *AaBb* × *AaBb*?
i. What will be the phenotypic ratios among the offspring of a cross *BbDd*× *BbDd*?
j. What will be the phenotypic ratios among the offspring of a cross *AaDd* × *AaDd*?

6.3 If1 is colorless, 2 is yellow and 3 is blue and 4 is red, what will be the phenotypes associated with the following genotypes?

a. *aa*
b. *bb*
c. *dd*
d. *aabb*
e. *aadd*
f. *bbdd*
g. *aabbdd*
h. What will be the phenotypic ratios among the offspring of a cross *AaBb* × *AaBb*?

i. What will be the phenotypic ratios among the offspring of a cross *BbDd*× *BbDd*?
j. What will be the phenotypic ratios among the offspring of a cross *AaDd* × *AaDd*?

6.4 Which of the situations in questions 6.1 – 6.3 demonstrate epistasis?

6.5 If the genotypes written within the Punnett Square are from the F_2 generation, what would be the phenotypes and genotypes of the F_1 and P generations for:

a. Figure 6.6
b. Figure 6.8
c. Figure 6.10
d. Figure 6.12

6.6 To better understand how genes control the development of three-dimensional structures, you conducted a mutant screen in Arabidopsis and identified a recessive point mutation allele of a single gene (*g*) that causes leaves to develop as narrow tubes rather than the broad flat surfaces that develop in wild-type (*G*). Allele *g* causes a complete loss of function. Now you want to identify more genes involved in the same process. Diagram a process you could use to identify other genes that interact with gene *g*. Show all of the possible genotypes that could arise in the Fgeneration.

6.7 With reference to question 6.6, if the recessive allele, *g* is mutated again to make allele *g**, what are the possible phenotypes of a homozygous *g* g** individual?

6.8 Again, in reference to question 6.7, what are the possible phenotypes of a homozygous *aagg* individual, where *a* is a recessive allele of a second gene? In each case, also specify the phenotypic ratios that would be observed among the F_1 progeny of a cross of *AaGg* x *AaGg*

6.9 Calculate the phenotypic ratios from a dihybrid cross involving the two loci shown in Figure 6.13.There may be more than one possible set of ratios, depending on the assumptions you make about the phenotype of allele *b*.

6.10 Use the product rule to calculate the phenotypic ratios expected from a trihybrid cross.Assume independent assortment and no epistasis/gene interactions.

Chapter 6 - Answers

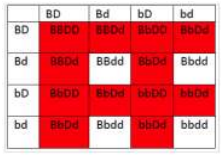
6.1 If 1 and 2 and 3 are all colorless, and 4 is red, what will be the phenotypes associated with the following genotypes? All of these mutations are recessive. As always, if the genotype for a particular gene is not listed, you can assume that alleles for that gene are wild-type.

a. red(because A and B are redundant, so products 3 and then 4 can be made)
b. red(because A and B are redundant, so products 3 and then 4 can be made)
c. white (because product3 will accumulate and it is colorless)
d. white (because only product 1 and 2 will be present and both are colorless)
e. white (because only product 1 and 3 will be present and both are colorless)
f. white (because only product 2 and 3 will be present and both are colorless)
g. white (because only product 1 and 2 will be present and both are colorless)
h. 15 red : 1 white

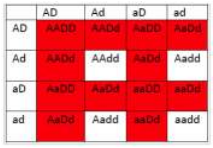


i. 12 red : 4 white

https://bio.libretexts.org/@go/page/27266



j. 12 red :4 white

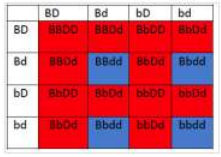


6.2

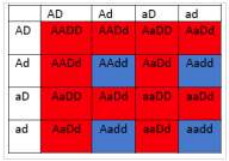
a. red(because A and B are redundant, so products 3 and then 4 can be made)
b. red(because A and B are redundant, so products 3 and then 4 can be made)
c. blue (because product 3 will accumulate, and it is blue)
d. white (because only product 1 and 2 will be present and both are colorless)
e. blue (because only product 1 and 3 will be present and 1 is colorless and 3 is blue)
f. blue(because only product 2 and 3 will be present and 2 is colorless and 3 is blue)
g. white (because only product 1 and 2 will be present and both are colorless)
h. 15 red : 1 white



i. 12 red : 4 blue



j. 12 red : 4 blue



6.3

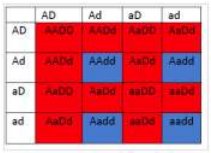
a. red(because A and B are redundant, so products 3 and then 4 can be made)
b. red(because A and B are redundant, so products 3 and then 4 can be made)
c. blue (because product 3 will accumulate, and it is blue)
d. yellow (because only product 1 and 2 will be present and 1 is colorless and 2 is yellow)
e. blue (because only product 1 and 3 will be present and 1 is colorless and 3 is blue)
f. green? (because only product 2 and 3 will be present and 2 is yellow and 3 is blue, so probably the fruit will be some combination of those two colors)
g. yellow (because only product 1 and 2 will be present and 1 is colorless and 2 is yellow)
h. 15 red : 1 yellow



i. 12 red : 3 blue:1 green



j. 12 red : 4 blue



6.4 Epistasis is demonstrated when the phenotype for a homozygous mutant in one gene is the same as the phenotype for a homozygous mutant in two genes. So, the following situations from questions 6.1-6.3 demonstrated epistasis:

6.1 No epistasis is evident from the phenotypes, even though we know from the pathway provided that D is downstream of A and B.

6.2 The phenotypes show that D is epistatic to A, because *aadd* looks like *AAdd* or *Aadd*. Also D is epistatic to B, because *bbdd* looks like *BBdd* or *Bbdd*.

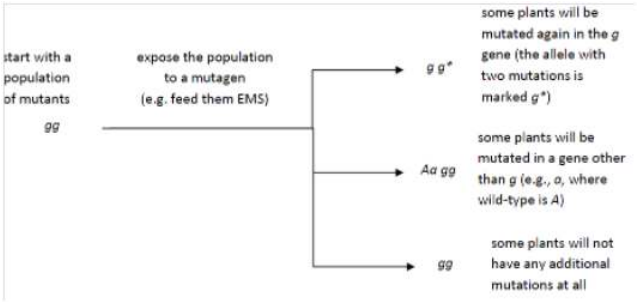
6.3 The phenotypes show that D is epistatic to A, because aadd looks like AAdd or Aadd. The phenotypes do not provide evidence for epistasis between B and D.

6.5 The answer is the same for a) – d)

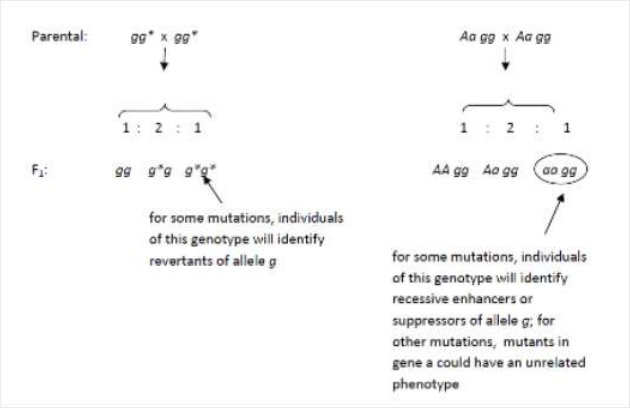
P could have been either: *AABB* x*aabb*or*aaBB* x *AAbb*;

F_1 was : *AaBb* x *AaBb*

6.6 Conduct an enhancer/suppressor screen (which can also result in the identification of revertants, as well)



allow the plants to self-pollinate in order to make any new, recessive mutations homozygous



6.7 Depending which amino acids were altered, and how they were altered, a second mutation in *g*g** could either have no effect (in which case the phenotype would be the same as *gg*), or it could possibly cause a reversion of the phenotype to wild-type, so that *g*g** and *GG* have the same phenotype.

6.8 Depending on the normal function of gene *A*, and which amino acids were altered in allele *a*, there are many potential phenotypes for *aagg*:

Case 1: If the normal function of gene *A* is in an unrelated process (e.g. *A* is required for root development, but not the development of leaves), then the phenotype of *aagg* will be: short roots and narrow leaves. The phenotypic ratios among the progeny of a dihybrid cross will be:

9	3	3	1
*A_G_*	*A_gg*	*aaG_*	*aagg*
wild-type	tubular leaves normal roots	short roots normal leaves	tubular leaves short roots

Case 2: If the normal function of gene *A* is in the same process as *G*, such that *a* is a recessive allele that **increases** the severity of the *gg* mutant (**i.e. *a* is an enhancer of *g***) then the phenotype of *aagg* could be : no leaves. The phenotypic ratios among the progeny of a dihybrid cross depend on whether *aa* mutants have a phenotype independent of *gg*, in other words, do *aaG_* plants have a phenotype that is different from wild-type or from *A_gg*. There is no way to know this without doing the experiment, since it depends on the biology of the particular gene, mutation and pathway involved, so there are three possible outcomes:

Case 2a) If *aa* is an enhancer of *gg*, and *aaG_* plants have a mutant phenotype that differs from wild-type or (*A_gg*) then the phenotypic ratios among the progeny of a dihybrid cross will be:

9	3	3	1
*A_G_*	*A_gg*	*aaG_*	*aagg*
wild-type	tubular leaves (some phenotype that differs from gg; maybe small twisted leaves)	abnormal leaves	no leaves

Case 2b) If *aa* is an enhancer of *gg*, and *aaG_* plants have a mutant phenotype that is the same as *A_gg* , the phenotypic ratios among the progeny of a dihybrid cross will be:

9	6	1
*A_G_*	*A_gg aaG_*	*aagg*
wild-type	tubular leaves	no leaves

Case 2c) If *aa* is an enhancer of *gg*, and *aaG_* do not have a phenotype that differs from wild-type then the phenotypic ratios among the progeny of a dihybrid cross will be:

12	3	1
*A_G_ aaG_*	*A_gg*	*aagg*
wild-type	tubular leaves	no leaves

Case 3: If the normal function of gene *A* is in the same process as *G*, such that *a* is a recessive allele that **decreases** the severity of the *gg* mutant (**i.e. *a* is an suppressor of *g***) then the phenotype of *aagg* could be : wild-type. The phenotypic ratios among the progeny of a dihybrid cross depend on whether *aa* mutants have a phenotype independent of *gg*, in other words, do *aaG_* plants have a phenotype that is different from wild-type or from *A_gg*. There is no way to know this without doing the experiment, since it depends on the biology of the particular gene, mutation and pathway involved, so there are three possible outcomes:

Case 3a) If *aa* is a suppressor of *gg*, and *aaG_* plants have a mutant phenotype that differs from wild-type or (*A_gg*) then the phenotypic ratios among the progeny of a dihybrid cross will be:

10	3	3
A_G_ aagg	*A_gg*	*aaG_*
wild-type	tubular leaves (some phenotype that differs from *gg*)	no leaves

Case 3b) If *aa* is an suppressor of *gg*, and *aaG_* plants have a mutant phenotype that is the same as *A_gg* the phenotypic ratios among the progeny of a dihybrid cross will be:

10	6
A_G_ aagg	*A_gg aaG_*
wild-type	tubular leaves

Case 3c) If *aa* is an suppressor of *gg*, and *aaG_* plants do not have a phenotype that differs from wild-type then the phenotypic ratios among the progeny of a dihybrid cross will be:

13	3
A_G_ aaG_ aagg	*A_gg*
wild-type	tubular leaves

Case 4: If the normal function of gene *A* is in the same process as *G*, such that *a* is a recessive allele that with a phenotype that **is epistatic** to the *gg* mutant then the phenotype of both *aaG_* and *aagg* could be : no leaves. The phenotypic ratios among

the progeny of a dihybrid cross will be:

9	4	3
*A_G_*	*aaG_ aagg*	*A_gg*
wild-type	no leaves	tubular leaves

Case … ?: There are many more phenotypes and ratios that could be imagined (e.g. different types of dominance relationships, different types of epistasis, lethality…etc). Isn't genetics wonderful? It is sometimes shocking that more people don't want to become geneticists.

The point of this exercise is to show that many different ratios can be generated, depending on the biology of the genes involved. On an exam, you could be asked to calculate the ratio, given particular biological parameters. So, this exercise is also meant to demonstrate that it is better to learn how to calculate ratios than just trying to memorize which ratios match which parameters. In a real genetic screen, you would observe the ratios, and then try to deduce something about the biology from those ratios.

6.9 Calculate the phenotypic ratios from a dihybrid cross involving the two loci shown in Figure 6.13. There may be more than one possible set of ratios, depending on the assumptions you make about the phenotype of allele *b*.

Assuming that bb has no phenotype on its own (i.e. *A_bb* looks like *A_B_)*, then *aaB_* will have the mutant phenotype, and *A_bb*, *A_B_*, and aabb will appear phenotypically wild-type. The phenotypic ratio will be 13 wild-type : 3 mutant.

6.10 For a dihybrid cross, there are 4 classes, 9:3:3:1. In a trihybrid cross without gene interactions, each of these 4 classes will be further split into a 3:1 ratio based on the phenotype at the third locus. For example, 9 x 3 =27 and 9 x 1 = 9. This explains the first two terms of the complete ratio: 27:9:9:9:3:3:3:1.

This page titled 5.E: Genetic Analysis of Multiple Genes (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

5.S: Genetic Analysis of Multiple Genes (Summary)

- The alleles of different loci are inherited independently of each other, unless they are genetically linked.
- The expected phenotypic ratio of a dihybrid cross is 9:3:3:1, except in cases of linkage or gene interactions that modify this ratio.
- Modified ratios from 9:3:3:1 are seen in the case of recessive and dominant epistasis, duplicate genes, and complementary gene action. This usually indicates that the two genes interact within the same biological pathway.

Key Terms

Mendel's Second Law	linkage	
Dihybrid	recessive epistasis	duplicate gene action
9:3:3:19:3:4	dominant epistasis	Orange, long, dilute, White masking, piebald spotting, calico
12:3:1	complementary action	
independent assortment	redundancy	

This page titled 5.S: Genetic Analysis of Multiple Genes (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **6.S: Genetic Analysis of Multiple Genes (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

6: Sex Chromosomes Sex Linkage

6.1: Introduction
6.2: Autosomes and Sex Chromosomes
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
6.4: Sex Linkage- An Exception to Mendel's First Law
6.5: Y-Linked Genes
6.6: Chapter 10 Summary
6.7: Key Terms in Chapter 10
6.8: Chapter 10 Study Questions
6.9: Chapter 10 Answers

This page titled 6: Sex Chromosomes Sex Linkage is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Introduction

Learning Objectives

- Distinguish between autosomes and sex chromosomes.
- Discuss sex-linkage as an exception to Mendel's First Law.
- Describe sex-linked pattern of inheritance, with specific reference to *Drosophila melanogaster* as a model genetic organism.
- Discuss the role of the Y-chromosome in sex-determination.
- Recall and describe selected Y-linked characteristics.

Previously, Mendel, working with plants, showed patterns of inheritance derived from gene loci on **autosomal** chromosomes. One complication to this model of inheritance in animals is that loci present on sex chromosomes (see **Figure 10.1.1** for example), called **sex-linked** loci, don't follow this pattern. This chapter covers the various patterns of inheritance for various sex-linked loci.



Figure 10.1.1 The E/e Gene in Turkeys is Responsible for Bronze or Brown Feather Colour, and is Located on the Z-Chromosome. [Long description]

Figure 10.1.2 shows that most of the chromosomes in humans are present in two copies. Each copy has the same length, centromere location, and banding pattern. As mentioned before, these are called autosomes. However, note that two of the chromosomes, the X and the Y, do not look alike. These are **sex chromosomes** . In mammals, males have one of each while females have two X chromosomes.

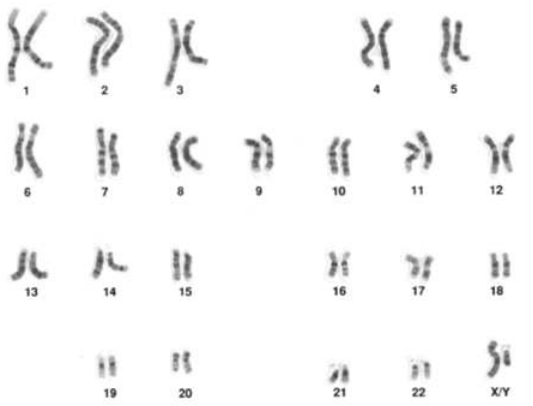


Figure 10.1.2 Karyogram of a Normal Human Male [Long description]

Watch the video below, *Sex Determination | Genetics | Biology | FuseSchool*, presented by FuseSchool – Global Education (2017) on YouTube,which describes how the sex-chromosomes play a role in sex-determination in humans.



Media Attribution

- **Figure 10.1.1**Wild Turkey strut by Steve Voght, CC BY-SA 2.0, via Flickr
- **Figure 10.1.2**Human male karyotpe high resolution by National Human Genome Research Institute/ Talking Glossary of Genetics, public domain, via Wikipedia

FuseSchool – Global Education. (2017, November 27). Sex determination | genetics | biology | FuseSchool (video file). YouTube. https://youtu.be/D2hVgujy2E8

Long Descriptions

- **Figure 10.1.1** A turkey demonstrates the phenotype of a bronze feather color, which is a gene located on the Z-chromosome or a sex chromosome of this animal. Sex-linked characteristics (e.g., bronze or brown feather color in turkeys) show different inheritance patterns than autosomal linked traits. [Back to Figure 10.1.1]
- **Figure 10.1.2** A karyogram, which is a visualization of all the chromosomes in their condensed state of a cell in a particular organism. Here, 23 pairs of chromosomes are observed in a somatic cell of a human male: we see 22 pairs of autosomes and one pair of sex-chromosomes which are labelled XY. The Y chromosome is significantly smaller than the X chromosome. [Back to Figure 10.1.2]

This page titled 6.1: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.1: Introduction** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.2: Autosomes and Sex Chromosomes

In diploids, most chromosomes exist in pairs (same length, centromere location, and banding pattern), with one set coming from each parent. These chromosomes are called **autosomes**. However, many species have an additional pair of chromosomes that do not look alike. These are **sex chromosomes** because they differ between the sexes. In humans, males have one of each while females have two X chromosomes. Autosomes are those chromosomes present in the same number in males and females, while sex chromosomes are those that are not. When sex chromosomes were first discovered, their function was unknown, and the name X was used to indicate this mystery. The next ones were named Y, then Z, and then W (depending on the species).

The combination of sex chromosomes within a species is associated with either male or female individuals. In mammals, fruit flies, and some dioecious plants, those with two X chromosomes are females, while those with an X and a Y are males. In birds, moths, and butterflies, males are ZZ and females are ZW. Because sex chromosomes have arisen multiple times during evolution the molecular mechanism(s) through which they determine sex differs among those organisms. For example, although humans and Drosophila both have X and Y sex chromosomes, they have different mechanisms for determining sex (see Chapter 11).

How do the sex chromosomes behave during meiosis? Well, in those individuals with two of the same chromosomes (i.e., **homogametic** sexes: XX females and ZZ males), the chromosomes pair and segregate during meiosis I, the same as autosomes. During meiosis in XY males or ZW females (**heterogametic** sexes), the sex chromosomes pair with each other.

Take a look at this video, *7R – Sex chromosomes in Meiosis*, produced by Professor Redfield of UBC (Useful Genetics, 2015), which discusses what occurs with sex chromosomes during meiosis.



In mammals (XX, XY), the consequence is all egg cells will carry an X chromosome, while the sperm cells will carry either an X or a Y chromosome. Half of the offspring will receive two X chromosomes and become female, while half will receive an X and a Y and become male (**Figure 10.3**). In species with ZZ males, all sperm carry a Z chromosome, while in females, ZW, half will have a Z and half a W.

It is a popular misconception that the X and Y chromosomes were named based upon their shapes; physically each looks like any other chromosome. A Y-chromosome doesn't look like a Y any more than a chromosome 4 looks like a 4. The combination of sex chromosomes within a species is associated with either male or female individuals. In mammals, fruit flies, and some flowering plants, XX individuals are females, while XY individuals are males.

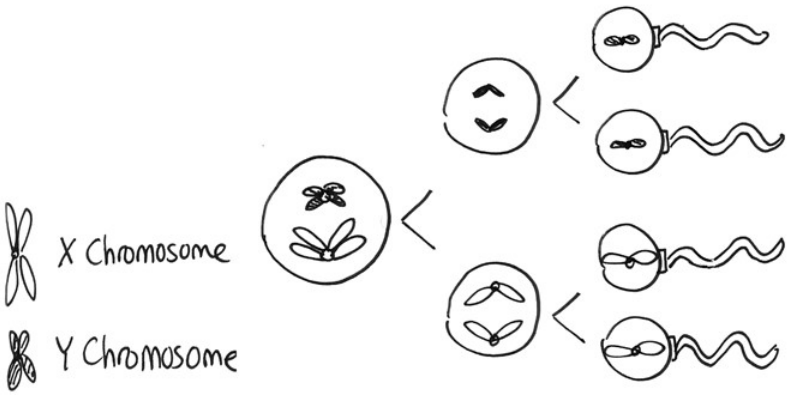


Figure 10.2.1 Meiosis in an XY Mammal. The stages shown are anaphase I, anaphase II, and mature sperm. Note how half of the sperm contain Y chromosomes and half contain X chromosomes. [Long description]

Media Attributions

- **Figure 10.2.1**Original by M. Harrington (2017), CC BY-NC 3.0, Open Genetics Lectures

References

Harrington, M. (2017). Figure 2. Meiosis in an XY mammal [digital image]. In Locke, J., Harrington, M., Canham, L. and Min Ku Kang (Eds.), *Open Genetics Lectures, Fall 2017* (Chapter 20, p. 2). Dataverse/ BCcampus. http://solr.bccampus.ca:8001/bcc/file/7a7b00f9-fb56-4c49-81a9-cfa3ad80e6d8/1/OpenGeneticsLectures_Fall2017.pdf

Redfield, R./ UBC [Useful Genetics]. (2015, August 23). 7R – Sex chromosomes in meiosis (video file). YouTube. https://www.youtube.com/watch?v=kiZWI_yuGDM

Long Description

- **Figure 10.2.1** The process of meiosis in a human male demonstrates the stages of anaphase I, anaphase II, and mature sperm cells. Half of the sperm cells produced contain an X chromosome, and half of the sperm cells contain a Y chromosome. This indicates males can only pass on their Y linked traits to their sons, not their daughters. [Back to Figure 10.2.1

This page titled 6.2: Autosomes and Sex Chromosomes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.2: Autosomes and Sex Chromosomes** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes

In evolution, before the X and Y chromosomes differentiated, they used to be equivalent homologs, like an autosome. Over time, the Y chromosome lost most of its genes (hence the reduced size), but the X chromosome retained all its genes. Thus, even though the Y chromosome has lost most of its genes, it still shares some regions with the X chromosome. This is the reason why although X and Y chromosomes are **heteromorphic** (morphologically dissimilar), they are able to act as a homologous pair in meiosis and undergo crossover. These common regions, which contain similar genes, permit the X and Y to pair up, and are called the "**pseudo-autosomal regions**". The name comes from the observation that genes in these regions behave like autosomes in their inheritance. Alleles of the genes in this region crossover just like those on the autosomes. Thus, genes in this region are not inherited in a sex-linked pattern, even though they are located on the X chromosome.

The genes found in pseudo-autosomal region are present in two copies in both XY males and XX females, and, thus, expressed from both active and inactive X chromosomes. These genes may explain clinical features in sex chromosome aneuploidy (addition or subtraction of a sex chromosome; e.g., XXY) as gene products may be either under- or over-expressed in relation to normal females and males.

One of the genes in this region is called *SHOX*. It makes a protein that promotes bone growth. 46,XX and 46,XY people have two functioning copies and have average height. People with 47,XYY and 47,XXX genomes have three copies and are taller than average. And people with 45,X have one copy and are short. It is the single copy of *SHOX* and a few other genes in the pseudo-autosomal region that causes health problems for women with Turner syndrome.

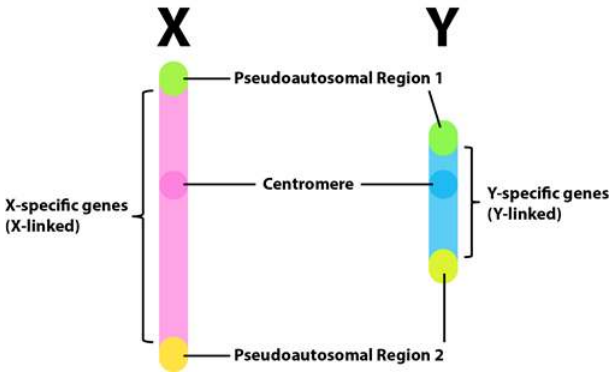
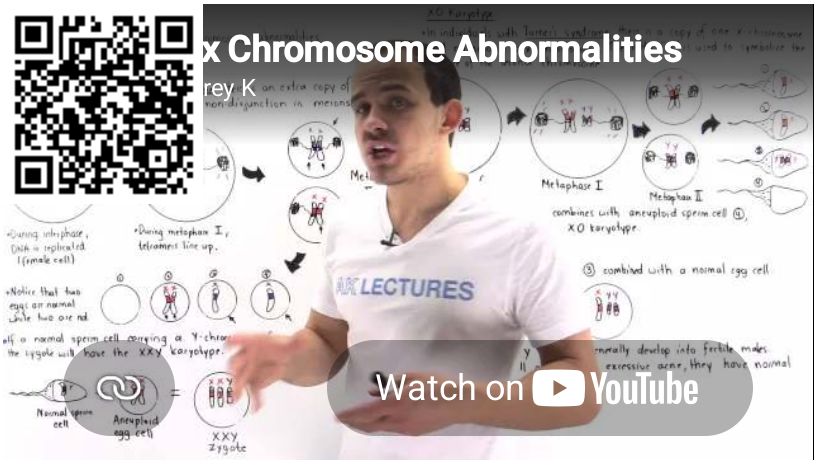


Figure 10.3.1 X and Y Chromosome have Pseudoautosomal Regions, Which are Capable of Pairing During Meiosis and Recombination. [Long description]

Take a look at this video, *Sex Chromosome Abnormalities*, part of the AK Lectures (2015) series which discusses abnormalities which can occur with the sex chromosomes.



Media Attribution

- **Figure 10.3.1**Original by Locke/Kang (2017), CC BY-NC 3.0, Open Genetics Lectures

References

AK Lectures. (2015, January 15). Sex chromosome abnormalities (video file). https://www.youtube.com/watch?v=gdXHq8FrfHI

Locke, J., Kang, M.K. (2017). Figure 3. X and Y chromosome have pseudoautosomal regions [digital image]. In Locke, J., Harrington, M., Canham, L. and Min Ku Kang (Eds.), *Open Genetics Lectures*, *Fall 2017* (Chapter 20, p. 2). Dataverse/BCcampus. http://solr.bccampus.ca:8001/bcc/file/7a7b00f9-fb56-4c49-81a9-cfa3ad80e6d8/1/OpenGeneticsLectures_Fall2017.pdf

Long Description

- **Figure 10.3.1** An X chromosome and a Y chromosome, showing centromeres, and Pseudoautosomal regions, are able to pair up during meiosis. Crossing over does not occur, since the X and Y chromosomes are heteromorphic. However, pairing up homologous chromosomes is necessary for proper division of nuclei during meiosis. [Back to Figure 10.3.1]

This page titled 6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.3: Pseudo-Autosomal Regions on the X and Y Chromosomes** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.4: Sex Linkage- An Exception to Mendel's First Law

We have introduced sex chromosomes and autosomes (non-sex-linked chromosomes). For loci on autosomes, the alleles follow the classic Mendelian pattern of inheritance. However, for loci on the sex chromosomes this doesn't follow because most (not all) of the loci on the typical X-chromosome are absent from the Y-chromosome, even though they act as a homologous pair during meiosis. Instead, they will follow a **sex-linked** pattern of inheritance. An X-linked allele in the father will always be passed on to his daughters only, but an X-linked allele in the mother will be passed on to both daughters and sons equally.

X-Linked Genes: The White Gene in Drosophila Melanogaster

A well-studied, sex-linked gene is the *white* gene on the X chromosome of *Drosophila melanogaster*. Normally flies have red eyes, but flies with a mutant allele of this gene called *white–* (*w–*) have white eyes because the red pigments are absent. Because this mutation is recessive to the wild type *w+* allele, females that are heterozygous have normal red eyes. Female flies that are homozygous for the mutant allele have white eyes. Because there is no *white* gene on the Y chromosome, male flies can only be hemizygous for the wild type allele or the mutant allele.

A researcher may not know beforehand whether a novel mutation is sex-linked. The definitive method to test for sex-linkage is **reciprocal crosses** (**Figure 10.4.2**). This means to cross a male and a female that have different phenotypes, and then conduct a second set of crosses, in which the phenotypes are reversed relative to the sex of the parents in the first cross. For example, if you were to set up reciprocal crosses with flies from pure-breeding *w+* and *w–* strains the results would be as shown in **Figure 10.4.2**. Whenever reciprocal crosses give different results in the F_1 and F_2 and whenever the male and female offspring have different phenotypes the usual explanation is sex-linkage. Remember, if the locus were autosomal, the F_1 and F_2 progeny would be different from either of these crosses.

A similar pattern of sex-linked inheritance is seen for X-chromosome loci in other species with an XX-XY sex chromosome system, including mammals and humans.

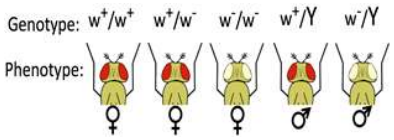


Figure 10.4.1 – Relationship between genotype and phenotype for the white gene on the X-linked gene in Drosophila melanogaster. The Y chromosome is indicated with a capital Y because it does not have a copy of the white gene. [Long description]

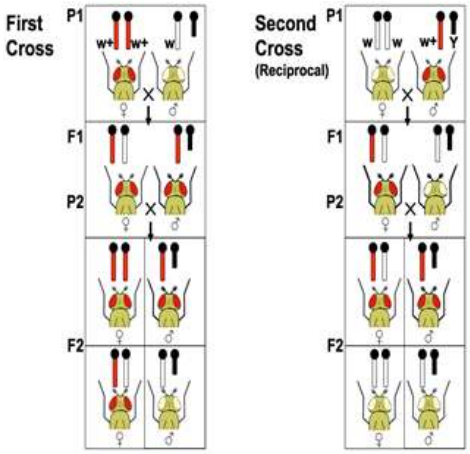
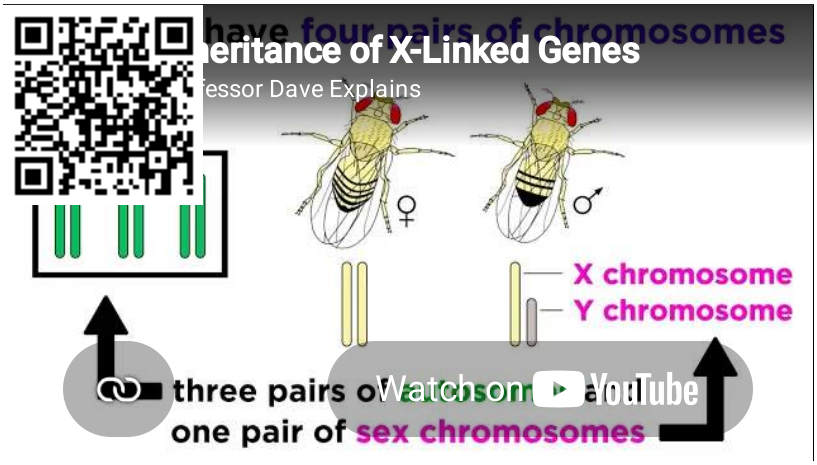


Figure 10.4.2 Reciprocal crosses involving an X-linked gene in Drosophila melanogaster. In the first cross (left), all offspring have red eyes. In second (reciprocal) cross (right), all female offspring have red eyes and male offspring all have white eyes. If the F1 progeny are crossed (to make the P2), the F2 progeny will be different in each cross. The first cross has all red-eyed females and half red-eyed males. The reciprocal cross has half red-eyed males and females. [Long description]

Thomas Morgan was awarded the Nobel Prize, in part, for using crosses like the ones represented in **Figure 10.4.2,** to demonstrate that genes (such as white) were on chromosomes — in this case the X-chromosome.

Take a look at the video below, *Inheritance of X-Linked Genes*, presented by Professor Dave Explains (2020), which discusses the inheritance of sex-linked traits.



Media Attributions

- **Figure 10.4.1** Original by Deyholos/Harrington/Locke (2017), CC BY-NC 3.0, Open Genetics Lectures
- **Figure 10.4.2** Original by Deyholos/Harrington/Locke (2017), CC BY-NC 3.0, Open Genetics Lectures

Reference

Deyholos, M., Harrington, M., & Locke, J. (2017). Figures: 4. Relationship between genotype and phenotype…; and 5. Reciprocal crosses involving an X-linked gene [digital image]. In Locke, J., Harrington, M., Canham, L. and Min Ku Kang (Eds.), *Open Genetics Lectures, Fall 2017* (Chapter 20, p. 3). Dataverse/ BCcampus. http://solr.bccampus.ca:8001/bcc/file/7a7b00f9-fb56-4c49-81a9-cfa3ad80e6d8/1/OpenGeneticsLectures_Fall2017.pdf

Professor Dave Explains. (2020, March 11). Inheritance of X-linked genes (video file). YouTube. https://www.youtube.com/watch?v=IShS60Azqjg

Long Descriptions

- **Figure 10.4.1** Relationship between genotype and phenotype for the white gene on the X-linked gene in Drosophila melanogaster. The Y chromosome is indicated with a capital Y because it does not have a copy of the white gene. This particular trait (eye colour) is sex-linked in Drosophila. [Back to Figure 10.4.1]
- **Figure 10.4.2** Reciprocal crosses involving an X-linked gene in Drosophila melanogaster. In the first cross (left), all offspring have red eyes. In second (reciprocal) cross (right), all female offspring have red eyes and male offspring all have white eyes. If the F1 progeny are crossed (to make the P2), the F2 progeny will be different in each cross. The first cross has all red-eyed females and half red-eyed males. The reciprocal cross has half red-eyed males and females. [Back to Figure 10.4.2]

This page titled 6.4: Sex Linkage- An Exception to Mendel's First Law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.4: Sex Linkage- An Exception to Mendel's First Law** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.5: Y-Linked Genes

In humans, the Y chromosome has been studied and is known to contain approximately 200 genes which provide instructions for making proteins. Because only males have the Y chromosome, the genes on this chromosome tend to be involved in male sex determination and development. Sex is determined by the SRY gene, which is the **sex-determining region** of the Y-chromosome.

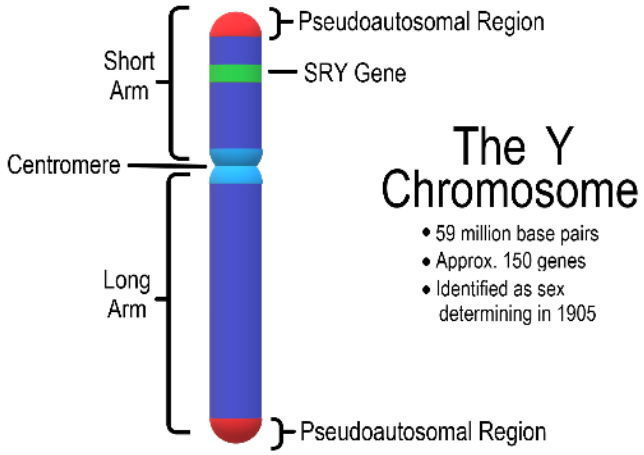
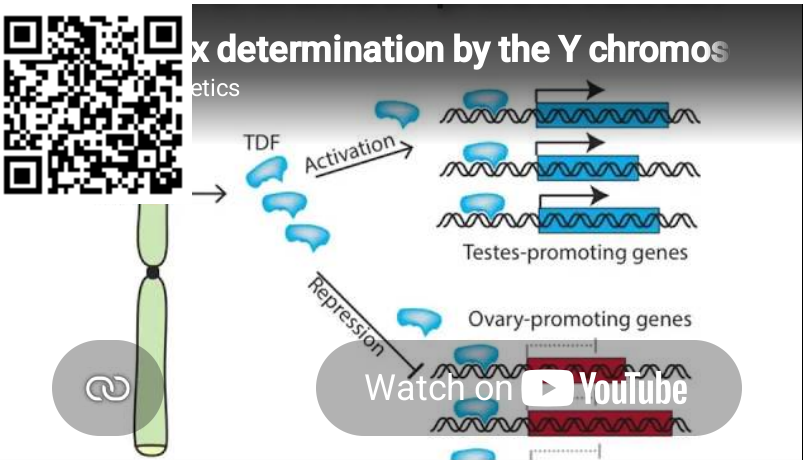


Figure 10.5.1 Structure of the Y-Chromosome in Humans [Long description]

Other genes on the Y chromosome are important for enabling men to father biological children (male fertility). Many genes are unique to the Y chromosome, but genes in areas known as pseudoautosomal regions are present on both sex chromosomes. As a result, men and women each have two functional copies of these genes. Many genes in the pseudoautosomal regions are essential for normal development. Although the Y-chromosome is sex-determining in humans and some other species, not all genes that play a role in sex determination are Y-linked.

Y-linked traits, while few in number, do exist. For instance, the Y-linked trait of “webbed toes” causes a web-like connection between second and third toes, and “porcupine man” occurs when the skin thickens and gradually becomes darker, scaly, rough, and with bristle-like outgrowths. Since Y-linked inheritance involves the Y chromosome, Y-linked inheritance is passed on from father to son. Of course, Y-linked traits never occur in females, and occur in all male descendants of an affected male. The concepts of dominant and recessive do not apply to Y-linked traits, as only one allele (on the Y) is ever present in any one (male) individual (this, is of course, ignoring XYY syndrome, which is a rare chromosomal disorder that affects males. It is caused by the presence of an extra Y chromosome. Males normally have one X and one Y chromosome. However, individuals with this syndrome have one X and two Y chromosomes).

Take a look at the video, *Sex determination by the Y chromosome*, by Genetics/ UC Davis (2017) on YouTube, which looks at sex determination by the Y-chromosome.



Media Attribution

- **Figure 10.5.1** Y Chromosome by Christinelmiller, CC BY-SA 4.0, via Wikimedia Commons

Reference

Genetics/ UC Davis (2017, December 31). Sex determination by the Y chromosome (video file). YouTube. https://www.youtube.com/watch?v=Fd31_gaRyME

Long Description

- **Figure 10.5.1**The structure of the Y-chromosome in humans shows pseudoautosomal regions at the ends, the position of the centromere, the short arm, which contains the loci for the SRY gene – the sex-determining region of the Y-chromosome, and the long arm. The Y chromosome contains approximately 59 million base pairs, and carries approximately 150 genes, and was identified as a sex chromosome in 1905. [Back to Figure 10.5.1]

This page titled 6.5: Y-Linked Genes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

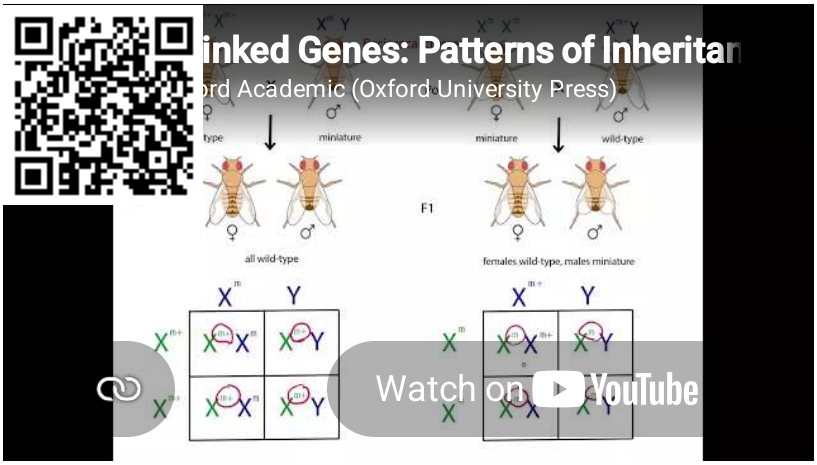
- **10.5: Y-Linked Genes** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.6: Chapter 10 Summary

The topics covered in this chapter can be summarized as follows:

- Autosomes and sex chromosomes differ in that the former exist in pairs but the latter depends on the sex of the chromosome.
- Pseudo-autosomal regions are regions on X and Y chromosome that can pair up and recombine.
- Sex-linked genes are an exception to standard Mendelian inheritance. Their phenotypes are influenced by the type of sex chromosome system and the type of dosage compensation system found in the species.
- Some of the examples of sex-linked genes are: *white* gene on the Drosophila's X chromosome, *TDF* gene on Y chromosome, *E/e* gene on Z chromosome (birds).

For further interest, take a look at this video, *X-Linked Genes: Patterns of Inheritance* by Oxford Academic (Oxford University Press, 2017) on YouTube.



Reference

Oxford Academic (Oxford University Press). (2017). X-linked genes: Patterns of inheritance (video file). YouTube. https://www.youtube.com/watch?v=WNEoT7KhQPI

This page titled 6.6: Chapter 10 Summary is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.6: Chapter 10 Summary** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.7: Key Terms in Chapter 10

Key Terms – Sex Chromosomes & Sex Linkage

- autosome
- heterogametic
- heteromorphic
- homogametic
- pseudoautosomal regions
- reciprocal cross
- sex chromosome
- sex-linked
- X-linked genes

This page titled 6.7: Key Terms in Chapter 10 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.7: Key Terms in Chapter 10** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.8: Chapter 10 Study Questions

1. A rare dominant mutation causes a neurological disease that appears late in life in all people that carry the mutation. If a father has this disease, what is the probability that his daughter will also have the disease?
2. Construct Punnett Squares to accompany the crosses shown in **Figure 10.4.2.**
3. Draw reciprocal crosses that would show that the turkey E-gene is on the Z-chromosome.

Some extra reading and other activities!

4. Read "The Y chromosome is disappearing – so what will happen to men?" a 2018 article published in *The Conversation* by Darren Griffin and Peter Ellis from the University of Kent. It provides a discussion about the rate of disappearance of the Y-chromosome in the human genome. Also, take a look at the following animation video on YouTube, *Recombination of the Y Chromosome | HHMI BioInteractive Video*, which relates to this article.
5. Visit the HHMI *Sex Verification Testing of Athletes* interactive, by BioInteractive. Relate this back to crossing over in meiosis (*In Recombination of the Y Chromosome | HHMI BioInteractive* Video on YouTube).

References

Griffin, D., Ellis, P. (2018, January 17). The Y chromosome is disappearing – so what will happen to men. *The Conversation* (accessed January 11, 2021). https://theconversation.com/the-y-chromosome-is-disappearing-so-what-will-happen-to-men-90125

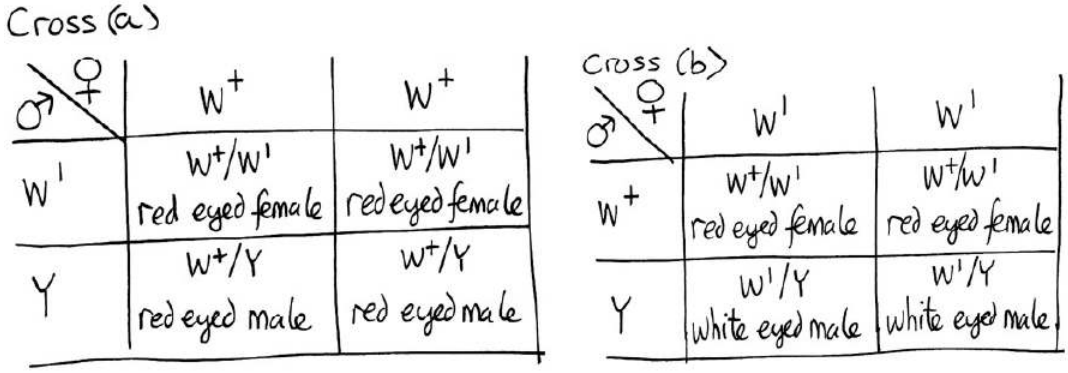
HHMI BioInteractive. (2018, November 29). Recombination of the Y chromosome | HHMI BioInteractive video (video file). YouTube. https://www.youtube.com/watch?v=nIBPBM2z3Kg&feature=emb_logo

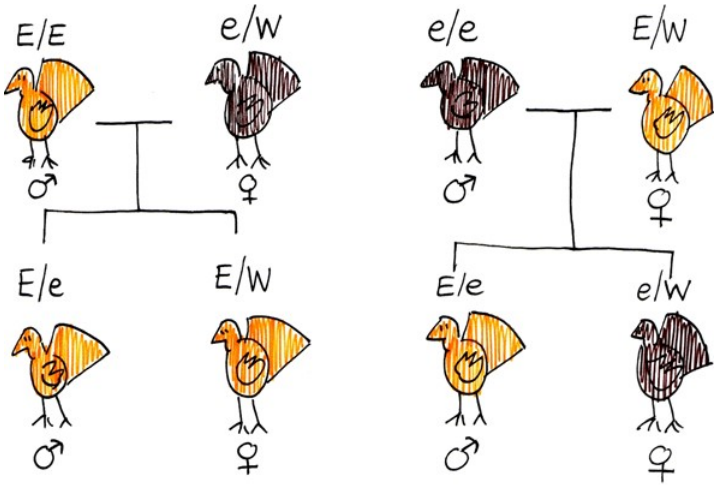
HHMI BioInteractive. (n.d.). Sex verification testing of athletes [interactive]. HHMI BioInteractive (accessed January 11, 2021) https://www.biointeractive.org/classroom-resources/sex-verification-testing-athletes

This page titled 6.8: Chapter 10 Study Questions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.8: Chapter 10 Study Questions** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

6.9: Chapter 10 Answers

1. It depends on the chromosomal location of the disease locus. If the gene is autosomal, the probability is 50%. If it is sex-linked, that is on the X-chromosome, it would be 100%. If it is Y-linked, then 0%. In both situations the probability would decrease if the penetrance was less than 100%.
2. 

3. 

This page titled 6.9: Chapter 10 Answers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.9: Chapter 10 Answers** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

CHAPTER OVERVIEW

7: DNA STRUCTURE AND FUNCTION

Each human cell has 23 pairs of chromosomes: one set of chromosomes is inherited from the mother and the other set is inherited from the father. There is also a mitochondrial genome, inherited exclusively from the mother, which can be involved in inherited genetic disorders. On each chromosome, there are thousands of genes that are responsible for determining the genotype and phenotype of the individual. A gene is defined as a sequence of DNA that codes for a functional product. The human haploid genome contains 3 billion base pairs and has between 20,000 and 25,000 functional genes.

7.1: Introduction
7.2: Historical Basis of Modern Understanding
7.3: DNA Structure and Sequencing
7.4: Basics of DNA Replication
7.5: DNA Replication in Prokaryotes
7.6: DNA Replication in Eukaryotes
7.7: DNA Repair
7.8: Key Terms
7.9: Chapter Summary
7.10: Visual Connection Questions
7.11: Review Questions
7.12: Critical Thinking Questions

This page titled 7: DNA Structure and Function is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.1: INTRODUCTION



Figure 14.1 Dolly the sheep was the first large mammal to be cloned.

CHAPTER OUTLINE

14.1 Historical Basis of Modern Understanding
14.2 DNA Structure and Sequencing
14.3 Basics of DNA Replication
14.4 DNA Replication in Prokaryotes
14.5 DNA Replication in Eukaryotes
14.6 DNA Repair

The three letters “DNA” have now become synonymous with crime solving and genetic testing. DNA can be retrieved from hair, blood, or saliva. Each person’s DNA is unique, and it is possible to detect differences between individuals within a species on the basis of these unique features.

DNA analysis has many practical applications beyond forensics. In humans, DNA testing is applied to numerous uses: determining paternity, tracing genealogy, identifying pathogens, archeological research, tracing disease outbreaks, and studying human migration patterns. In the medical field, DNA is used in diagnostics, new vaccine development, and cancer therapy. It is now possible to determine predisposition to diseases by looking at genes.

Each human cell has 23 pairs of chromosomes: one set of chromosomes is inherited from the female parent and the other set is inherited from the male parent. There is also a mitochondrial genome, inherited exclusively from the female parent, which can be involved in inherited genetic disorders. On each chromosome, there are thousands of genes that are responsible for determining the genotype and phenotype of the individual. A gene is defined as a sequence of DNA that codes for a functional product. The human haploid genome contains 3 billion base pairs and has between 20,000 and 25,000 functional genes.

This page titled 7.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.2: HISTORICAL BASIS OF MODERN UNDERSTANDING

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Explain transformation of DNA
- Describe the key experiments that helped identify that DNA is the genetic material
- State and explain Chargaff's rules

Our current understanding of DNA began with the discovery of nucleic acids followed by the development of the double-helix model. In the 1860s, Friedrich Miescher (Figure 14.2), a physician by profession, isolated phosphate-rich chemicals from white blood cells (leukocytes). He named these chemicals (which would eventually be known as DNA) *nuclein* because they were isolated from the nuclei of the cells.



Figure 14.2 Friedrich Miescher (1844–1895) discovered nucleic acids.

LINK TO LEARNING

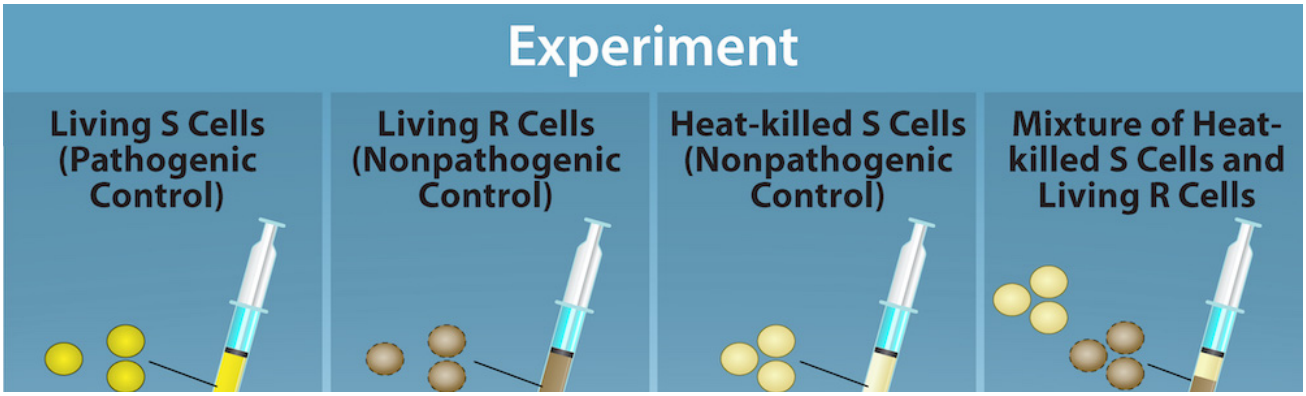
LINK TO LEARNING

To see Miescher conduct his experiment that led to his discovery of DNA and associated proteins in the nucleus, click through this review.

A half century later, in 1928, British bacteriologist Frederick Griffith reported the first demonstration of bacterial transformation—a process in which external DNA is taken up by a cell, thereby changing its morphology and physiology. Griffith conducted his experiments with *Streptococcus pneumoniae*, a bacterium that causes pneumonia. Griffith worked with two strains of this bacterium called rough (R) and smooth (S). (The two cell types were called "rough" and "smooth" after the appearance of their colonies grown on a nutrient agar plate.)

The R strain is non-pathogenic (does not cause disease). The S strain is pathogenic (disease-causing), and has a capsule outside its cell wall. The capsule allows the cell to escape the immune responses of the host mouse.

When Griffith injected the living S strain into mice, they died from pneumonia. In contrast, when Griffith injected the live R strain into mice, they survived. In another experiment, when he injected mice with the heat-killed S strain, they also survived. This experiment showed that the capsule alone was not the cause of death. In a third set of experiments, a mixture of live R strain and heat-killed S strain were injected into mice, and—to his surprise—the mice died. Upon isolating the live bacteria from the dead mouse, only the S strain of bacteria was recovered. When this isolated S strain was injected into fresh mice, the mice died. Griffith concluded that something had passed from the heat-killed S strain into the live R strain and transformed it into the pathogenic S strain. He called this the *transforming principle* (Figure 14.3). These experiments are now known as Griffith's transformation experiments.



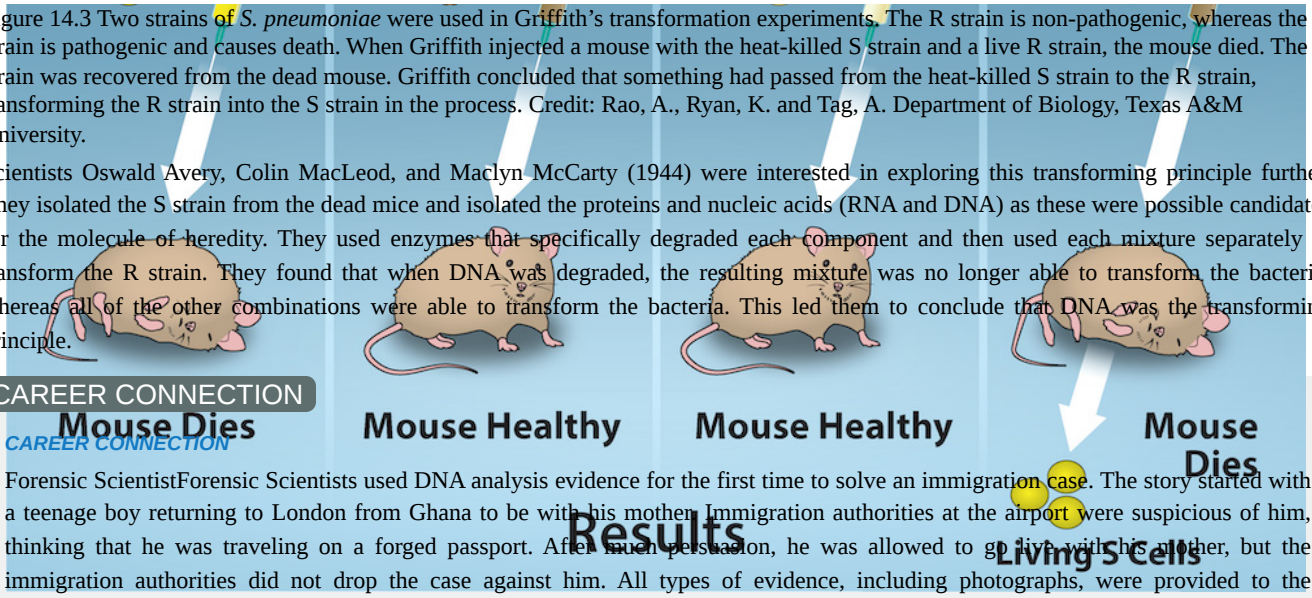


Figure 14.3 Two strains of *S. pneumoniae* were used in Griffith's transformation experiments. The R strain is non-pathogenic, whereas the S strain is pathogenic and causes death. When Griffith injected a mouse with the heat-killed S strain and a live R strain, the mouse died. The S strain was recovered from the dead mouse. Griffith concluded that something had passed from the heat-killed S strain to the R strain, transforming the R strain into the S strain in the process. Credit: Rao, A., Ryan, K. and Tag, A. Department of Biology, Texas A&M University.

Scientists Oswald Avery, Colin MacLeod, and Maclyn McCarty (1944) were interested in exploring this transforming principle further. They isolated the S strain from the dead mice and isolated the proteins and nucleic acids (RNA and DNA) as these were possible candidates for the molecule of heredity. They used enzymes that specifically degraded each component and then used each mixture separately to transform the R strain. They found that when DNA was degraded, the resulting mixture was no longer able to transform the bacteria, whereas all of the other combinations were able to transform the bacteria. This led them to conclude that DNA was the transforming principle.

CAREER CONNECTION

CAREER CONNECTION

Forensic ScientistForensic Scientists used DNA analysis evidence for the first time to solve an immigration case. The story started with a teenage boy returning to London from Ghana to be with his mother. Immigration authorities at the airport were suspicious of him, thinking that he was traveling on a forged passport. After much persuasion, he was allowed to go live with his mother, but the immigration authorities did not drop the case against him. All types of evidence, including photographs, were provided to the authorities, but deportation proceedings were started nevertheless. Around the same time, Dr. Alec Jeffreys of Leicester University in the United Kingdom had invented a technique known as DNA fingerprinting. The immigration authorities approached Dr. Jeffreys for help. He took DNA samples from the mother and three of her children, as well as an unrelated mother, and compared the samples with the boy's DNA. Because the biological father was not in the picture, DNA from the three children was compared with the boy's DNA. He found a match in the boy's DNA for both the mother and his three siblings. He concluded that the boy was indeed the mother's son.

Forensic scientists analyze many items, including documents, handwriting, firearms, and biological samples. They analyze the DNA content of hair, semen, saliva, and blood, and compare it with a database of DNA profiles of known criminals. Analysis includes DNA isolation, sequencing, and sequence analysis. Forensic scientists are expected to appear at court hearings to present their findings. They are usually employed in crime labs of city and state government agencies. Geneticists experimenting with DNA techniques also work for scientific and research organizations, pharmaceutical industries, and college and university labs. Students wishing to pursue a career as a forensic scientist should have at least a bachelor's degree in chemistry, biology, or physics, and preferably some experience working in a laboratory.

Although the experiments of Avery, McCarty and McLeod had demonstrated that DNA was the informational component transferred during transformation, DNA was still considered to be too simple a molecule to carry biological information. Proteins, with their 20 different amino acids, were regarded as more likely candidates. The decisive experiment, conducted by Martha Chase and Alfred Hershey in 1952, provided confirmatory evidence that DNA was indeed the genetic material and not proteins. Chase and Hershey were studying a bacteriophage—a virus that infects bacteria. Viruses typically have a simple structure: a protein coat, called the capsid, and a nucleic acid core that contains the genetic material (either DNA or RNA). The bacteriophage infects the host bacterial cell by attaching to its surface, and then it injects its nucleic acids inside the cell. The phage DNA makes multiple copies of itself using the host machinery, and eventually the host cell bursts, releasing a large number of bacteriophages. Hershey and Chase selected radioactive elements that would specifically distinguish the protein from the DNA in infected cells. They labeled one batch of phage with radioactive sulfur, ^{35}S, to label the protein coat. Another batch of phage were labeled with radioactive phosphorus, ^{32}P. Because phosphorous is found in DNA, but not protein, the DNA and not the protein would be tagged with radioactive phosphorus. Likewise, sulfur is absent from DNA, but present in several amino acids such as methionine and cysteine.

Each batch of phage was allowed to infect the cells separately. After infection, the phage bacterial suspension was put in a blender, which caused the phage coat to detach from the host cell. Cells exposed long enough for infection to occur were then examined to see which of the two radioactive molecules had entered the cell. The phage and bacterial suspension was spun down in a centrifuge. The heavier bacterial cells settled down and formed a pellet, whereas the lighter phage particles stayed in the supernatant. In the tube that contained phage labeled with ^{35}S, the supernatant contained the radioactively labeled phage, whereas no radioactivity was detected in the pellet. In the tube that contained the phage labeled with ^{32}P, the radioactivity was detected in the pellet that contained the heavier bacterial cells, and no radioactivity was detected in the supernatant. Hershey and Chase concluded that it was the phage DNA that was injected into the cell and carried information to produce more phage particles, thus providing evidence that DNA was the genetic material and not proteins (Figure 14.4).

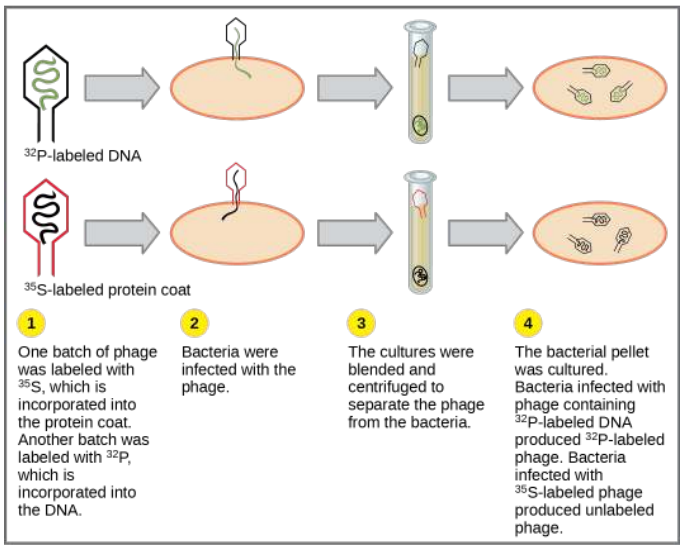


Figure 14.4 In Hershey and Chase's experiments, bacteria were infected with phage radiolabeled with either ^{35}S, which labels protein, or ^{32}P, which labels DNA. Only ^{32}P entered the bacterial cells, indicating that DNA is the genetic material.

Around this same time, Austrian biochemist Erwin Chargaff examined the content of DNA in different species and found that the amounts of adenine, thymine, guanine, and cytosine were not found in equal quantities, and that relative concentrations of the four nucleotide bases varied from species to species, but not within tissues of the same individual or between individuals of the same species. He also discovered something unexpected: That the amount of adenine equaled the amount of thymine, and the amount of cytosine equaled the amount of guanine (that is, A = T and G = C). Different species had equal amounts of *purines* (A+G) and pyrimidines (T + C), but different ratios of A+T to G+C. These observations became known as Chargaff's rules. Chargaff's findings proved immensely useful when Watson and Crick were getting ready to propose their DNA double helix model! You can see after reading the past few pages how science builds upon previous discoveries, sometimes in a slow and laborious process.

This page titled 7.2: Historical Basis of Modern Understanding is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.3: DNA STRUCTURE AND SEQUENCING

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Describe the structure of DNA
- Explain the Sanger method of DNA sequencing
- Discuss the similarities and differences between eukaryotic and prokaryotic DNA

The building blocks of DNA are nucleotides. The important components of the nucleotide are a nitrogenous (nitrogen-bearing) base, a 5-carbon sugar (pentose), and a phosphate group (Figure 14.5). The nucleotide is named depending on the nitrogenous base. The nitrogenous base can be a purine such as adenine (A) and guanine (G), or a pyrimidine such as cytosine (C) and thymine (T).

VISUAL CONNECTION

VISUAL CONNECTION

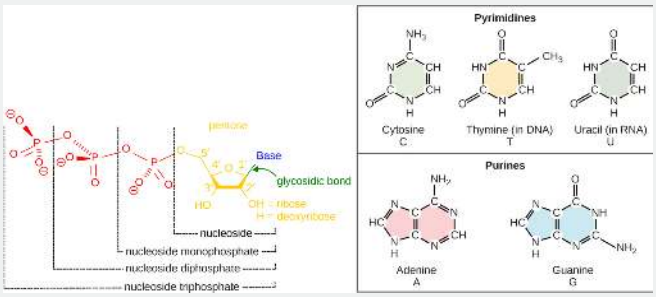


Figure 14.5 The purines have a double ring structure with a six-membered ring fused to a five-membered ring. Pyrimidines are smaller in size; they have a single six-membered ring structure.

The images above illustrate the five bases of DNA and RNA. Examine the images and explain why these are called “nitrogenous bases.” How are the purines different from the pyrimidines? How is one purine or pyrimidine different from another, e.g., adenine from guanine? How is a nucleoside different from a nucleotide?

The purines have a double ring structure with a six-membered ring fused to a five-membered ring. Pyrimidines are smaller in size; they have a single six-membered ring structure.

The sugar is deoxyribose in DNA and ribose in RNA. The carbon atoms of the five-carbon sugar are numbered 1', 2', 3', 4', and 5' (1' is read as “one prime”). The phosphate, which makes DNA and RNA acidic, is connected to the 5' carbon of the sugar by the formation of an ester linkage between phosphoric acid and the 5'-OH group (an ester is an acid + an alcohol). In DNA nucleotides, the 3' carbon of the sugar deoxyribose is attached to a hydroxyl (OH) group. In RNA nucleotides, the 2' carbon of the sugar ribose also contains a hydroxyl group. The base is attached to the 1'carbon of the sugar.

The nucleotides combine with each other to produce phosphodiester bonds. The phosphate residue attached to the 5' carbon of the sugar of one nucleotide forms a second ester linkage with the hydroxyl group of the 3' carbon of the sugar of the next nucleotide, thereby forming a 5'-3' phosphodiester bond. In a polynucleotide, one end of the chain has a free 5' phosphate, and the other end has a free 3'-OH. These are called the 5' and 3' ends of the chain.

In the 1950s, Francis Crick and James Watson worked together to determine the structure of DNA at the University of Cambridge, England. Other scientists like Linus Pauling and Maurice Wilkins were also actively exploring this field. Pauling previously had discovered the secondary structure of proteins using X-ray crystallography. In Wilkins’ lab, researcher Rosalind Franklin was using X-ray diffraction methods to understand the structure of DNA. Watson and Crick were able to piece together the puzzle of the DNA molecule on the basis of Franklin's data because Crick had also studied X-ray diffraction (Figure 14.6). In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine. Unfortunately, by then Franklin had died, and Nobel prizes are not awarded posthumously.



Figure 14.6 The X-ray diffraction pattern of DNA, which helped to elucidate its double-helix structure.

Watson and Crick proposed that DNA is made up of two strands that are twisted around each other to form a right-handed helix. Base pairing takes place between a purine and pyrimidine on opposite strands, so that A pairs with T, and G pairs with C (suggested by Chargaff's Rules). Thus, adenine and thymine are complementary base pairs, and cytosine and guanine are also complementary base pairs. The base pairs are stabilized by hydrogen bonds: *adenine and thymine form two hydrogen bonds and cytosine and guanine form three hydrogen bonds.* The two strands are anti-parallel in nature; that is, the 3' end of one strand faces the 5' end of the other strand. The sugar and phosphate of the nucleotides form the backbone of the structure, whereas the nitrogenous bases are stacked inside, like the rungs of a ladder. Each base pair is separated from the next base pair by a distance of 0.34 nm, and each turn of the helix measures 3.4 nm. Therefore, 10 base pairs are present per turn of the helix. The diameter of the DNA double helix is 2 nm, and it is uniform throughout. Only the pairing between a purine and pyrimidine and the antiparallel orientation of the two DNA strands can explain the uniform diameter. The twisting of the two strands around each other results in the formation of uniformly spaced major and minor grooves (Figure 14.7).

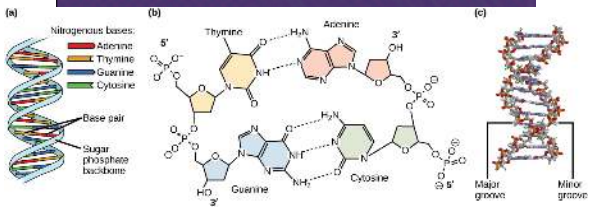


Figure 14.7 DNA has (a) a double helix structure and (b) phosphodiester bonds; the dotted lines between Thymine and Adenine and Guanine and Cytosine represent hydrogen bonds. The (c) major and minor grooves are binding sites for DNA binding proteins during processes such as transcription (the copying of RNA from DNA) and replication.

DNA SEQUENCING TECHNIQUES

Until the 1990s, the sequencing of DNA (reading the sequence of DNA) was a relatively expensive and long process. Using radiolabeled nucleotides also compounded the problem through safety concerns. With currently available technology and automated machines, the process is cheaper, safer, and can be completed in a matter of hours. Fred Sanger developed the sequencing method used for the human genome sequencing project, which is widely used today (Figure 14.8).

LINK TO LEARNING

LINK TO LEARNING

Visit this site to watch a video explaining the DNA sequence-reading technique that resulted from Sanger's work.

The sequencing method is known as the dideoxy chain termination method. The method is based on the use of chain terminators, the dideoxynucleotides (ddNTPs). The ddNTPSs differ from the deoxynucleotides by the lack of a free 3' OH group on the five-carbon sugar. If a ddNTP is added to a growing DNA strand, the chain cannot be extended any further because the free 3' OH group needed to add another nucleotide is not available. By using a predetermined ratio of deoxynucleotides to dideoxynucleotides, it is possible to generate DNA fragments of different sizes.

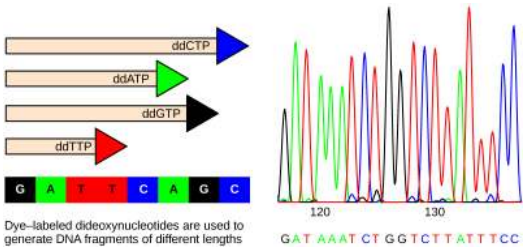


Figure 14.8 In Frederick Sanger's dideoxy chain termination method, dye-labeled dideoxynucleotides are used to generate DNA fragments that terminate at different points. The DNA is separated by capillary electrophoresis (not defined) on the basis of size, and from the order of fragments formed, the DNA sequence can be read. The DNA sequence readout is shown on an electropherogram (not defined) that is generated by a laser scanner.

The DNA sample to be sequenced is denatured (separated into two strands by heating it to high temperatures). The DNA is divided into four tubes in which a primer, DNA polymerase, and all four nucleoside triphosphates (A, T, G, and C) are added. In addition, limited quantities of one of the four dideoxynucleoside triphosphates (ddCTP, ddATP, ddGTP, and ddTTP) are added to each tube respectively. The tubes are labeled as A, T, G, and C according to the ddNTP added. For detection purposes, each of the four dideoxynucleotides carries a different fluorescent label. Chain elongation continues until a fluorescent dideoxy nucleotide is incorporated, after which no further elongation takes place. After the reaction is over, electrophoresis is performed. Even a difference in length of a single base can be detected. The sequence is

read from a laser scanner that detects the fluorescent marker of each fragment. For his work on DNA sequencing, Sanger received a Nobel Prize in Chemistry in 1980.

LINK TO LEARNING

LINK TO LEARNING

Sanger's genome sequencing has led to a race to sequence human genomes at rapid speed and low cost. Learn more by viewing the animation here.

Gel electrophoresis is a technique used to separate DNA fragments of different sizes. Usually the gel is made of a chemical called *agarose* (a polysaccharide polymer extracted from seaweed that is high in galactose residues). Agarose powder is added to a buffer and heated. After cooling, the gel solution is poured into a casting tray. Once the gel has solidified, the DNA is loaded on the gel and electric current is applied. The DNA has a net negative charge and moves from the negative electrode toward the positive electrode. The electric current is applied for sufficient time to let the DNA separate according to size; the smallest fragments will be farthest from the well (where the DNA was loaded), and the heavier molecular weight fragments will be closest to the well. Once the DNA is separated, the gel is stained with a DNA-specific dye for viewing it (Figure 14.9).

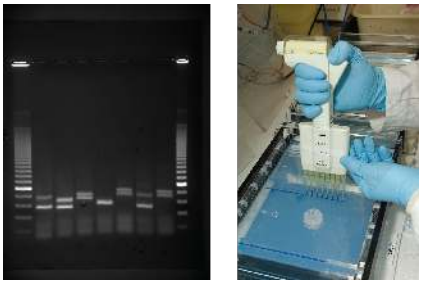


Figure 14.9 DNA can be separated on the basis of size using gel electrophoresis. (credit: James Jacob, Tompkins Cortland Community College)

EVOLUTION CONNECTION

EVOLUTION CONNECTION

Neanderthal Genome: How Are We Related?The first draft sequence of the Neanderthal genome was recently published by Richard E. Green et al. in 2010.[1] Neanderthals are the closest ancestors of present-day humans. They were known to have lived in Europe and Western Asia (and now, perhaps, in Northern Africa) before they disappeared from fossil records approximately 30,000 years ago. Green's team studied almost 40,000-year-old fossil remains that were selected from sites across the world. Extremely sophisticated means of sample preparation and DNA sequencing were employed because of the fragile nature of the bones and heavy microbial contamination. In their study, the scientists were able to sequence some four billion base pairs. The Neanderthal sequence was compared with that of present-day humans from across the world. After comparing the sequences, the researchers found that the Neanderthal genome had 2 to 3 percent greater similarity to people living outside Africa than to people in Africa. While current theories have suggested that all present-day humans can be traced to a small ancestral population in Africa, the data from the Neanderthal genome suggest some interbreeding between Neanderthals and early modern humans.

Green and his colleagues also discovered DNA segments among people in Europe and Asia that are more similar to Neanderthal sequences than to other contemporary human sequences. Another interesting observation was that Neanderthals are as closely related to people from Papua New Guinea as to those from China or France. This is surprising because Neanderthal fossil remains have been located only in Europe and West Asia. Most likely, genetic exchange took place between Neanderthals and modern humans as modern humans emerged out of Africa, before the divergence of Europeans, East Asians, and Papua New Guineans.

Several genes seem to have undergone changes from Neanderthals during the evolution of present-day humans. These genes are involved in cranial structure, metabolism, skin morphology, and cognitive development. One of the genes that is of particular interest is *RUNX2*, which is different in modern day humans and Neanderthals. This gene is responsible for the prominent frontal bone, bell-shaped rib cage, and dental differences seen in Neanderthals. It is speculated that an evolutionary change in *RUNX2* was important in the origin of modern-day humans, and this affected the cranium and the upper body.

LINK TO LEARNING

LINK TO LEARNING

Watch Svante Pääbo's talk explaining the Neanderthal genome research at the 2011 annual TED (Technology, Entertainment, Design) conference.

DNA PACKAGING IN CELLS

Prokaryotes are much simpler than eukaryotes in many of their features (Figure 14.10). Most prokaryotes contain a single, circular chromosome that is found in an area of the cytoplasm called the *nucleoid region*.

VISUAL CONNECTION

VISUAL CONNECTION

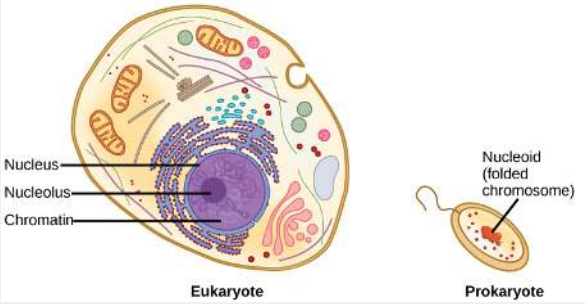


Figure 14.10 A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

In eukaryotic cells, DNA and RNA synthesis occur in a separate compartment from protein synthesis. In prokaryotic cells, both processes occur together. What advantages might there be to separating the processes? What advantages might there be to having them occur together?

The size of the genome in one of the most well-studied prokaryotes, *E.coli*, is 4.6 million base pairs (approximately 1.1 mm, if cut and stretched out). So how does this fit inside a small bacterial cell? The DNA is twisted by what is known as supercoiling. Supercoiling suggests that DNA is either "under-wound" (less than one turn of the helix per 10 base pairs) or "over-wound" (more than 1 turn per 10 base pairs) from its normal relaxed state. Some proteins are known to be involved in the supercoiling; other proteins and enzymes such as DNA gyrase help in maintaining the supercoiled structure.

Eukaryotes, whose chromosomes each consist of a linear DNA molecule, employ a different type of packing strategy to fit their DNA inside the nucleus (Figure 14.11). At the most basic level, DNA is wrapped around proteins known as **histones** to form structures called nucleosomes. The histones are evolutionarily conserved proteins that are rich in basic amino acids and form an octamer composed of two molecules of each of four different histones. Their composition and properties are important to understanding gene expression, and were partially uncovered based on research by Marie M. Daly and Alfred E. Mirsky in the early 1950s. The DNA (remember, it is negatively charged because of the phosphate groups) is wrapped tightly around the histone core. This nucleosome is linked to the next one with the help of a *linker DNA*. This is also known as the "beads on a string" structure. With the help of a fifth histone, a string of nucleosomes is further compacted into a 30-nm fiber, which is the diameter of the structure. Metaphase chromosomes are even further condensed by association with scaffolding proteins. At the metaphase stage, the chromosomes are at their most compact, approximately 700 nm in width.

In interphase, eukaryotic chromosomes have two distinct regions that can be distinguished by staining. The tightly packaged region is known as heterochromatin, and the less dense region is known as euchromatin. Heterochromatin usually contains genes that are not expressed, and is found in the regions of the centromere and telomeres. The euchromatin usually contains genes that are transcribed, with DNA packaged around nucleosomes but not further compacted.

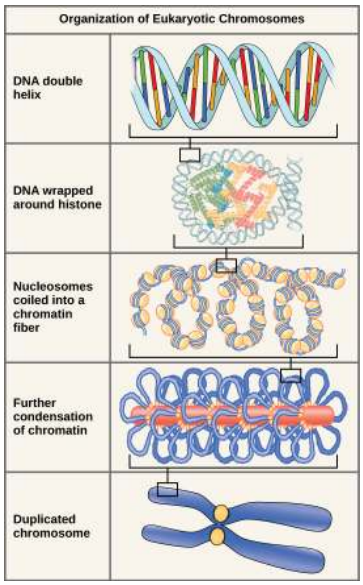


Figure 14.11 These figures illustrate the compaction of the eukaryotic chromosome.

FOOTNOTES

- 1Richard E. Green et al., “A Draft Sequence of the Neandertal Genome,” *Science* 328 (2010): 710-22.

This page titled 7.3: DNA Structure and Sequencing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.4: BASICS OF DNA REPLICATION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Explain how the structure of DNA reveals the replication process
- Describe the Meselson and Stahl experiments

The elucidation of the structure of the double helix provided a hint as to how DNA divides and makes copies of itself. In their 1953 paper, Watson and Crick penned an incredible understatement: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." With specific base pairs, the sequence of one DNA strand can be predicted from its complement. The double-helix model suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied. What was not clear was how the replication took place. There were three models suggested (Figure 14.12): *conservative, semi-conservative, and dispersive*.

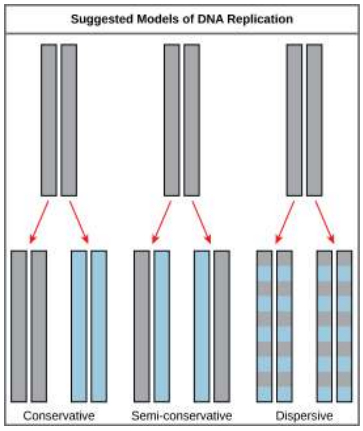


Figure 14.12 The three suggested models of DNA replication. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

In conservative replication, the parental DNA remains together, and the newly formed daughter strands are together. The semi-conservative method suggests that each of the two parental DNA strands acts as a template for new DNA to be synthesized; after replication, each double-stranded DNA includes one parental or “old” strand and one “new” strand. In the dispersive model, both copies of DNA have double-stranded segments of parental DNA and newly synthesized DNA interspersed.

Meselson and Stahl were interested in understanding how DNA replicates. They grew *E. coli* for several generations in a medium containing a “heavy” isotope of nitrogen (^{15}N), which gets incorporated into nitrogenous bases, and eventually into the DNA (Figure 14.13).

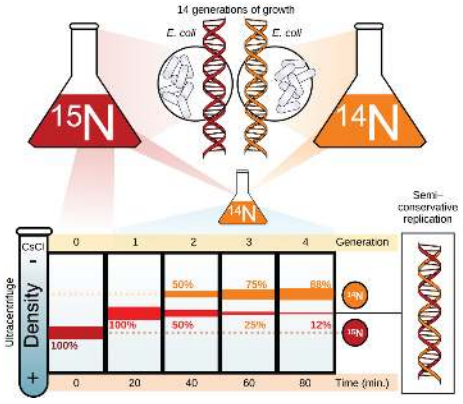


Figure 14.13 Meselson and Stahl experimented with *E. coli* grown first in heavy nitrogen (^{15}N) then in ^{14}N. DNA grown in ^{15}N (red band) is heavier than DNA grown in ^{14}N (orange band), and sediments to a lower level in cesium chloride solution in an ultracentrifuge. When DNA grown in ^{15}N is switched to media containing ^{14}N, after one round of cell division the DNA sediments halfway between the ^{15}N and ^{14}N levels, indicating that it now contains fifty percent ^{14}N. In subsequent cell divisions, an increasing amount of DNA contains ^{14}N only. These data support the semi-conservative replication model. (credit: modification of work by Mariana Ruiz Villareal)

The *E. coli* culture was then placed into medium containing ^{14}N and allowed to grow for several generations. After each of the first few generations, the cells were harvested and the DNA was isolated, then centrifuged at high speeds in an ultracentrifuge. During the centrifugation, the DNA was loaded into a *gradient* (typically a solution of salt such as cesium chloride or sucrose) and spun at high speeds of 50,000 to 60,000 rpm. Under these circumstances, the DNA will form a band according to its *buoyant density*: the density within the gradient at which it floats. DNA grown in ^{15}N will form a band at a higher density position (i.e., farther down the centrifuge tube) than that grown in ^{14}N. Meselson and Stahl noted that after one generation of growth in ^{14}N after they had been shifted from ^{15}N, the single band observed was intermediate in position in between DNA of cells grown exclusively in ^{15}N and ^{14}N. This suggested either a semi-conservative or dispersive mode of replication. The DNA harvested from cells grown for two generations in ^{14}N formed two bands: one DNA band was at the intermediate position between ^{15}N and ^{14}N, and the other corresponded to the band of ^{14}N DNA. These results could only be explained if DNA replicates in a semi-conservative manner. And for this reason, therefore, the other two models were ruled out.

During DNA replication, each of the two strands that make up the double helix serves as a template from which new strands are copied. The new strands will be complementary to the parental or “old” strands. When two daughter DNA copies are formed, they have the same sequence and are divided equally into the two daughter cells.

LINK TO LEARNING

LINK TO LEARNING

View this video on DNA replication.

This page titled 7.4: Basics of DNA Replication is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.5: DNA REPLICATION IN PROKARYOTES

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Explain the process of DNA replication in prokaryotes
- Discuss the role of different enzymes and proteins in supporting this process

DNA replication has been well studied in prokaryotes primarily because of the small size of the genome and because of the large variety of mutants that are available. *E. coli* has 4.6 million base pairs in a single circular chromosome and all of it gets replicated in approximately 42 minutes, starting from a single site along the chromosome and proceeding around the circle in both directions. This means that approximately 1000 nucleotides are added per second. Thus, the process is quite rapid and occurs without many mistakes.

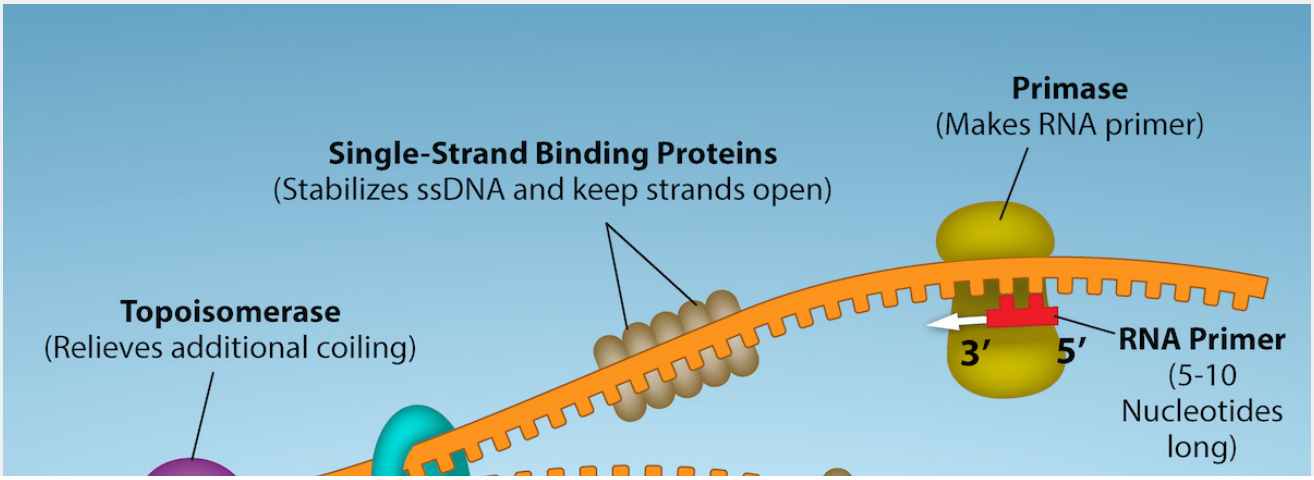
DNA replication employs a large number of structural proteins and enzymes, each of which plays a critical role during the process. One of the key players is the enzyme **DNA polymerase**, also known as DNA pol, which adds nucleotides one-by-one to the growing DNA chain that is complementary to the template strand. The addition of nucleotides requires energy; this energy is obtained from the nucleoside triphosphates ATP, GTP, TTP and CTP. Like ATP, the other **NTPs** (nucleoside triphosphates) are high-energy molecules that can serve both as the source of DNA nucleotides and the source of energy to drive the polymerization. When the bond between the phosphates is "broken," the energy released is used to form the phosphodiester bond between the incoming nucleotide and the growing chain. In prokaryotes, three main types of polymerases are known: DNA pol I, DNA pol II, and DNA pol III. It is now known that DNA pol III is the enzyme required for DNA synthesis; DNA pol I is an important accessory enzyme in DNA replication, and along with DNA pol II, is primarily required for repair.

How does the replication machinery know where to begin? It turns out that there are specific nucleotide sequences called *origins of replication* where replication begins. In *E. coli*, which has a single origin of replication on its one chromosome (as do most prokaryotes), this origin of replication is approximately 245 base pairs long and is rich in AT sequences. The origin of replication is recognized by certain proteins that bind to this site. An enzyme called *helicase* unwinds the DNA by breaking the hydrogen bonds between the nitrogenous base pairs. ATP hydrolysis is required for this process. As the DNA opens up, Y-shaped structures called *replication forks* are formed. Two replication forks are formed at the origin of replication and these get extended bi-directionally as replication proceeds. Single-strand binding proteins coat the single strands of DNA near the replication fork to prevent the single-stranded DNA from winding back into a double helix.

DNA polymerase has two important restrictions: it is able to add nucleotides only in the 5' to 3' direction (a new DNA strand can be only extended in this direction). It also requires a free 3'-OH group to which it can add nucleotides by forming a phosphodiester bond between the 3'-OH end and the 5' phosphate of the next nucleotide. This essentially means that it cannot add nucleotides if a free 3'-OH group is not available. Then how does it add the first nucleotide? The problem is solved with the help of a primer that provides the free 3'-OH end. Another enzyme, RNA primase, synthesizes an RNA segment that is about five to ten nucleotides long and complementary to the template DNA. Because this sequence primes the DNA synthesis, it is appropriately called the primer. DNA polymerase can now extend this RNA primer, adding nucleotides one-by-one that are complementary to the template strand (Figure 14.14).

VISUAL CONNECTION

VISUAL CONNECTION



https://bio.libretexts.org/@go/page/143279

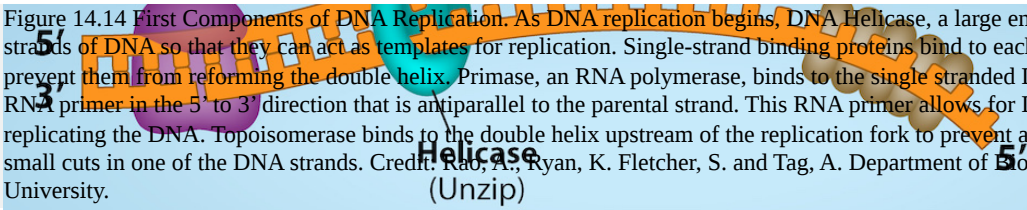


Figure 14.14 First Components of DNA Replication. As DNA replication begins, DNA Helicase, a large enzyme, separates the two strands of DNA so that they can act as templates for replication. Single-strand binding proteins bind to each strand to stabilize and prevent them from reforming the double helix. Primase, an RNA polymerase, binds to the single stranded DNA and synthesizes a short RNA primer in the 5’ to 3’ direction that is antiparallel to the parental strand. This RNA primer allows for DNA polymerase to begin replicating the DNA. Topoisomerase binds to the double helix upstream of the replication fork to prevent additional coiling by making small cuts in one of the DNA strands. Credit: Rao, A., Ryan, K. Fletcher, S. and Tag, A. Department of Biology, Texas A&M University.

Question: You isolate a cell strain in which the joining of Okazaki fragments is impaired and suspect that a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

The replication fork moves at the rate of 1000 nucleotides per second. Topoisomerase prevents the over-winding of the DNA double helix ahead of the replication fork as the DNA is opening up; it does so by causing temporary nicks in the DNA helix and then resealing it. Because DNA polymerase can only extend in the 5' to 3' direction, and because the DNA double helix is *antiparallel*, there is a slight problem at the replication fork. *The two template DNA strands have opposing orientations:* one strand is in the 5' to 3' direction and the other is oriented in the 3' to 5' direction. Only one new DNA strand, the one that is complementary to the 3' to 5' parental DNA strand, can be synthesized continuously towards the replication fork. This continuously synthesized strand is known as the leading strand. The other strand, complementary to the 5' to 3' parental DNA, is extended away from the replication fork, in small fragments known as Okazaki fragments, each requiring a primer to start the synthesis. New primer segments are laid down in the direction of the replication fork, but each pointing away from it. (Okazaki fragments are named after the Japanese scientist who first discovered them. The strand with the Okazaki fragments is known as the lagging strand.)

The leading strand can be extended from a single primer, whereas the lagging strand needs a new primer for each of the short Okazaki fragments. The overall direction of the lagging strand will be 3' to 5', and that of the leading strand 5' to 3'. A protein called the sliding clamp holds the DNA polymerase in place as it continues to add nucleotides. The sliding clamp is a ring-shaped protein that binds to the DNA and holds the polymerase in place. As synthesis proceeds, the RNA primers are replaced by DNA. The primers are removed by the exonuclease activity of DNA pol I, which uses DNA behind the RNA as its own primer and fills in the gaps left by removal of the RNA nucleotides by the addition of DNA nucleotides. The nicks that remain between the newly synthesized DNA (that replaced the RNA primer) and the previously synthesized DNA are sealed by the enzyme DNA ligase, which catalyzes the formation of phosphodiester linkages between the 3'-OH end of one nucleotide and the 5' phosphate end of the other fragment.

Once the chromosome has been completely replicated, the two DNA copies move into two different cells during cell division.

The process of DNA replication can be summarized as follows:

1. DNA unwinds at the origin of replication.
2. Helicase opens up the DNA-forming replication forks; these are extended bidirectionally.
3. Single-strand binding proteins coat the DNA around the replication fork to prevent rewinding of the DNA.
4. Topoisomerase binds at the region ahead of the replication fork to prevent supercoiling.
5. Primase synthesizes RNA primers complementary to the DNA strand.
6. DNA polymerase III starts adding nucleotides to the 3'-OH end of the primer.
7. Elongation of both the lagging and the leading strand continues.
8. RNA primers are removed by exonuclease activity.
9. Gaps are filled by DNA pol I by adding dNTPs.
10. The gap between the two DNA fragments is sealed by DNA ligase, which helps in the formation of phosphodiester bonds.

Table 14.1 summarizes the enzymes involved in prokaryotic DNA replication and the functions of each.

Prokaryotic DNA Replication: Enzymes and Their Function

Enzyme/protein	Specific Function
DNA pol I	Removes RNA primer and replaces it with newly synthesized DNA
DNA pol III	Main enzyme that adds nucleotides in the 5'-3' direction
Helicase	Opens the DNA helix by breaking hydrogen bonds between the nitrogenous bases
Ligase	Seals the gaps between the Okazaki fragments to create one continuous DNA strand
Primase	Synthesizes RNA primers needed to start replication
Sliding Clamp	Helps to hold the DNA polymerase in place when nucleotides are being added
Topoisomerase	Helps relieve the strain on DNA when unwinding by causing breaks, and then resealing the DNA
Single-strand binding proteins (SSB)	Binds to single-stranded DNA to prevent DNA from rewinding back.

Table 14.1

LINK TO LEARNING

LINK TO LEARNING

Review the full process of DNA replication here.

This page titled 7.5: DNA Replication in Prokaryotes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.6: DNA REPLICATION IN EUKARYOTES

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Discuss the similarities and differences between DNA replication in eukaryotes and prokaryotes
- State the role of telomerase in DNA replication

Eukaryotic genomes are much more complex and larger in size than prokaryotic genomes. Eukaryotes also have a number of different linear chromosomes. The human genome has 3 billion base pairs per haploid set of chromosomes, and 6 billion base pairs are replicated during the S phase of the cell cycle. There are multiple origins of replication on each eukaryotic chromosome; humans can have up to 100,000 origins of replication across the genome. The rate of replication is approximately 100 nucleotides per second, much slower than prokaryotic replication. In yeast, which is a eukaryote, special sequences known as autonomously replicating sequences (ARS) are found on the chromosomes. These are equivalent to the origin of replication in *E. coli*.

The number of DNA polymerases in eukaryotes is much more than in prokaryotes: 14 are known, of which five are known to have major roles during replication and have been well studied. They are known as pol *α*, pol *β*, pol *γ*, pol *δ*, and pol *ε*.

The essential steps of replication are the same as in prokaryotes. Before replication can start, the DNA has to be made available as a template. Eukaryotic DNA is bound to basic proteins known as histones to form structures called nucleosomes. Histones must be removed and then replaced during the replication process, which helps to account for the lower replication rate in eukaryotes. The chromatin (the complex between DNA and proteins) may undergo some chemical modifications, so that the DNA may be able to slide off the proteins or be accessible to the enzymes of the DNA replication machinery. At the origin of replication, a pre-replication complex is made with other initiator proteins. Helicase and other proteins are then recruited to start the replication process (Table 14.2).

Difference between Prokaryotic and Eukaryotic Replication

Property	Prokaryotes	Eukaryotes
Origin of replication	Single	Multiple
Rate of replication	1000 nucleotides/s	50 to 100 nucleotides/s
DNA polymerase types	5	14
Telomerase	Not present	Present
RNA primer removal	DNA pol I	RNase H
Strand elongation	DNA pol III	Pol α, pol δ, pol ε
Sliding clamp	Sliding clamp	PCNA

Table 14.2

A helicase using the energy from ATP hydrolysis opens up the DNA helix. Replication forks are formed at each replication origin as the DNA unwinds. The opening of the double helix causes over-winding, or supercoiling, in the DNA ahead of the replication fork. These are resolved with the action of topoisomerases. Primers are formed by the enzyme primase, and using the primer, DNA pol can start synthesis. Three major DNA polymerases are then involved: α, δ and ε. DNA pol α adds a short (20 to 30 nucleotides) DNA fragment to the RNA primer on both strands, and then hands off to a second polymerase. While the leading strand is continuously synthesized by the enzyme pol *ε*, the lagging strand is synthesized by pol *δ*. A sliding clamp protein known as PCNA (proliferating cell nuclear antigen) holds the DNA pol in place so that it does not slide off the DNA. As pol δ runs into the primer RNA on the lagging strand, it displaces it from the DNA template. The displaced primer RNA is then removed by RNase H (AKA flap endonuclease) and replaced with DNA nucleotides. The Okazaki fragments in the lagging strand are joined after the replacement of the RNA primers with DNA. The gaps that remain are sealed by DNA ligase, which forms the phosphodiester bond.

TELOMERE REPLICATION

Unlike prokaryotic chromosomes, eukaryotic chromosomes are linear. As you've learned, the enzyme DNA pol can add nucleotides only in the 5' to 3' direction. In the leading strand, synthesis continues until the end of the chromosome is reached. On the lagging strand, DNA is synthesized in short stretches, each of which is initiated by a separate primer. When the replication fork reaches the end of the linear chromosome, there is no way to replace the primer on the 5' end of the lagging strand. The DNA at the ends of the chromosome thus remains unpaired, and over time these ends, called **telomeres,** may get progressively shorter as cells continue to divide.

Telomeres comprise repetitive sequences that code for no particular gene. In humans, a six-base-pair sequence, TTAGGG, is repeated 100 to 1000 times in the telomere regions. In a way, these telomeres protect the genes from getting deleted as cells continue to divide. The telomeres are added to the ends of chromosomes by a separate enzyme, telomerase (Figure 14.16), whose discovery helped in the understanding of how these repetitive chromosome ends are maintained. The telomerase enzyme contains a catalytic part and a built-in RNA template. It attaches to the end of the chromosome, and DNA nucleotides complementary to the RNA template are added on the 3' end of the DNA strand. Once the 3' end of the lagging strand template is sufficiently elongated, DNA polymerase can add the nucleotides complementary to the ends of the chromosomes. Thus, the ends of the chromosomes are replicated.

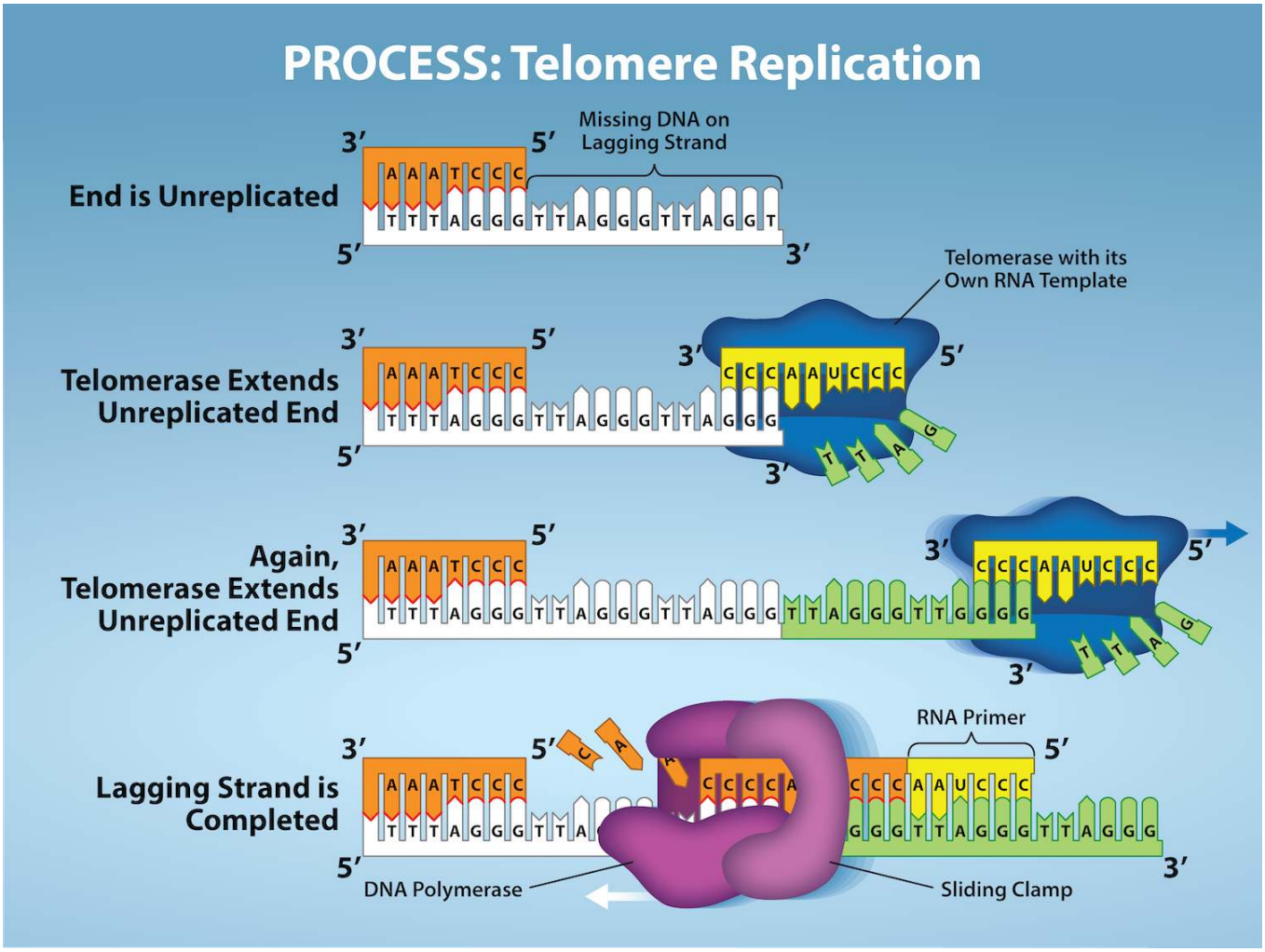


Figure 14.15 The ends of linear chromosomes are maintained by the action of the telomerase enzyme. Credit: Rao, A. and Fletcher, S. Department of Biology, Texas A&M University.

Telomerase is typically active in germ cells and adult stem cells. It is not active in adult somatic cells. For their discovery of telomerase and its action, Elizabeth Blackburn, Carol W. Greider, and Jack W. Szostak (Figure 14.16) received the Nobel Prize for Medicine and Physiology in 2009. Later research using HeLa cells (obtained from Henrietta Lacks) confirmed that telomerase is present in human cells. And in 2001, researchers including Diane L. Wright found that telomerase is necessary for cells in human embryos to rapidly proliferate.



Figure 14.16 Elizabeth Blackburn, 2009 Nobel Laureate, is one of the scientists who discovered how telomerase works. (credit: US Embassy Sweden)

TELOMERASE AND AGING

Cells that undergo cell division continue to have their telomeres shortened because most somatic cells do not make telomerase. This essentially means that telomere shortening is associated with aging. With the advent of modern medicine, preventative health care, and healthier lifestyles, the human life span has increased, and there is an increasing demand for people to look younger and have a better quality of life as they grow older.

In 2010, scientists found that telomerase can reverse some age-related conditions in mice. This may have potential in regenerative medicine.[2] Telomerase-deficient mice were used in these studies; these mice have tissue atrophy, stem cell depletion, organ system failure, and impaired tissue injury responses. Telomerase reactivation in these mice caused extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved the function of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

Cancer is characterized by uncontrolled cell division of abnormal cells. The cells accumulate mutations, proliferate uncontrollably, and can migrate to different parts of the body through a process called metastasis. Scientists have observed that cancerous cells have considerably shortened telomeres and that telomerase is active in these cells. Interestingly, only after the telomeres were shortened in the cancer cells did the telomerase become active. If the action of telomerase in these cells can be inhibited by drugs during cancer therapy, then the cancerous cells could potentially be stopped from further division.

FOOTNOTES

- 2Jaskelioff et al., “Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice,” *Nature* 469 (2011): 102-7.

This page titled 7.6: DNA Replication in Eukaryotes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.7: DNA REPAIR

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Discuss the different types of mutations in DNA
- Explain DNA repair mechanisms

DNA replication is a highly accurate process, but mistakes can occasionally occur, such as a DNA polymerase inserting a wrong base. Uncorrected mistakes may sometimes lead to serious consequences, such as cancer. Repair mechanisms correct the mistakes. In rare cases, mistakes are not corrected, leading to mutations; in other cases, repair enzymes are themselves mutated or defective.

Most of the mistakes during DNA replication are promptly corrected by the proofreading ability of DNA polymerase itself. (Figure 14.17). In proofreading, the DNA pol reads the newly added base before adding the next one, so a correction can be made. The polymerase checks whether the newly added base has paired correctly with the base in the template strand. If it is the right base, the next nucleotide is added. If an incorrect base has been added, the enzyme makes a cut at the phosphodiester bond and releases the wrong nucleotide. This is performed by the 3' exonuclease action of DNA pol. Once the incorrect nucleotide has been removed, it can be replaced by the correct one.

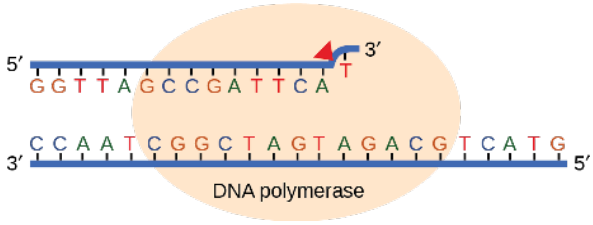


Figure 14.17 Proofreading by DNA polymerase corrects errors during replication.

Some errors are not corrected during replication, but are instead corrected after replication is completed; this type of repair is known as mismatch repair (Figure 14.18). Specific repair enzymes recognize the mispaired nucleotide and excise part of the strand that contains it; the excised region is then resynthesized. If the mismatch remains uncorrected, it may lead to more permanent damage when the mismatched DNA is replicated. How do mismatch repair enzymes recognize which of the two bases is the incorrect one? In *E. coli*, after replication, the nitrogenous base adenine acquires a methyl group; the parental DNA strand will have methyl groups, whereas the newly synthesized strand lacks them. Thus, DNA polymerase is able to remove the wrongly incorporated bases from the newly synthesized, non-methylated strand. In eukaryotes, the mechanism is not very well understood, but it is believed to involve recognition of unsealed nicks in the new strand, as well as a short-term continuing association of some of the replication proteins with the new daughter strand after replication has completed.

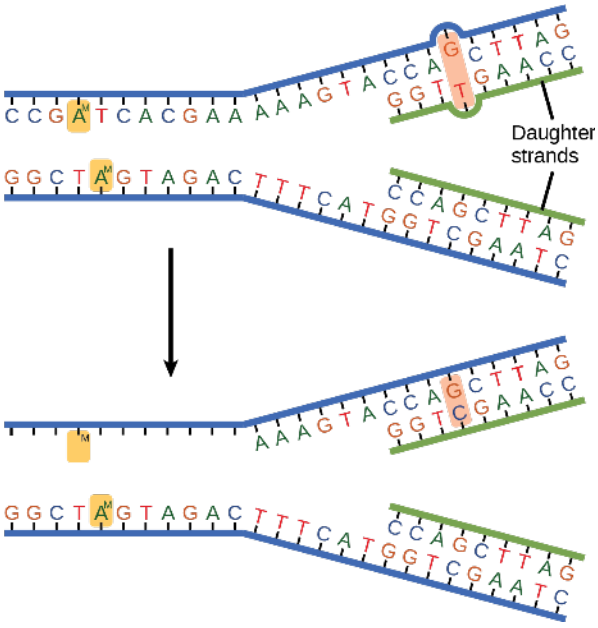


Figure 14.18 In mismatch repair, the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base.

Another type of repair mechanism, nucleotide excision repair, is similar to mismatch repair, except that it is used to remove damaged bases rather than mismatched ones. The repair enzymes replace abnormal bases by making a cut on both the 3' and 5' ends of the damaged base (Figure 14.19). The segment of DNA is removed and replaced with the correctly paired nucleotides by the action of DNA pol. Once the bases are filled in, the remaining gap is sealed with a phosphodiester linkage catalyzed by DNA ligase. This repair mechanism is often employed when UV exposure causes the formation of pyrimidine dimers.

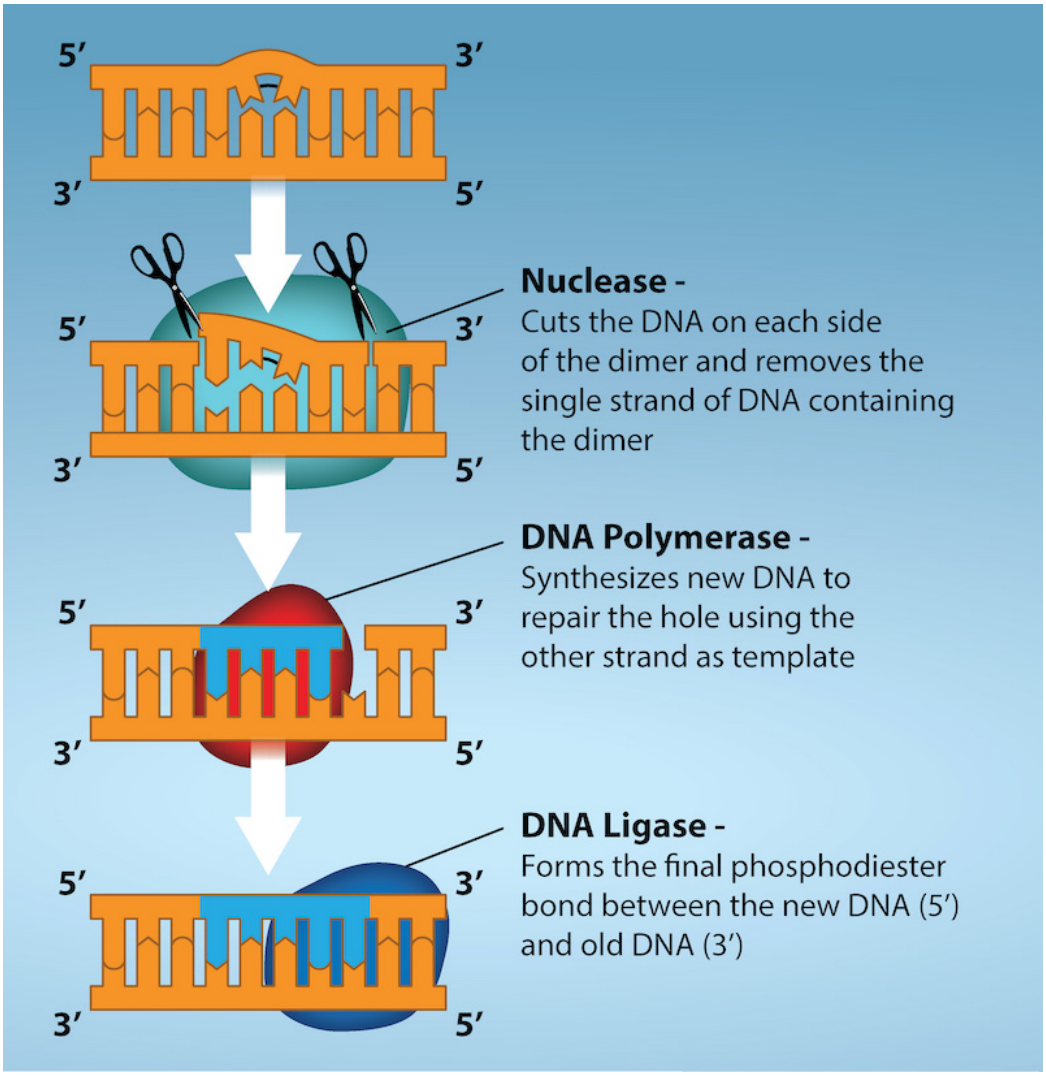


Figure 14.19 Nucleotide excision repairs thymine dimers. When exposed to UV light, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced. Credit: Rao, A., Fletcher, S. and Tag, A. Department of Biology, Texas A&M University.

A well-studied example of mistakes not being corrected is seen in people suffering from xeroderma pigmentosa (Figure 14.20). Affected individuals have skin that is highly sensitive to UV rays from the sun. When individuals are exposed to UV light, pyrimidine dimers, especially those of thymine, are formed; people with xeroderma pigmentosa are not able to repair the damage. These are not repaired because of a defect in the nucleotide excision repair enzymes, whereas in normal individuals, the thymine dimers are excised and the defect is corrected. The thymine dimers distort the structure of the DNA double helix, and this may cause problems during DNA replication. People with xeroderma pigmentosa may have a higher risk of contracting skin cancer than those who don't have the condition.



Figure 14.20 Xeroderma pigmentosa is a condition in which thymine dimerization from exposure to UV light is not repaired. Exposure to sunlight results in skin lesions. (credit: James Halpern et al.)

Errors during DNA replication are not the only reason why mutations arise in DNA. Mutations, variations in the nucleotide sequence of a genome, can also occur because of damage to DNA. Such mutations may be of two types: induced or spontaneous. Induced mutations are those that result from an exposure to chemicals, UV rays, x-rays, or some other environmental agent. For example, Charlotte Auerbach and J.M Robson discovered the mutation-inducing effects of mustard gas. Spontaneous mutations occur without any exposure to any environmental agent; they are a result of natural reactions taking place within the body.

Mutations may have a wide range of effects. Point mutations are those mutations that affect a single base pair. The most common nucleotide mutations are substitutions, in which one base is replaced by another. These substitutions can be of two types, either transitions or transversions. Transition substitution refers to a purine or pyrimidine being replaced by a base of the same kind; for example, a purine such as adenine may be replaced by the purine guanine. Transversion substitution refers to a purine being replaced by a pyrimidine, or vice versa; for example, cytosine, a pyrimidine, is replaced by adenine, a purine. Some point mutations are not detectable in the final product; these are known as silent mutations. Silent mutations are usually due to a substitution in the third base of a codon, which often represents the same amino acid as the original codon. Other point mutations can result in the replacement of one amino acid by another, which may alter the function of the protein. Point mutations that generate a stop codon can terminate a protein early.

Some mutations can result in an increased number of copies of the same codon. These are called trinucleotide repeat expansions and result in repeated regions of the same amino acid. Mutations can also be the result of the addition of a base, known as an insertion, or the removal of a base, also known as deletion. If an insertion or deletion results in the alteration of the translational reading frame (a frameshift mutation), the resultant protein is usually nonfunctional. Sometimes a piece of DNA from one chromosome may get translocated to another chromosome or to another region of the same chromosome; this is also known as translocation. These mutation types are shown in Figure 14.21.

VISUAL CONNECTION

VISUAL CONNECTION

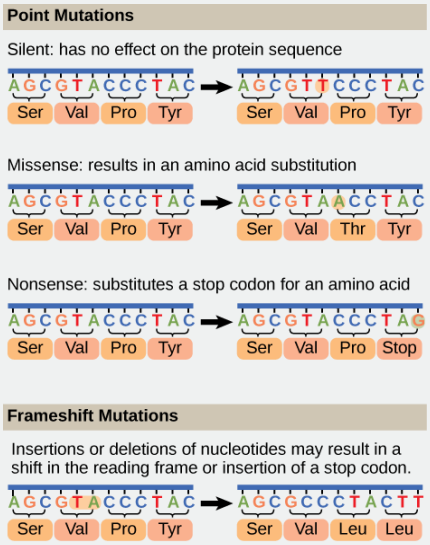


Figure 14.21 Mutations can lead to changes in the protein sequence encoded by the DNA.

A frameshift mutation that results in the insertion of three nucleotides is often less deleterious than a mutation that results in the insertion of one nucleotide. Why?

Mutations in repair genes have been known to cause cancer. Many mutated repair genes have been implicated in certain forms of pancreatic cancer, colon cancer, and colorectal cancer. Mutations can affect either somatic cells or germ cells. If many mutations accumulate in a somatic cell, they may lead to problems such as the uncontrolled cell division observed in cancer. If a mutation takes place in germ cells, the mutation will be passed on to the next generation, as in the case of hemophilia and xeroderma pigmentosa.

This page titled 7.7: DNA Repair is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

7.8: KEY TERMS

electrophoresis
technique used to separate DNA fragments according to size

helicase
during replication, this enzyme helps to open up the DNA helix by breaking the hydrogen bonds

induced mutation
mutation that results from exposure to chemicals or environmental agents

lagging strand
during replication, the strand that is replicated in short fragments and away from the replication fork

leading strand
strand that is synthesized continuously in the 5'-3' direction, which is synthesized in the direction of the replication fork

ligase
enzyme that catalyzes the formation of a phosphodiester linkage between the 3' OH and 5' phosphate ends of the DNA

mismatch repair
type of repair mechanism in which mismatched bases are removed after replication

mutation
variation in the nucleotide sequence of a genome

nucleotide excision repair
type of DNA repair mechanism in which the wrong base, along with a few nucleotides upstream or downstream, are removed

Okazaki fragment
DNA fragment that is synthesized in short stretches on the lagging strand

point mutation
mutation that affects a single base

primase
enzyme that synthesizes the RNA primer; the primer is needed for DNA pol to start synthesis of a new DNA strand

primer
short stretch of nucleotides that is required to initiate replication; in the case of replication, the primer has RNA nucleotides

proofreading
function of DNA pol in which it reads the newly added base before adding the next one

replication fork
Y-shaped structure formed during initiation of replication

silent mutation
mutation that is not expressed

single-strand binding protein
during replication, protein that binds to the single-stranded DNA; this helps in keeping the two strands of DNA apart so that they may serve as templates

sliding clamp
ring-shaped protein that holds the DNA pol on the DNA strand

spontaneous mutation
: mutation that takes place in the cells as a result of chemical reactions taking place naturally without exposure to any external agent

telomerase
: enzyme that contains a catalytic part and an inbuilt RNA template; it functions to maintain telomeres at chromosome ends

telomere
: DNA at the end of linear chromosomes

topoisomerase
: enzyme that prevents overwinding of DNA when DNA replication is taking place

transformation
: process in which external DNA is taken up by a cell

transition substitution
: when a purine is replaced with a purine or a pyrimidine is replaced with another pyrimidine

transversion substitution
: when a purine is replaced by a pyrimidine or a pyrimidine is replaced by a purine

---

This page titled 7.8: Key Terms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu.

7.9: CHAPTER SUMMARY

14.1 HISTORICAL BASIS OF MODERN UNDERSTANDING

DNA was first isolated from white blood cells by Friedrich Miescher, who called it nuclein because it was isolated from nuclei. Frederick Griffith's experiments with strains of *Streptococcus pneumoniae* provided the first hint that DNA may be the transforming principle. Avery, MacLeod, and McCarty showed that DNA is required for the transformation of bacteria. Later experiments by Hershey and Chase using bacteriophage T2 proved that DNA is the genetic material. Chargaff found that the ratio of A = T and C = G, and that the percentage content of A, T, G, and C is different for different species.

14.2 DNA STRUCTURE AND SEQUENCING

The currently accepted model of the double-helix structure of DNA was proposed by Watson and Crick. Some of the salient features are that the two strands that make up the double helix have complementary base sequences and anti-parallel orientations. Alternating deoxyribose sugars and phosphates form the backbone of the structure, and the nitrogenous bases are stacked like rungs inside. The diameter of the double helix, 2 nm, is uniform throughout. A purine always pairs with a pyrimidine; A pairs with T, and G pairs with C. One turn of the helix has 10 base pairs. Prokaryotes are much simpler than eukaryotes in many of their features. Most prokaryotes contain a single, circular chromosome. In general, eukaryotic chromosomes contain a linear DNA molecule packaged into nucleosomes, and have two distinct regions that can be distinguished by staining, reflecting different states of packaging and compaction.

14.3 BASICS OF DNA REPLICATION

During cell division, each daughter cell receives a copy of each molecule of DNA by a process known as DNA replication. The single chromosome of a prokaryote or each chromosome of a eukaryote consists of a single continuous double helix. The model for DNA replication suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied. In the conservative model of replication, the parental DNA is conserved, and the daughter DNA is newly synthesized. The semi-conservative model suggests that each of the two parental DNA strands acts as template for new DNA to be synthesized; after replication, each double-stranded DNA retains the parental or “old” strand and one “new” strand. The dispersive model suggested that the two copies of the DNA would have segments of parental DNA and newly synthesized DNA. The Meselson and Stahl experiment supported the semi-conservative model of replication, in which an entire replicated chromosome consists of one parental strand and one newly synthesized strand of DNA.

14.4 DNA REPLICATION IN PROKARYOTES

Replication in prokaryotes starts from a sequence found on the chromosome called the origin of replication—the point at which the DNA opens up. Helicase opens up the DNA double helix, resulting in the formation of the replication fork. Single-strand binding proteins bind to the single-stranded DNA near the replication fork to keep the fork open. Primase synthesizes an RNA primer to initiate synthesis by DNA polymerase, which can add nucleotides only to the 3' end of a previously synthesized primer strand. Both new DNA strands grow according to their respective 5'-3' directions. One strand is synthesized continuously in the direction of the replication fork; this is called the leading strand. The other strand is synthesized in a direction away from the replication fork, in short stretches of DNA known as Okazaki fragments. This strand is known as the lagging strand. Once replication is completed, the RNA primers are replaced by DNA nucleotides and the DNA is sealed with DNA ligase, which creates phosphodiester bonds between the 3'-OH of one end and the 5' phosphate of the other strand.

14.5 DNA REPLICATION IN EUKARYOTES

Replication in eukaryotes starts at multiple origins of replication. The mechanism is quite similar to that in prokaryotes. A primer is required to initiate synthesis, which is then extended by DNA polymerase as it adds nucleotides one by one to the growing chain. The leading strand is synthesized continuously, whereas the lagging strand is synthesized in short stretches called Okazaki fragments. The RNA primers are replaced with DNA nucleotides; the DNA Okazaki fragments are linked into one continuous strand by DNA ligase. The ends of the chromosomes pose a problem as the primer RNA at the 5’ ends of the DNA cannot be replaced with DNA, and the chromosome is progressively shortened. Telomerase, an enzyme with an inbuilt RNA template, extends the ends by copying the RNA template and extending one strand of the chromosome. DNA polymerase can then fill in the complementary DNA strand using the regular replication enzymes. In this way, the ends of the chromosomes are protected.

14.6 DNA REPAIR

DNA polymerase can make mistakes while adding nucleotides. It edits the DNA by proofreading every newly added base. Incorrect bases are removed and replaced by the correct base before proceeding with elongation. Most mistakes are corrected during replication, although when this does not happen, the mismatch repair mechanism is employed. Mismatch repair enzymes recognize the wrongly incorporated base and excise it from the DNA, replacing it with the correct base. In yet another type of repair, nucleotide excision repair, a damaged base

is removed along with a few bases on the 5' and 3' end, and these are replaced by copying the template with the help of DNA polymerase. The ends of the newly synthesized fragment are attached to the rest of the DNA using DNA ligase, which creates a phosphodiester bond.

Most mistakes are corrected, and if they are not, they may result in a mutation, defined as a permanent change in the DNA sequence. Mutations can be of many types, such as substitution, deletion, insertion, and trinucleotide repeat expansions. Mutations in repair genes may lead to serious consequences such as cancer. Mutations can be induced or may occur spontaneously.

This page titled 7.9: Chapter Summary is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

7.10: VISUAL CONNECTION QUESTIONS

1.

Figure 14.10 In eukaryotic cells, DNA and RNA synthesis occur in a separate compartment from protein synthesis. In prokaryotic cells, both processes occur together. What advantages might there be to separating the processes? What advantages might there be to having them occur together?

2.

Figure 14.14 You isolate a cell strain in which the joining of Okazaki fragments is impaired and suspect that a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

3.

Figure 14.21 A frameshift mutation that results in the insertion of three nucleotides is often less deleterious than a mutation that results in the insertion of one nucleotide. Why?

This page titled 7.10: Visual Connection Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

7.11: REVIEW QUESTIONS

4.

If DNA of a particular species was analyzed and it was found that it contains 27 percent A, what would be the percentage of C?

1. 27 percent
2. 30 percent
3. 23 percent
4. 54 percent

5.

The experiments by Hershey and Chase helped confirm that DNA was the hereditary material on the basis of the finding that:

1. radioactive phage were found in the pellet
2. radioactive cells were found in the supernatant
3. radioactive sulfur was found inside the cell
4. radioactive phosphorus was found in the cell

6.

Bacterial transformation is a major concern in many medical settings. Why might health care providers be concerned?

1. Pathogenic bacteria could introduce disease-causing genes in non-pathogenic bacteria.
2. Antibiotic resistance genes could be introduced to new bacteria to create “superbugs.”
3. Bacteriophages could spread DNA encoding toxins to new bacteria.
4. All of the above.

7.

DNA double helix does not have which of the following?

1. antiparallel configuration
2. complementary base pairing
3. major and minor grooves
4. uracil

8.

In eukaryotes, what is the DNA wrapped around?

1. single-stranded binding proteins
2. sliding clamp
3. polymerase
4. histones

9.

Meselson and Stahl's experiments proved that DNA replicates by which mode?

1. conservative
2. semi-conservative
3. dispersive
4. none of the above

10.

If the sequence of the 5'-3' strand is AATGCTAC, then the complementary sequence has the following sequence:

1. 3'-AATGCTAC-5'
2. 3'-CATCGTAA-5'
3. 3'-TTACGATG-5'
4. 3'-GTAGCATT-5'

11.

How did Meselson and Stahl support Watson and Crick’s double-helix model?

1. They demonstrated that each strand serves as a template for synthesizing a new strand of DNA.
2. They showed that the DNA strands break and recombine without losing genetic material.

3. They proved that DNA maintains a double-helix structure while undergoing semi-conservative replication.
4. They demonstrated that conservative replication maintains the complementary base pairing of each DNA helix.

12.

Which of the following components is not involved during the formation of the replication fork?

1. single-strand binding proteins
2. helicase
3. origin of replication
4. ligase

13.

Which of the following does the enzyme primase synthesize?

1. DNA primer
2. RNA primer
3. Okazaki fragments
4. phosphodiester linkage

14.

In which direction does DNA replication take place?

1. 5'-3'
2. 3'-5'
3. 5'
4. 3'

15.

A scientist randomly mutates the DNA of a bacterium. She then sequences the bacterium's daughter cells, and finds that the daughters have many errors in their replicated DNA. The parent bacterium likely acquired a mutation in which enzyme?

1. DNA ligase
2. DNA pol II
3. Primase
4. DNA pol I

16.

The ends of the linear chromosomes are maintained by

1. helicase
2. primase
3. DNA pol
4. telomerase

17.

Which of the following is not a true statement comparing prokaryotic and eukaryotic DNA replication?

1. Both eukaryotic and prokaryotic DNA polymerases build off RNA primers made by primase.
2. Eukaryotic DNA replication requires multiple replication forks, while prokaryotic replication uses a single origin to rapidly replicate the entire genome.
3. DNA replication always occurs in the nucleus.
4. Eukaryotic DNA replication involves more polymerases than prokaryotic replication.

18.

During proofreading, which of the following enzymes reads the DNA?

1. primase
2. topoisomerase
3. DNA pol
4. helicase

19.

The initial mechanism for repairing nucleotide errors in DNA is _____.

1. mismatch repair
2. DNA polymerase proofreading
3. nucleotide excision repair
4. thymine dimers

20.

A scientist creates fruit fly larvae with a mutation that eliminates the exonuclease function of DNA pol. Which prediction about the mutational load in the adult fruit flies is most likely to be correct?

1. The adults with the DNA pol mutation will have significantly more mutations than average.
2. The adults with the DNA pol mutation will have slightly more mutations than average.
3. The adults with the DNA pol mutation will have the same number of mutations as average.
4. The adults with the DNA pol mutation will have fewer mutations than average.

This page titled 7.11: Review Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

7.12: CRITICAL THINKING QUESTIONS

21.

Explain Griffith's transformation experiments. What did he conclude from them?

22.

Why were radioactive sulfur and phosphorous used to label bacteriophage in Hershey and Chase's experiments?

23.

When Chargaff was performing his experiments, the tetranucleotide hypothesis, which stated that DNA was composed of GACT nucleotide repeats, was the most widely accepted view of DNA's composition. How did Chargaff disprove this hypothesis?

24.

Provide a brief summary of the Sanger sequencing method.

25.

Describe the structure and complementary base pairing of DNA.

26.

Prokaryotes have a single circular chromosome while eukaryotes have linear chromosomes. Describe one advantage and one disadvantage to the eukaryotic genome packaging compared to the prokaryotes.

27.

How did the scientific community learn that DNA replication takes place in a semi-conservative fashion?

28.

Imagine the Meselson and Stahl experiments had supported conservative replication instead of semi-conservative replication. What results would you predict to observe after two rounds of replication? Be specific regarding percent distributions of DNA incorporating ^{15}N and ^{14}N in the gradient.

29.

DNA replication is bidirectional and discontinuous; explain your understanding of those concepts.

30.

What are Okazaki fragments and how they are formed?

31.

If the rate of replication in a particular prokaryote is 900 nucleotides per second, how long would it take 1.2 million base pair genomes to make two copies?

32.

Explain the events taking place at the replication fork. If the gene for helicase is mutated, what part of replication will be affected?

33.

What is the role of a primer in DNA replication? What would happen if you forgot to add a primer in a tube containing the reaction mix for a DNA sequencing reaction?

34.

Quinolone antibiotics treat bacterial infections by blocking the activity of topoisomerase. Why does this treatment work? Explain what occurs at the molecular level.

35.

How do the linear chromosomes in eukaryotes ensure that its ends are replicated completely?

36.

What is the consequence of mutation of a mismatch repair enzyme? How will this affect the function of a gene?

37.

An adult with a history of tanning has his genome sequenced. The beginning of a protein-coding region of his DNA reads ATGGGGATATGGCAT. If the protein-coding region of a healthy adult reads ATGGGGATATGAGCAT, identify the site and type of mutation.

This page titled 7.12: Critical Thinking Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

CHAPTER OVERVIEW

8: GENES AND PROTEINS

Since the rediscovery of Mendel's work in 1900, the definition of the gene has progressed from an abstract unit of heredity to a tangible molecular entity capable of replication, expression, and mutation. Genes are composed of DNA and are linearly arranged on chromosomes. Genes specify the sequences of amino acids, which are the building blocks of proteins. In turn, proteins are responsible for orchestrating nearly every function of the cell. Both genes and the proteins they encode are absolutely essential to life as we know it.

- 8.1: Introduction
- 8.2: The Genetic Code
- 8.3: Prokaryotic Transcription
- 8.4: Eukaryotic Transcription
- 8.5: RNA Processing in Eukaryotes
- 8.6: Ribosomes and Protein Synthesis
- 8.7: Gene Expression
  - 8.7.1: Introduction
  - 8.7.2: Regulation of Gene Expression
  - 8.7.3: Prokaryotic Gene Regulation
  - 8.7.4: Eukaryotic Epigenetic Gene Regulation
  - 8.7.5: Eukaryotic Transcription Gene Regulation
  - 8.7.6: Eukaryotic Post-transcriptional Gene Regulation
  - 8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
  - 8.7.8: Cancer and Gene Regulation
  - 8.7.9: Key Terms
  - 8.7.10: Chapter Summary
  - 8.7.11: Visual Connection Questions
  - 8.7.12: Review Questions
  - 8.7.13: Critical Thinking Questions
- 8.8: Key Terms
- 8.9: Chapter Summary
- 8.10: Visual Connection Questions
- 8.11: Regulation of Gene Expression
  - 8.11.1: The lac Operon
  - 8.11.2: The Use of Mutants to Study the lac Operon
  - 8.11.3: Eukaryotic Gene Regulation
  - 8.11.4: Regulatory Elements in Evolution
  - 8.11.5: Additional Levels of Regulating Transcription
  - 8.11.6: Epigenetics
  - 8.11.7: Regulation of Gene Expression (Exercises)
  - 8.11.8: Gene Expression Essentials
  - 8.11.S: Regulation of Gene Expression (Summary)
- 8.12: Review Questions
- 8.13: Critical Thinking Questions

This page titled 8: Genes and Proteins is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.1: INTRODUCTION

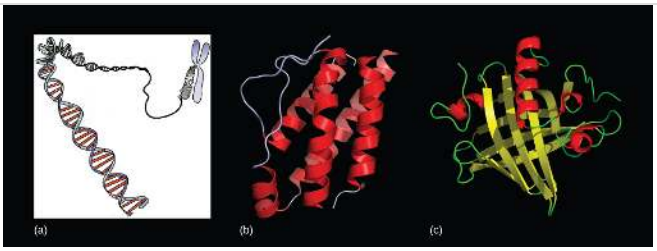


Figure 15.1 Genes, which are carried on (a) chromosomes, are linearly organized instructions for making the RNA and protein molecules that are necessary for all of the processes of life. The (b) interleukin-2 protein and (c) alpha-2u-globulin protein are just two examples of the array of different molecular structures that are encoded by genes. (credit “chromosome: National Human Genome Research Institute; credit “interleukin-2”: Ramin Herati/Created from PDB 1M47 and rendered with Pymol; credit “alpha-2u-globulin”: Darren Logan/rendered with AISMIG)

CHAPTER OUTLINE

15.1 The Genetic Code
15.2 Prokaryotic Transcription
15.3 Eukaryotic Transcription
15.4 RNA Processing in Eukaryotes
15.5 Ribosomes and Protein Synthesis

Since the rediscovery of Mendel’s work in 1900, the definition of the gene has progressed from an abstract unit of heredity to a tangible molecular entity capable of replication, expression, and mutation (Figure 15.1). Genes are composed of DNA and are linearly arranged on chromosomes. Genes specify the sequences of amino acids, which are the building blocks of proteins. In turn, proteins are responsible for orchestrating nearly every function of the cell. Both genes and the proteins they encode are absolutely essential to life as we know it.

This page titled 8.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.2: THE GENETIC CODE

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Explain the “central dogma” of DNA-protein synthesis
- Describe the genetic code and how the nucleotide sequence prescribes the amino acid and the protein sequence

The cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template on ribosomes converts nucleotide-based genetic information into a protein product. That is the central dogma of DNA-protein synthesis. Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 “letters” (Figure 15.2). Different amino acids have different chemistries (such as acidic versus basic, or polar and nonpolar) and different structural constraints. Variation in amino acid sequence is responsible for the enormous variation in protein structure and function.

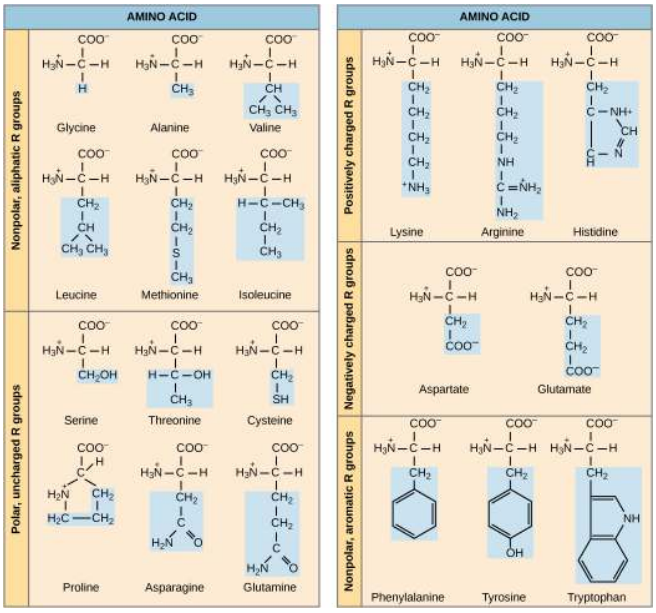
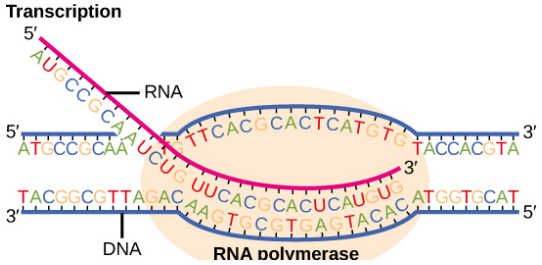


Figure 15.2 Structures of the 20 amino acids found in proteins are shown. Each amino acid is composed of an amino group (NH_3^+), a carboxyl group (COO^-), and a side chain (blue). The side chain may be nonpolar, polar, or charged, as well as large or small. It is the variety of amino acid side chains that gives rise to the incredible variation of protein structure and function.

THE CENTRAL DOGMA: DNA ENCODES RNA; RNA ENCODES PROTEIN

The flow of genetic information in cells from DNA to mRNA to protein is described by the central dogma (Figure 15.3), which states that genes specify the sequence of mRNAs, which in turn specify the sequence of amino acids making up all proteins. The decoding of one molecule to another is performed by specific proteins and RNAs. Because the information stored in DNA is so central to cellular function, it makes intuitive sense that the cell would make mRNA copies of this information for protein synthesis, while keeping the DNA itself intact and protected. The copying of DNA to RNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every nucleotide read in the DNA strand. The translation to protein is a bit more complex because three mRNA nucleotides correspond to one amino acid in the polypeptide sequence. However, the translation to protein is still systematic and colinear, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.



https://bio.libretexts.org/@go/page/143289

Figure 15.3 Instructions on DNA are transcribed onto messenger RNA. Ribosomes are able to read the genetic information inscribed on a strand of messenger RNA and use this information to string amino acids together into a protein.

THE GENETIC CODE IS DEGENERATE AND UNIVERSAL

Each amino acid is defined by a three-nucleotide sequence called the triplet codon. Given the different numbers of "letters" in the mRNA and protein "alphabets," scientists theorized that single amino acids must be represented by combinations of nucleotides. Nucleotide doublets would not be sufficient to specify every amino acid because there are only 16 possible two-nucleotide combinations (4^2). In contrast, there are 64 possible nucleotide triplets (4^3), which is far more than the number of amino acids. Scientists theorized that amino acids were encoded by nucleotide triplets and that the genetic code was "degenerate." In other words, a given amino acid could be encoded by more than one nucleotide triplet. This was later confirmed experimentally: Francis Crick and Sydney Brenner used the chemical mutagen proflavin to insert one, two, or three nucleotides into the gene of a virus. When one or two nucleotides were inserted, the normal proteins were not produced. When three nucleotides were inserted, the protein was synthesized and functional. This demonstrated that the amino acids must be specified by groups of three nucleotides. These nucleotide triplets are called codons. The insertion of one or two nucleotides completely changed the triplet reading frame, thereby altering the message for every subsequent amino acid (Figure 15.5). Though insertion of three nucleotides caused an extra amino acid to be inserted during translation, the integrity of the rest of the protein was maintained.

Scientists painstakingly solved the genetic code by translating synthetic mRNAs in vitro and sequencing the proteins they specified (Figure 15.4).

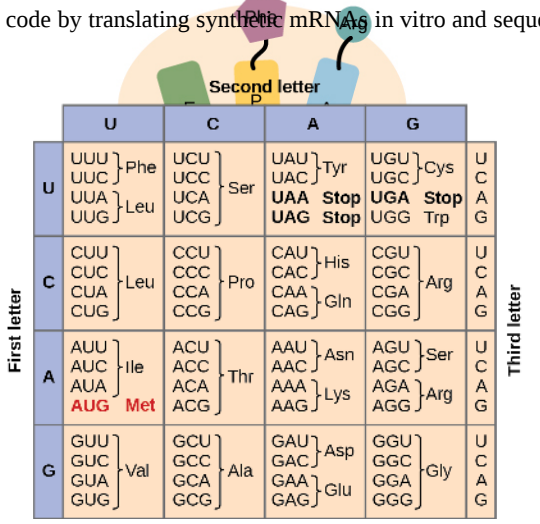


Figure 15.4 This figure shows the genetic code for translating each nucleotide triplet in mRNA into an amino acid or a termination signal in a protein. (credit: modification of work by NIH)

In addition to codons that instruct the addition of a specific amino acid to a polypeptide chain, three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called nonsense codons, or *stop codons*. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the start codon to initiate translation. The reading frame for translation is set by the AUG start codon near the 5' end of the mRNA. Following the start codon, the mRNA is read in groups of three until a stop codon is encountered.

The arrangement of the coding table reveals the structure of the code. There are sixteen "blocks" of codons, each specified by the first and second nucleotides of the codons within the block, e.g., the "AC*" block that corresponds to the amino acid threonine (Thr). Some blocks are divided into a pyrimidine half, in which the codon ends with U or C, and a purine half, in which the codon ends with A or G. Some amino acids get a whole block of four codons, like alanine (Ala), threonine (Thr) and proline (Pro). Some get the pyrimidine half of their block, like histidine (His) and asparagine (Asn). Others get the purine half of their block, like glutamate (Glu) and lysine (Lys). Note that some amino acids get a block and a half-block for a total of six codons.

The specification of a single amino acid by multiple similar codons is called "degeneracy." Degeneracy is believed to be a cellular mechanism to reduce the negative impact of random mutations. Codons that specify the same amino acid typically only differ by one nucleotide. In addition, amino acids with chemically similar side chains are encoded by similar codons. For example, aspartate (Asp) and glutamate (Glu), which occupy the GA* block, are both negatively charged. This nuance of the genetic code ensures that a single-nucleotide substitution mutation might specify the same amino acid but have no effect or specify a similar amino acid, preventing the protein from being rendered completely nonfunctional.

The genetic code is nearly universal. With a few minor exceptions, virtually all species use the same genetic code for protein synthesis. Conservation of codons means that a purified mRNA encoding the globin protein in horses could be transferred to a tulip cell, and the tulip

would synthesize horse globin. That there is only one genetic code is powerful evidence that all of life on Earth shares a common origin, especially considering that there are about 10^{84} possible combinations of 20 amino acids and 64 triplet codons.

LINK TO LEARNING

LINK TO LEARNING

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at this site.

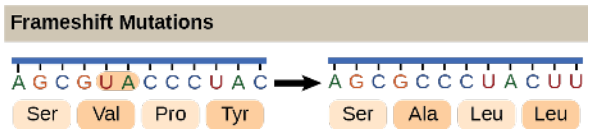


Figure 15.5 The deletion of two nucleotides shifts the reading frame of an mRNA and changes the entire protein message, creating a nonfunctional protein or terminating protein synthesis altogether.

SCIENTIFIC METHOD CONNECTION

SCIENTIFIC METHOD CONNECTION

Which Has More DNA: A Kiwi or a Strawberry?



Figure 15.6 Do you think that a kiwi or a strawberry has more DNA per fruit? (credit “kiwi”: "Kelbv"/Flickr; credit: “strawberry”: Alisdair McDiarmid)

Question: Would a kiwi and strawberry that are approximately the same size (Figure 15.6) also have approximately the same amount of DNA?

Background: Genes are carried on chromosomes and are made of DNA. All mammals are diploid, meaning they have two copies of each chromosome. However, not all plants are diploid. The common strawberry is octoploid ($8n$) and the cultivated kiwi is hexaploid ($6n$). Research the total number of chromosomes in the cells of each of these fruits and think about how this might correspond to the amount of DNA in these fruits’ cell nuclei. What other factors might contribute to the total amount of DNA in a single fruit? Read about the technique of DNA isolation to understand how each step in the isolation protocol helps liberate and precipitate DNA.

Hypothesis: Hypothesize whether you would be able to detect a difference in DNA quantity from similarly sized strawberries and kiwis. Which fruit do you think would yield more DNA?

Test your hypothesis: Isolate the DNA from a strawberry and a kiwi that are similarly sized. Perform the experiment in at least triplicate for each fruit

1. Prepare a bottle of DNA extraction buffer from 900 mL water, 50 mL dish detergent, and two teaspoons of table salt. Mix by inversion (cap it and turn it upside down a few times).
2. Grind a strawberry and a kiwi by hand in a plastic bag, or using a mortar and pestle, or with a metal bowl and the end of a blunt instrument. Grind for at least two minutes per fruit.
3. Add 10 mL of the DNA extraction buffer to each fruit, and mix well for at least one minute.
4. Remove cellular debris by filtering each fruit mixture through cheesecloth or porous cloth and into a funnel placed in a test tube or an appropriate container.
5. Pour ice-cold ethanol or isopropanol (rubbing alcohol) into the test tube. You should observe white, precipitated DNA.
6. Gather the DNA from each fruit by winding it around separate glass rods.

Analyze your data: Did you notice an obvious difference in the amount of DNA produced by each fruit? Were your results reproducible?

Draw a conclusion: Given what you know about the number of chromosomes in each fruit, can you conclude that chromosome number necessarily correlates to DNA amount? Can you identify any drawbacks to this procedure? If you had access to a laboratory, how could you standardize your comparison and make it more quantitative?

This page titled 8.2: The Genetic Code is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.3: PROKARYOTIC TRANSCRIPTION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- List the different steps in prokaryotic transcription
- Discuss the role of promoters in prokaryotic transcription
- Describe how and when transcription is terminated

The prokaryotes, which include Bacteria and Archaea, are mostly single-celled organisms that, by definition, lack membrane-bound nuclei and other organelles. A bacterial chromosome is a closed circle that, unlike eukaryotic chromosomes, is not organized around histone proteins. The central region of the cell in which prokaryotic DNA resides is called the nucleoid region. In addition, prokaryotes often have abundant plasmids, which are shorter, circular DNA molecules that may only contain one or a few genes. Plasmids can be transferred independently of the bacterial chromosome during cell division and often carry traits such as those involved with antibiotic resistance.

Transcription in prokaryotes (and in eukaryotes) requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a transcription bubble. Transcription always proceeds from the same DNA strand for each gene, which is called the template strand. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the nontemplate strand, or the coding strand. The only nucleotide difference is that in mRNA, all of the T nucleotides are replaced with U nucleotides (Figure 15.7). In an RNA double helix, A can bind U via two hydrogen bonds, just as in A–T pairing in a DNA double helix.

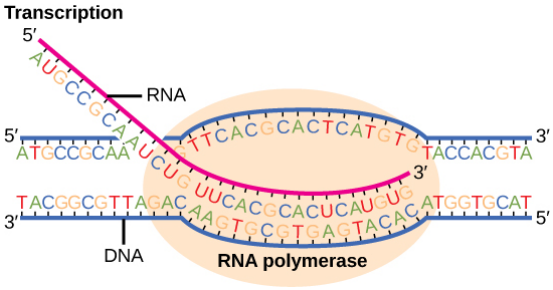


Figure 15.7 Messenger RNA is a copy of protein-coding information in the coding strand of DNA, with the substitution of U in the RNA for T in the coding sequence. However, new RNA nucleotides base pair with the nucleotides of the template strand. RNA is synthesized in its 5'-3' direction, using the enzyme RNA polymerase. As the template is read, the DNA unwinds ahead of the polymerase and then rewinds behind it.

The nucleotide pair in the DNA double helix that corresponds to the site from which the first 5' mRNA nucleotide is transcribed is called the +1 site, or the initiation site. Nucleotides preceding the initiation site are denoted with a "-" and are designated *upstream nucleotides*. Conversely, nucleotides following the initiation site are denoted with "+" numbering and are called *downstream nucleotides*.

INITIATION OF TRANSCRIPTION IN PROKARYOTES

Prokaryotes do not have membrane-enclosed nuclei. Therefore, the processes of transcription, translation, and mRNA degradation can all occur simultaneously. The intracellular level of a bacterial protein can quickly be amplified by multiple transcription and translation events that occur concurrently on the same DNA template. Prokaryotic genomes are very compact, and prokaryotic transcripts often cover more than one gene or cistron (a coding sequence for a single protein). Polycistronic mRNAs are then translated to produce more than one kind of protein.

Our discussion here will exemplify transcription by describing this process in *Escherichia coli*, a well-studied eubacterial species. Although some differences exist between transcription in *E. coli* and transcription in archaea, an understanding of *E. coli* transcription can be applied to virtually all bacterial species.

PROKARYOTIC RNA POLYMERASE

Prokaryotes use the same RNA polymerase to transcribe all of their genes. In *E. coli*, the polymerase is composed of five polypeptide subunits, two of which are identical. Four of these subunits, denoted α, α, β, and β', comprise the polymerase core enzyme. These subunits assemble every time a gene is transcribed, and they disassemble once transcription is complete. Each subunit has a unique role; the two α-subunits are necessary to assemble the polymerase on the DNA; the β-subunit binds to the ribonucleoside triphosphate that will become part of the nascent mRNA molecule; and the β' subunit binds the DNA template strand. The fifth subunit, σ, is involved only in transcription

initiation. It confers transcriptional specificity such that the polymerase begins to synthesize mRNA from an appropriate initiation site. Without σ, the core enzyme would transcribe from random sites and would produce mRNA molecules that specified protein gibberish. The polymerase comprised of all five subunits is called the holoenzyme.

PROKARYOTIC PROMOTERS

A promoter is a DNA sequence onto which the transcription machinery, including RNA polymerase, binds and initiates transcription. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very important because it determines whether the corresponding gene is transcribed all the time, some of the time, or infrequently. Although promoters vary among prokaryotic genomes, a few elements are evolutionarily conserved in many species. At the -10 and -35 regions upstream of the initiation site, there are two *promoter consensus sequences*, or regions that are similar across all promoters and across various bacterial species (Figure 15.8). The -10 sequence, called the -10 region, has the consensus sequence TATAAT. The -35 sequence has the consensus sequence TTGACA. These consensus sequences are recognized and bound by σ. Once this interaction is made, the subunits of the core enzyme bind to the site. The A–T-rich -10 region facilitates unwinding of the DNA template, and several phosphodiester bonds are made. The transcription initiation phase ends with the production of abortive transcripts, which are polymers of approximately 10 nucleotides that are made and released.

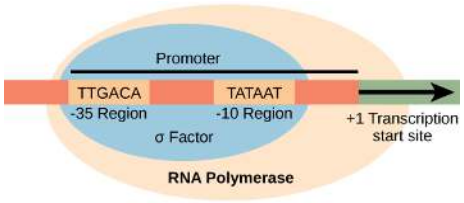


Figure 15.8 The σ subunit of prokaryotic RNA polymerase recognizes consensus sequences found in the promoter region upstream of the transcription start site. The σ subunit dissociates from the polymerase after transcription has been initiated.

LINK TO LEARNING

LINK TO LEARNING

View this MolecularMovies animation to see the transcription process as it happens in the cell.

ELONGATION AND TERMINATION IN PROKARYOTES

The transcription elongation phase begins with the release of the σ subunit from the polymerase. The dissociation of σ allows the core enzyme to proceed along the DNA template, synthesizing mRNA in the 5' to 3' direction at a rate of approximately 40 nucleotides per second. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it. The base pairing between DNA and RNA is not stable enough to maintain the stability of the mRNA synthesis components. Instead, the RNA polymerase acts as a stable linker between the DNA template and the nascent RNA strands to ensure that elongation is not interrupted prematurely.

PROKARYOTIC TERMINATION SIGNALS

Once a gene is transcribed, the prokaryotic polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals. One is protein-based and the other is RNA-based. Rho-dependent termination is controlled by the rho protein, which tracks along behind the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

Rho-independent termination is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in C–G nucleotides. The mRNA folds back on itself, and the complementary C–G nucleotides bind together. The result is a stable hairpin that causes the polymerase to stall as soon as it begins to transcribe a region rich in A–T nucleotides. The complementary U–A region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

Upon termination, the process of transcription is complete. By the time termination occurs, the prokaryotic transcript would already have been used to begin synthesis of numerous copies of the encoded protein because these processes can occur concurrently. The unification of transcription, translation, and even mRNA degradation is possible because all of these processes occur in the same 5' to 3' direction, and because there is no membranous compartmentalization in the prokaryotic cell (Figure 15.9). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.

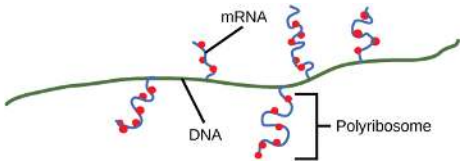


Figure 15.9 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

LINK TO LEARNING

LINK TO LEARNING

Visit this BioStudio animation to see the process of prokaryotic transcription.

This page titled 8.3: Prokaryotic Transcription is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.4: EUKARYOTIC TRANSCRIPTION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- List the steps in eukaryotic transcription
- Discuss the role of RNA polymerases in transcription
- Compare and contrast the three RNA polymerases
- Explain the significance of transcription factors

Prokaryotes and eukaryotes perform fundamentally the same process of transcription, with a few key differences. The most important difference between prokaryote and eukaryote transcription is due to the latter's membrane-bound nucleus and organelles. With the genes bound in a nucleus, the eukaryotic cell must be able to transport its mRNA to the cytoplasm and must protect its mRNA from degrading before it is translated. Eukaryotes also employ three different polymerases that each transcribe a different subset of genes. Eukaryotic mRNAs are usually *monogenic*, meaning that they specify a single protein.

INITIATION OF TRANSCRIPTION IN EUKARYOTES

Unlike the prokaryotic polymerase that can bind to a DNA template on its own, eukaryotes require several other proteins, called transcription factors, to first bind to the promoter region and then to help recruit the appropriate polymerase.

THE THREE EUKARYOTIC RNA POLYMERASES

The features of eukaryotic mRNA synthesis are markedly more complex than those of prokaryotes. Instead of a single polymerase comprising five subunits, the eukaryotes have three polymerases that are each made up of 10 subunits or more. Each eukaryotic polymerase also requires a distinct set of transcription factors to bring it to the DNA template.

RNA polymerase I is located in the nucleolus, a specialized nuclear substructure in which ribosomal RNA (rRNA) is transcribed, processed, and assembled into ribosomes (Table 15.1). The rRNA molecules are considered structural RNAs because they have a cellular role but are not translated into protein. The rRNAs are components of the ribosome and are essential to the process of translation. RNA polymerase I synthesizes all of the rRNAs from the tandemly duplicated set of 18S, 5.8S, and 28S ribosomal genes. (Note that the "S" designation applies to "Svedberg" units, a nonadditive value that characterizes the speed at which a particle sediments during centrifugation.)

Locations, Products, and Sensitivities of the Three Eukaryotic RNA Polymerases

RNA Polymerase	Cellular Compartment	Product of Transcription	α-Amanitin Sensitivity
I	Nucleolus	All rRNAs except 5S rRNA	Insensitive
II	Nucleus	All protein-coding nuclear pre-mRNAs	Extremely sensitive
III	Nucleus	5S rRNA, tRNAs, and small nuclear RNAs	Moderately sensitive

Table 15.1

RNA polymerase II is located in the nucleus and synthesizes all protein-coding nuclear pre-mRNAs. Eukaryotic pre-mRNAs undergo extensive processing after transcription but before translation. For clarity, this module's discussion of transcription and translation in eukaryotes will use the term "mRNAs" to describe only the mature, processed molecules that are ready to be translated. RNA polymerase II is responsible for transcribing the overwhelming majority of eukaryotic genes.

RNA polymerase III is also located in the nucleus. This polymerase transcribes a variety of structural RNAs that includes the 5S pre-rRNA, transfer pre-RNAs (pre-tRNAs), and small nuclear pre-RNAs. The tRNAs have a critical role in translation; they serve as the "adaptor molecules" between the mRNA template and the growing polypeptide chain. Small nuclear RNAs have a variety of functions, including "splicing" pre-mRNAs and regulating transcription factors.

A scientist characterizing a new gene can determine which polymerase transcribes it by testing whether the gene is expressed in the presence of α-amanitin, an oligopeptide toxin produced by the fly agaric toadstool mushroom and other species of *Amanita*. Interestingly, the α-amanitin affects the three polymerases very differently (Table 15.1). RNA polymerase I is completely insensitive to α-amanitin, meaning that the polymerase can transcribe DNA in vitro in the presence of this poison. RNA polymerase III is moderately sensitive to the toxin. In contrast, RNA polymerase II is extremely sensitive to α-amanitin. The toxin prevents the enzyme from progressing down the DNA, and thus inhibits transcription. Knowing the transcribing polymerase can provide clues as to the general function of the gene being studied. Because RNA polymerase II transcribes the vast majority of genes, we will focus on this polymerase in our subsequent discussions about eukaryotic transcription factors and promoters.

RNA POLYMERASE II PROMOTERS AND TRANSCRIPTION FACTORS

Eukaryotic promoters are much larger and more intricate than prokaryotic promoters. However, both have a sequence similar to the -10 sequence of prokaryotes. In eukaryotes, this sequence is called the TATA box, and has the consensus sequence TATAAA on the coding strand. It is located at -25 to -35 bases relative to the initiation (+1) site (Figure 15.10). This sequence is not identical to the *E. coli* -10 box, but it conserves the A–T rich element. The thermostability of A–T bonds is low and this helps the DNA template to locally unwind in preparation for transcription.

Instead of the simple σ factor that helps bind the prokaryotic RNA polymerase to its promoter, eukaryotes assemble a complex of transcription factors required to recruit RNA polymerase II to a protein coding gene. Transcription factors that bind to the promoter are called *basal transcription factors*. These basal factors are all called TFII (for Transcription Factor/polymerase II) plus an additional letter (A-J). The core complex is TFIID, which includes a TATA-binding protein (TBP). The other transcription factors systematically fall into place on the DNA template, with each one further stabilizing the pre-initiation complex and contributing to the recruitment of RNA polymerase II.

VISUAL CONNECTION

VISUAL CONNECTION

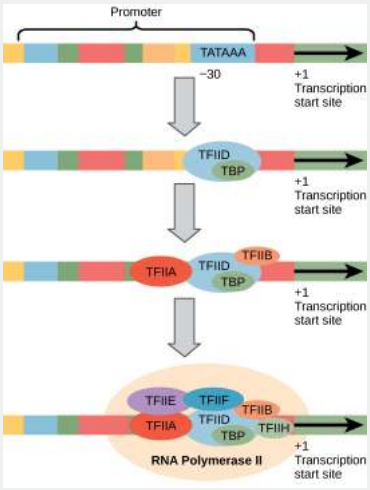


Figure 15.10 A generalized promoter of a gene transcribed by RNA polymerase II is shown. Transcription factors recognize the promoter. RNA polymerase II then binds and forms the transcription initiation complex.

A scientist splices a eukaryotic promoter in front of a bacterial gene and inserts the gene in a bacterial chromosome. Would you expect the bacteria to transcribe the gene?

Some eukaryotic promoters also have a conserved CAAT box (GGCCAATCT) at approximately -80. Further upstream of the TATA box, eukaryotic promoters may also contain one or more GC-rich boxes (GGCG) or octamer boxes (ATTTGCAT). These elements bind cellular factors that increase the efficiency of transcription initiation and are often identified in more “active” genes that are constantly being expressed by the cell.

Basal transcription factors are crucial in the formation of a preinitiation complex on the DNA template that subsequently recruits RNA polymerase II for transcription initiation. The complexity of eukaryotic transcription does not end with the polymerases and promoters. An army of other transcription factors, which bind to upstream enhancers and silencers, also help to regulate the frequency with which pre-mRNA is synthesized from a gene. Enhancers and silencers affect the efficiency of transcription but are not necessary for transcription to proceed.

PROMOTER STRUCTURES FOR RNA POLYMERASES I AND III

The processes of bringing RNA polymerases I and III to the DNA template involve slightly less complex collections of transcription factors, but the general theme is the same.

The conserved promoter elements for genes transcribed by polymerases I and III differ from those transcribed by RNA polymerase II. RNA polymerase I transcribes genes that have two GC-rich promoter sequences in the -45 to +20 region. These sequences alone are sufficient for transcription initiation to occur, but promoters with additional sequences in the region from -180 to -105 upstream of the initiation site will further enhance initiation. Genes that are transcribed by RNA polymerase III have upstream promoters or promoters that occur within the genes themselves.

Eukaryotic transcription is a tightly regulated process that requires a variety of proteins to interact with each other and with the DNA strand. Although the process of transcription in eukaryotes involves a greater metabolic investment than in prokaryotes, it ensures that the cell transcribes precisely the pre-mRNAs that it needs for protein synthesis.

EVOLUTION CONNECTION

EVOLUTION CONNECTION

The Evolution of Promoters The evolution of genes may be a familiar concept. Mutations can occur in genes during DNA replication, and the result may or may not be beneficial to the cell. By altering an enzyme, structural protein, or some other factor, the process of mutation can transform functions or physical features. However, eukaryotic promoters and other gene regulatory sequences may evolve as well. For instance, consider a gene that, over many generations, becomes more valuable to the cell. Maybe the gene encodes a structural protein that the cell needs to synthesize in abundance for a certain function. If this is the case, it would be beneficial to the cell for that gene's promoter to recruit transcription factors more efficiently and increase gene expression.

Scientists examining the evolution of promoter sequences have reported varying results. In part, this is because it is difficult to infer exactly where a eukaryotic promoter begins and ends. Some promoters occur within genes; others are located very far upstream, or even downstream, of the genes they are regulating. However, when researchers limited their examination to human core promoter sequences that were defined experimentally as sequences that bind the preinitiation complex, they found that promoters evolve even faster than protein-coding genes.

It is still unclear how promoter evolution might correspond to the evolution of humans or other complex organisms. However, the evolution of a promoter to effectively make more or less of a given gene product is an intriguing alternative to the evolution of the genes themselves.[1]

EUKARYOTIC ELONGATION AND TERMINATION

Following the formation of the preinitiation complex, the polymerase is released from the other transcription factors, and elongation is allowed to proceed as it does in prokaryotes with the polymerase synthesizing pre-mRNA in the 5' to 3' direction. As discussed previously, RNA polymerase II transcribes the major share of eukaryotic genes, so in this section we will focus on how this polymerase accomplishes elongation and termination.

Although the enzymatic process of elongation is essentially the same in eukaryotes and prokaryotes, the DNA template is considerably more complex. When eukaryotic cells are not dividing, their genes exist as a diffuse mass of DNA and proteins called chromatin. The DNA is tightly packaged around charged histone proteins at repeated intervals. These *DNA–histone complexes*, collectively called nucleosomes, are regularly spaced and include 146 nucleotides of DNA wound around eight histones like thread around a spool.

For polynucleotide synthesis to occur, the transcription machinery needs to move histones out of the way every time it encounters a nucleosome. This is accomplished by a special protein complex called FACT, which stands for "*facilitates chromatin transcription*." This complex pulls histones away from the DNA template as the polymerase moves along it. Once the pre-mRNA is synthesized, the FACT complex replaces the histones to recreate the nucleosomes.

The termination of transcription is different for the different polymerases. Unlike in prokaryotes, elongation by RNA polymerase II in eukaryotes takes place 1,000 to 2,000 nucleotides beyond the end of the gene being transcribed. This pre-mRNA tail is subsequently removed by cleavage during mRNA processing. On the other hand, RNA polymerases I and III require termination signals. Genes transcribed by RNA polymerase I contain a specific 18-nucleotide sequence that is recognized by a termination protein. The process of termination in RNA polymerase III involves an mRNA hairpin similar to rho-independent termination of transcription in prokaryotes.

FOOTNOTES

- 1H Liang et al., "Fast evolution of core promoters in primate genomes," *Molecular Biology and Evolution* 25 (2008): 1239–44.

This page titled 8.4: Eukaryotic Transcription is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.5: RNA PROCESSING IN EUKARYOTES

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Describe the different steps in RNA processing
- Understand the significance of exons, introns, and splicing for mRNAs
- Explain how tRNAs and rRNAs are processed

After transcription, eukaryotic pre-mRNAs must undergo several processing steps before they can be translated. Eukaryotic (and prokaryotic) tRNAs and rRNAs also undergo processing before they can function as components in the protein-synthesis machinery.

MRNA PROCESSING

The eukaryotic pre-mRNA undergoes extensive processing before it is ready to be translated. Eukaryotic protein-coding sequences are not continuous, as they are in prokaryotes. The coding sequences (exons) are interrupted by noncoding introns, which must be removed to make a translatable mRNA. The additional steps involved in eukaryotic mRNA maturation also create a molecule with a much longer half-life than a prokaryotic mRNA. Eukaryotic mRNAs last for several hours, whereas the typical *E. coli* mRNA lasts no more than five seconds.

Pre-mRNAs are first coated in RNA-stabilizing proteins; these protect the pre-mRNA from degradation while it is processed and exported out of the nucleus. The three most important steps of pre-mRNA processing are the addition of stabilizing and signaling factors at the 5' and 3' ends of the molecule, and the removal of the introns (Figure 15.11). In rare cases, the mRNA transcript can be “edited” after it is transcribed.

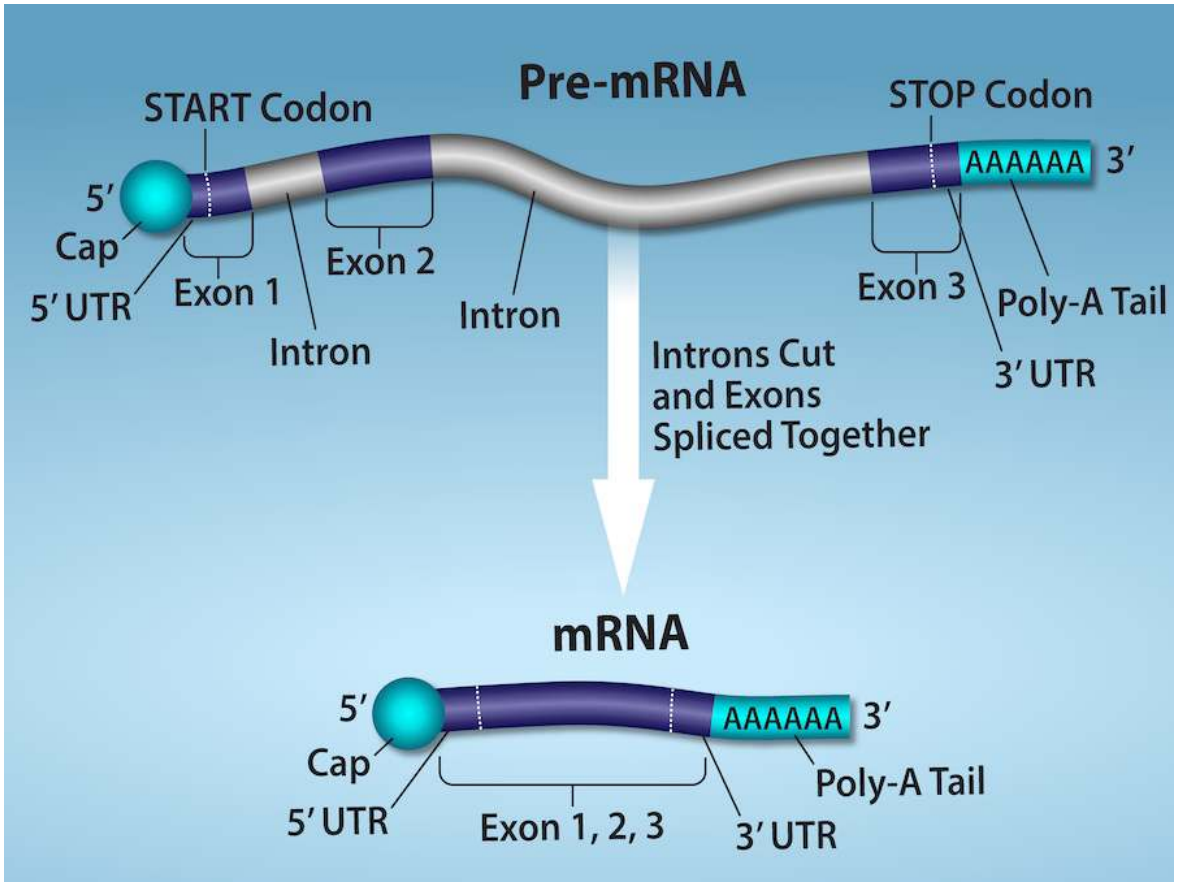


Figure 15.11 Eukaryotic pre-mRNA processing. In addition to 5’ Cap and 3’ Poly-A Tail addition, introns must be precisely removed and exons joined to generate a functional mRNA. Nucleotides upstream (towards the 5’cap) of the translation START codon are part of the 5’ untranslated region (5’ UTR). Nucleotides downstream (towards 3’end) of the STOP codon form the 3’ UTR. Both 5’ and 3’ UTRs are important for regulating translation initiation and mRNA stability. Credit: Rao, A., Ryan, K. Fletcher, S. and Tag, A. Department of Biology, Texas A&M University.

EVOLUTION CONNECTION

EVOLUTION CONNECTION

RNA Editing in TrypanosomesThe trypanosomes are a group of protozoa that include the pathogen *Trypanosoma brucei*, which causes nagana in cattle and sleeping sickness in humans throughout great areas of Africa (Figure 15.12). The trypanosome is carried by biting flies in the genus *Glossina* (commonly called tsetse flies). Trypanosomes, and virtually all other eukaryotes, have organelles called mitochondria that supply the cell with chemical energy. Mitochondria are organelles that express their own DNA and are believed to be the remnants of a symbiotic relationship between a eukaryote and an engulfed prokaryote. The mitochondrial DNA of trypanosomes exhibit an interesting exception to the central dogma: their pre-mRNAs do not have the correct information to specify a functional protein. Usually, this is because the mRNA is missing several U nucleotides. The cell performs an additional RNA processing step called RNA editing to remedy this.



Figure 15.12 *Trypanosoma brucei* is the causative agent of sleeping sickness in humans. The mRNAs of this pathogen must be modified by the addition of nucleotides before protein synthesis can occur. (credit: modification of work by Torsten Ochsenreiter)

Other genes in the mitochondrial genome encode 40- to 80-nucleotide guide RNAs. One or more of these molecules interacts by complementary base pairing with some of the nucleotides in the pre-mRNA transcript. However, the *guide RNA* has more A nucleotides than the pre-mRNA has U nucleotides with which to bind. In these regions, the guide RNA loops out. The 3' ends of guide RNAs have a long poly-U tail, and these U bases are inserted in regions of the pre-mRNA transcript at which the guide RNAs are looped. This process is entirely mediated by RNA molecules. That is, guide RNAs—rather than proteins—serve as the catalysts in RNA editing.

RNA editing is not just a phenomenon of trypanosomes. In the mitochondria of some plants, almost all pre-mRNAs are edited. RNA editing has also been identified in mammals such as rats, rabbits, and even humans. What could be the evolutionary reason for this additional step in pre-mRNA processing? One possibility is that the mitochondria, being remnants of ancient prokaryotes, have an equally ancient RNA-based method for regulating gene expression. In support of this hypothesis, edits made to pre-mRNAs differ depending on cellular conditions. Although speculative, the process of RNA editing may be a holdover from a primordial time when RNA molecules, instead of proteins, were responsible for catalyzing reactions.

5' CAPPING

While the pre-mRNA is still being synthesized, a 7-methylguanosine cap is added to the 5' end of the growing transcript by a phosphate linkage. This functional group protects the nascent mRNA from degradation. In addition, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

3' POLY-A TAIL

Once elongation is complete, the pre-mRNA is cleaved by an endonuclease between an AAUAAA consensus sequence and a GU-rich sequence, leaving the AAUAAA sequence on the pre-mRNA. An enzyme called poly-A polymerase then adds a string of approximately 200 A residues, called the poly-A tail. This modification further protects the pre-mRNA from degradation and is also the binding site for a protein necessary for exporting the processed mRNA to the cytoplasm.

PRE-MRNA SPLICING

Eukaryotic genes are composed of exons, which correspond to protein-coding sequences (*ex*-on signifies that they are *ex*pressed), and *int*ervening sequences called introns (*int*-ron denotes their *int*ervening role), which may be involved in gene regulation but are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins.

The discovery of introns came as a surprise to researchers in the 1970s who expected that pre-mRNAs would specify protein sequences without further processing, as they had observed in prokaryotes. The genes of higher eukaryotes very often contain one or more introns. These regions may correspond to regulatory sequences; however, the biological significance of having many introns or having very long

introns in a gene is unclear. It is possible that introns slow down gene expression because it takes longer to transcribe pre-mRNAs with lots of introns. Alternatively, introns may be nonfunctional sequence remnants left over from the fusion of ancient genes throughout the course of evolution. This is supported by the fact that separate exons often encode separate protein subunits or domains. For the most part, the sequences of introns can be mutated without ultimately affecting the protein product.

All of a pre-mRNA's introns must be completely and precisely removed before protein synthesis. If the process errs by even a single nucleotide, the reading frame of the rejoined exons would shift, and the resulting protein would be dysfunctional. The process of removing introns and reconnecting exons is called splicing (Figure 15.13). Introns are removed and degraded while the pre-mRNA is still in the nucleus. Splicing occurs by a sequence-specific mechanism that ensures introns will be removed and exons rejoined with the accuracy and precision of a single nucleotide. Although the intron itself is noncoding, the beginning and end of each intron is marked with specific nucleotides: GU at the 5' end and AG at the 3' end of the intron. The splicing of pre-mRNAs is conducted by complexes of proteins and RNA molecules called spliceosomes.

VISUAL CONNECTION

VISUAL CONNECTION

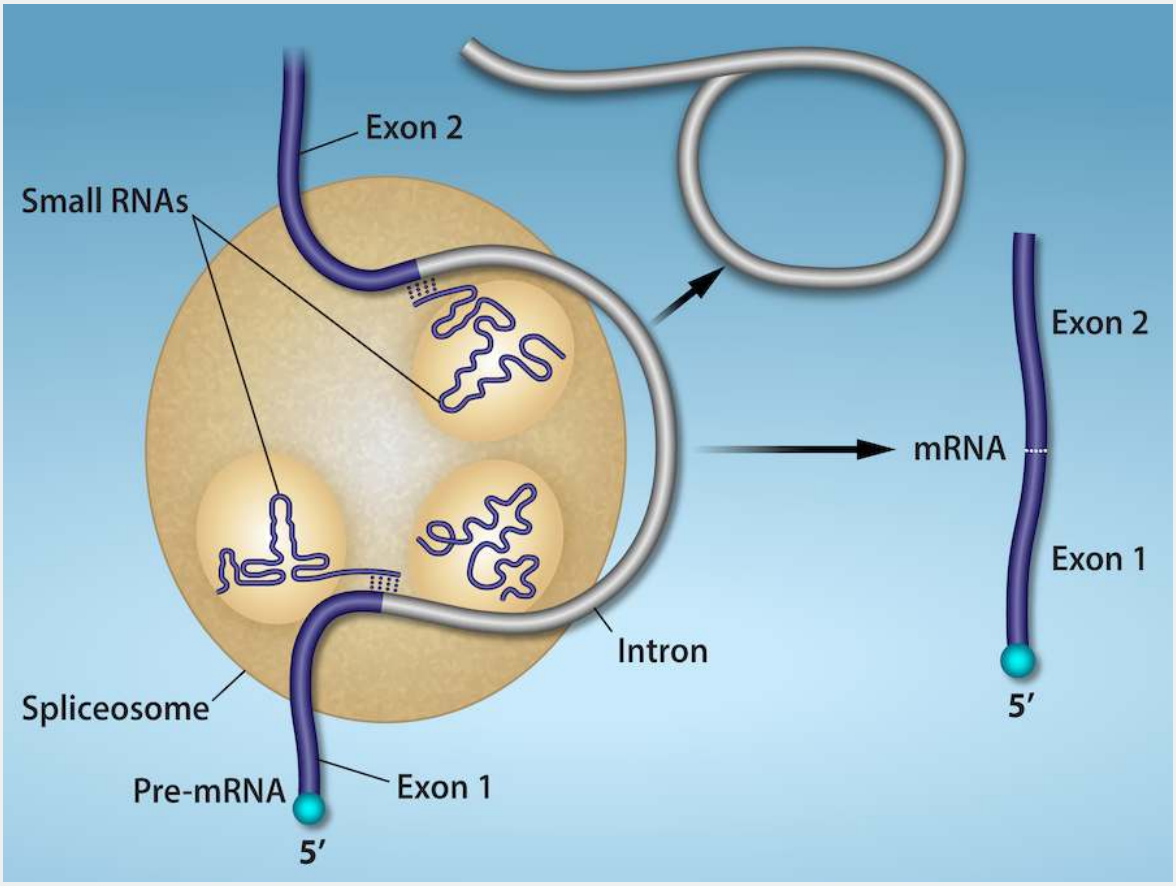


Figure 15.13 Pre-mRNA splicing involves the precise removal of introns from the primary RNA transcript. The splicing process is catalyzed by protein complexes called spliceosomes that are composed of proteins and RNA molecules called small nuclear RNAs (snRNAs). Spliceosomes recognize sequences at the 5' and 3' end of the intron. Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

Errors in splicing are implicated in cancers and other human diseases. What kinds of mutations might lead to splicing errors? Think of different possible outcomes if splicing errors occur.

Note that more than 70 individual introns can be present, and each has to undergo the process of splicing—in addition to 5' capping and the addition of a poly-A tail—just to generate a single, translatable mRNA molecule.

LINK TO LEARNING

> **LINK TO LEARNING**
>
> See how introns are removed during RNA splicing at this website.

PROCESSING OF TRNAS AND RRNAS

The tRNAs and rRNAs are structural molecules that have roles in protein synthesis; however, these RNAs are not themselves translated. Pre-rRNAs are transcribed, processed, and assembled into ribosomes in the nucleolus. Pre-tRNAs are transcribed and processed in the nucleus and then released into the cytoplasm where they are linked to free amino acids for protein synthesis.

Most of the tRNAs and rRNAs in eukaryotes and prokaryotes are first transcribed as a long precursor molecule that spans multiple rRNAs or tRNAs. Enzymes then cleave the precursors into subunits corresponding to each structural RNA. Some of the bases of pre-rRNAs are *methylated*; that is, a $–CH_3$ methyl functional group is added for stability. Pre-tRNA molecules also undergo methylation. As with pre-mRNAs, subunit excision occurs in eukaryotic pre-RNAs destined to become tRNAs or rRNAs.

Mature rRNAs make up approximately 50 percent of each ribosome. Some of a ribosome's RNA molecules are purely structural, whereas others have catalytic or binding activities. Mature tRNAs take on a three-dimensional structure through local regions of base pairing stabilized by intramolecular hydrogen bonding. The tRNA folds to position the amino acid binding site at one end and the anticodon at the other end (Figure 15.14). The anticodon is a three-nucleotide sequence in a tRNA that interacts with an mRNA codon through complementary base pairing.

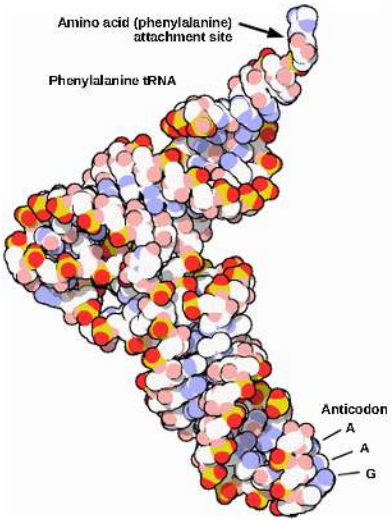


Figure 15.14 This is a space-filling model of a tRNA molecule that adds the amino acid phenylalanine to a growing polypeptide chain. The anticodon AAG binds the Codon UUC on the mRNA. The amino acid phenylalanine is attached to the other end of the tRNA.

This page titled 8.5: RNA Processing in Eukaryotes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.6: RIBOSOMES AND PROTEIN SYNTHESIS

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Describe the different steps in protein synthesis
- Discuss the role of ribosomes in protein synthesis

The synthesis of proteins consumes more of a cell's energy than any other metabolic process. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform virtually every function of a cell. The process of translation, or protein synthesis, involves the decoding of an mRNA message into a polypeptide product. Amino acids are covalently strung together by interlinking peptide bonds in lengths ranging from approximately 50 to more than 1000 amino acid residues. Each individual amino acid has an amino group (NH_2) and a carboxyl (COOH) group. Polypeptides are formed when the amino group of one amino acid forms an amide (i.e., peptide) bond with the carboxyl group of another amino acid (Figure 15.15). This reaction is catalyzed by ribosomes and generates one water molecule.

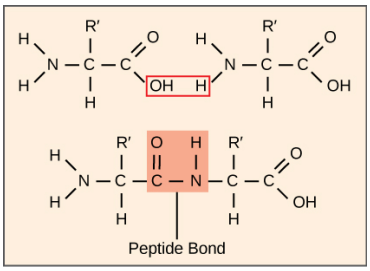


Figure 15.15 A peptide bond links the carboxyl end of one amino acid with the amino end of another, producing one water molecule during the process. For simplicity in this image, only the functional groups involved in the peptide bond are shown. The R and R' designations refer to the rest of each amino acid structure.

THE PROTEIN SYNTHESIS MACHINERY

In addition to the mRNA template, many molecules and macromolecules contribute to the process of translation. The composition of each component may vary across species; for example, ribosomes may consist of different numbers of rRNAs and polypeptides depending on the organism. However, the general structures and functions of the protein synthesis machinery are comparable from bacteria to human cells. Translation requires the input of an mRNA template, ribosomes, tRNAs, and various enzymatic factors. (Note: A ribosome can be thought of as an enzyme whose amino acid binding sites are specified by mRNA.)

LINK TO LEARNING

LINK TO LEARNING

Click through the steps of this PBS interactive to see protein synthesis in action.

RIBOSOMES

Even before an mRNA is translated, a cell must invest energy to build each of its ribosomes. In *E. coli*, there are between 10,000 and 70,000 ribosomes present in each cell at any given time. A ribosome is a complex macromolecule composed of structural and catalytic rRNAs, and many distinct polypeptides. In eukaryotes, the nucleolus is completely specialized for the synthesis and assembly of rRNAs.

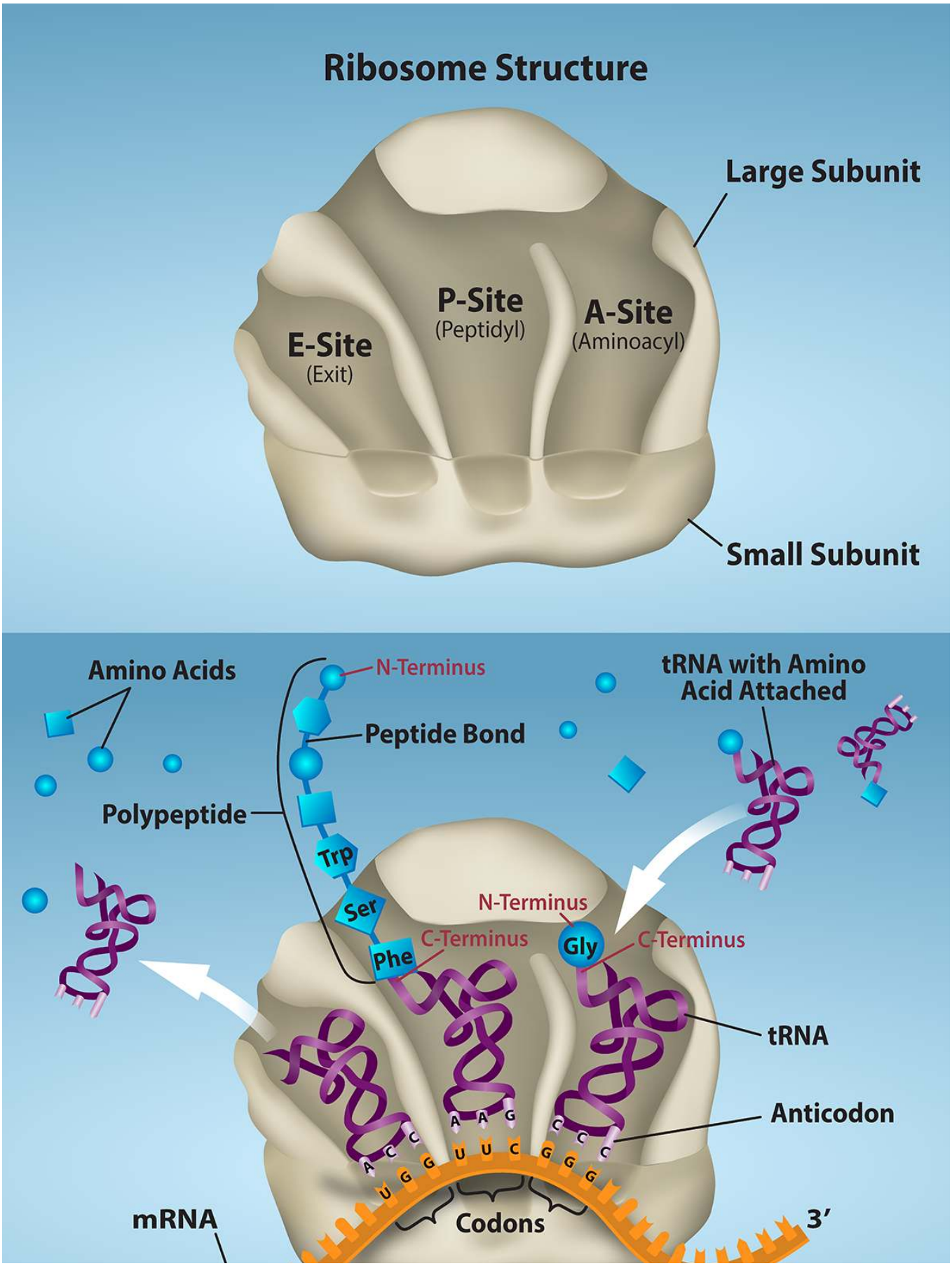
Ribosomes exist in the cytoplasm of prokaryotes and in the cytoplasm and rough endoplasmic reticulum of eukaryotes. Mitochondria and chloroplasts also have their own ribosomes in the matrix and stroma, which look more similar to prokaryotic ribosomes (and have similar drug sensitivities) than the ribosomes just outside their outer membranes in the cytoplasm. Ribosomes dissociate into large and small subunits when they are not synthesizing proteins and reassociate during the initiation of translation. *In E. coli, the small subunit is described as 30S, and the large subunit is 50S, for a total of 70S (recall that Svedberg units are not additive).* Mammalian ribosomes have a small 40S subunit and a large 60S subunit, for a total of 80S. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds tRNAs. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction: reading the mRNA from 5' to 3' and synthesizing the polypeptide from the N terminus to the C terminus. The complete mRNA/poly-ribosome structure is called a polysome.

TRNAS

The tRNAs are structural RNA molecules that were transcribed from genes by RNA polymerase III. Depending on the species, 40 to 60 types of tRNAs exist in the cytoplasm. Transfer RNAs serve as adaptor molecules. Each tRNA carries a specific amino acid and recognizes one or more of the mRNA codons that define the order of amino acids in a protein. Aminoacyl-tRNAs bind to the ribosome and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually “translate” the language of RNA into the language of proteins.

Of the 64 possible mRNA codons—or triplet combinations of A, U, G, and C—three specify the termination of protein synthesis and 61 specify the addition of amino acids to the polypeptide chain. Of these 61, one codon (AUG) also encodes the initiation of translation. Each tRNA anticodon can base pair with one or more of the mRNA codons for its amino acid. For instance, if the sequence CUA occurred on an mRNA template in the proper reading frame, it would bind a leucine tRNA expressing the complementary sequence, GAU. The ability of some tRNAs to match more than one codon is what gives the genetic code its blocky structure.

As the adaptor molecules of translation, it is surprising that tRNAs can fit so much specificity into such a small package. Consider that tRNAs need to interact with three factors: 1) they must be recognized by the correct aminoacyl synthetase (see below); 2) they must be recognized by ribosomes; and 3) they must bind to the correct sequence in mRNA.



https://bio.libretexts.org/@go/page/143293

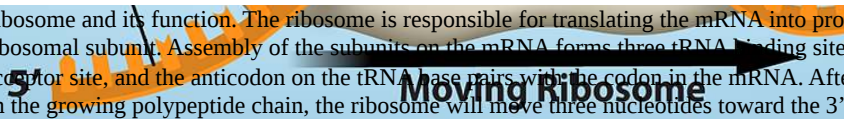


Figure 15.16 The ribosome and its function. The ribosome is responsible for translating the mRNA into protein. A. The ribosome consists of a large and small ribosomal subunit. Assembly of the subunits on the mRNA forms three tRNA binding sites. B. During translation, charged tRNAs enter the Acceptor site, and the anticodon on the tRNA base pairs with the codon in the mRNA. After the incoming amino acid forms a peptide bond with the growing polypeptide chain, the ribosome will move three nucleotides toward the 3' end of the mRNA. This movement will transfer the tRNA with the growing polypeptide to the Peptidyl-tRNA binding site and allow the empty tRNA to exit at the Exit site. Credit: Rao, A., Ryan, K. and Fletcher, S. Department of Biology, Texas A&M University.

AMINOACYL TRNA SYNTHETASES

The process of pre-tRNA synthesis by RNA polymerase III only creates the RNA portion of the adaptor molecule. The corresponding amino acid must be added later, once the tRNA is processed and exported to the cytoplasm. Through the process of tRNA "charging," each tRNA molecule is linked to its correct amino acid by one of a group of enzymes called aminoacyl tRNA synthetases. At least one type of aminoacyl tRNA synthetase exists for each of the 20 amino acids; the exact number of aminoacyl tRNA synthetases varies by species. These enzymes first bind and hydrolyze ATP to catalyze a high-energy bond between an amino acid and adenosine monophosphate (AMP); a pyrophosphate molecule is expelled in this reaction. The activated amino acid is then transferred to the tRNA, and AMP is released. The term "charging" is appropriate, since the high-energy bond that attaches an amino acid to its tRNA is later used to drive the formation of the peptide bond. Each tRNA is named for its amino acid.

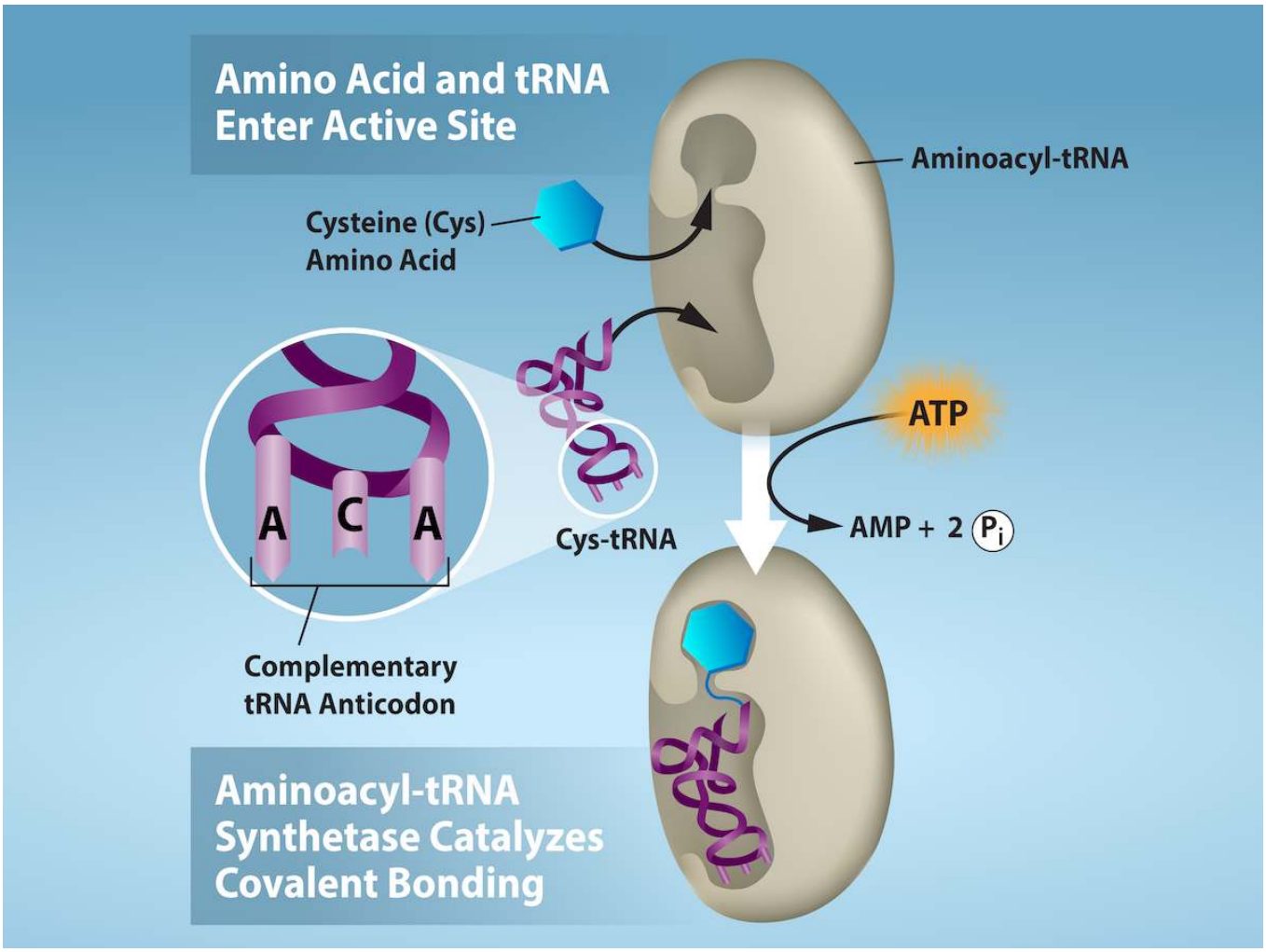


Figure 15.17 Charging of tRNAs with correct amino acids. Aminoacyl-tRNA synthetases catalyze covalent bond formation between the tRNA and the correct amino acid in preparation for translation. Because there are multiple amino acids, there are multiple different tRNA synthetases. All of the synthetases require energy, in the form of ATP, to make sure the correct amino acid is attached to the tRNA with the correct anticodon sequence. Credit: Rao, A., Ryan, K. and Tag, A. Department of Biology, Texas A&M University.

THE MECHANISM OF PROTEIN SYNTHESIS

As with mRNA synthesis, protein synthesis can be divided into three phases: *initiation*, *elongation*, *and termination*. The process of translation is similar in prokaryotes and eukaryotes. Here we'll explore how translation occurs in *E. coli*, a representative prokaryote, and specify any differences between prokaryotic and eukaryotic translation.

INITIATION OF TRANSLATION

Protein synthesis begins with the formation of an initiation complex. In *E. coli*, this complex involves the small 30S ribosome, the mRNA template, three initiation factors (IFs; IF-1, IF-2, and IF-3), and a special initiator tRNA, called $tRNA^{Metf}$.

In *E. coli* mRNA, a sequence upstream of the first AUG codon, called the Shine-Dalgarno sequence (AGGAGG), interacts with the rRNA molecules that compose the ribosome. This interaction anchors the 30S ribosomal subunit at the correct location on the mRNA template. Guanosine triphosphate (GTP), which is a purine nucleotide triphosphate, acts as an energy source during translation—both at the start of elongation and during the ribosome's translocation. Binding of the mRNA to the 30S ribosome also requires IF-III.

The initiator tRNA then interacts with the start codon AUG (or rarely, GUG). This tRNA carries the amino acid methionine, which is formylated after its attachment to the tRNA. The formylation creates a "faux" peptide bond between the formyl carboxyl group and the amino group of the methionine. Binding of the $fMet\text{-}tRNA^{Metf}$ is mediated by the initiation factor IF-2. The fMet begins every polypeptide chain synthesized by *E. coli*, but it is usually removed after translation is complete. When an in-frame AUG is encountered during translation elongation, a non-formylated methionine is inserted by a regular $Met\text{-}tRNA^{Met}$. After the formation of the initiation complex, the 30S ribosomal subunit is joined by the 50S subunit to form the translation complex. In eukaryotes, a similar initiation complex forms, comprising mRNA, the 40S small ribosomal subunit, eukaryotic IFs, and nucleoside triphosphates (GTP and ATP). The methionine on the charged initiator tRNA, called $Met\text{-}tRNA_i$, is not formylated. However, $Met\text{-}tRNA_i$ is distinct from other Met-tRNAs in that it can bind IFs.

Instead of depositing at the Shine-Dalgarno sequence, the eukaryotic initiation complex recognizes the 7-methylguanosine cap at the 5' end of the mRNA. A cap-binding protein (CBP) and several other IFs assist the movement of the ribosome to the 5' cap. Once at the cap, the initiation complex tracks along the mRNA in the 5' to 3' direction, searching for the AUG start codon. Many eukaryotic mRNAs are translated from the first AUG, but this is not always the case. According to Kozak's rules, the nucleotides around the AUG indicate whether it is the correct start codon. Kozak's rules state that the following consensus sequence must appear around the AUG of vertebrate genes: 5'-gccRccAUGG-3'. The R (for purine) indicates a site that can be either A or G, but cannot be C or U. Essentially, the closer the sequence is to this consensus, the higher the efficiency of translation.

Once the appropriate AUG is identified, the other proteins and CBP dissociate, and the 60S subunit binds to the complex of $Met\text{-}tRNA_i$, mRNA, and the 40S subunit. This step completes the initiation of translation in eukaryotes.

TRANSLATION, ELONGATION, AND TERMINATION

In prokaryotes and eukaryotes, the basics of elongation are the same, so we will review elongation from the perspective of *E. coli*. When the translation complex is formed, the tRNA binding region of the ribosome consists of three compartments. The A (aminoacyl) site binds incoming charged aminoacyl tRNAs. The P (peptidyl) site binds charged tRNAs carrying amino acids that have formed peptide bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The E (exit) site releases dissociated tRNAs so that they can be recharged with free amino acids. The initiating methionyl-tRNA, however, occupies the P site at the beginning of the elongation phase of translation in both prokaryotes and eukaryotes.

During translation elongation, the mRNA template provides tRNA binding specificity. As the ribosome moves along the mRNA, each mRNA codon comes into register, and specific binding with the corresponding charged tRNA anticodon is ensured. If mRNA were not present in the elongation complex, the ribosome would bind tRNAs nonspecifically and randomly.

Elongation proceeds with charged tRNAs sequentially entering and leaving the ribosome as each new amino acid is added to the polypeptide chain. Movement of a tRNA from A to P to E site is induced by conformational changes that advance the ribosome by three bases in the 3' direction. The energy for each step along the ribosome is donated by elongation factors that hydrolyze GTP. GTP energy is required both for the binding of a new aminoacyl-tRNA to the A site and for its translocation to the P site after formation of the peptide bond. Peptide bonds form between the amino group of the amino acid attached to the A-site tRNA and the carboxyl group of the amino acid attached to the P-site tRNA. The formation of each peptide bond is catalyzed by peptidyl transferase, an RNA-based enzyme that is integrated into the 50S ribosomal subunit. The energy for each peptide bond formation is derived from the high-energy bond linking each amino acid to its tRNA. After peptide bond formation, the A-site tRNA that now holds the growing peptide chain moves to the P site, and the P-site tRNA that is now empty moves to the E site and is expelled from the ribosome (Figure 15.18). Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino-acid protein can be translated in just 10 seconds.

VISUAL CONNECTION

VISUAL CONNECTION

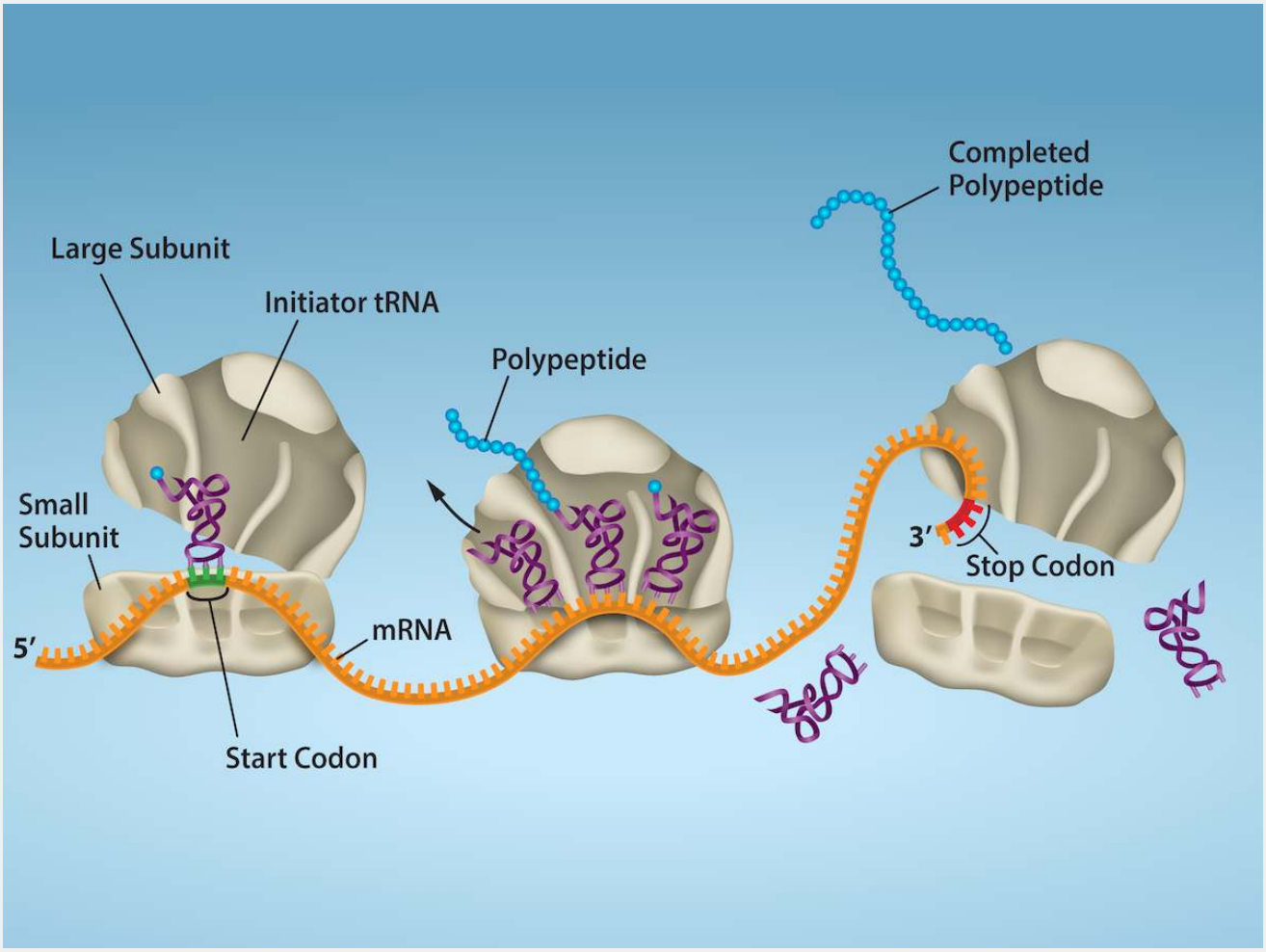


Figure 15.18 Translation begins when an initiator tRNA anticodon recognizes a start codon on mRNA bound to a small ribosomal subunit. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, successive tRNAs move through the ribosome and the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Many antibiotics inhibit bacterial protein synthesis. For example, tetracycline blocks the A site on the bacterial ribosome, and chloramphenicol blocks peptidyl transfer. What specific effect would you expect each of these antibiotics to have on protein synthesis?

Tetracycline would directly affect:

a. tRNA binding to the ribosome
b. ribosome assembly
c. growth of the protein chain

Chloramphenicol would directly affect:

a. tRNA binding to the ribosome
b. ribosome assembly
c. growth of the protein chain

Termination of translation occurs when a nonsense codon (UAA, UAG, or UGA) is encountered. Upon aligning with the A site, these nonsense codons are recognized by *protein release factors* that resemble tRNAs. The releasing factors in both prokaryotes and eukaryotes instruct peptidyl transferase to add a water molecule to the carboxyl end of the P-site amino acid. This reaction forces the P-site amino acid to detach from its tRNA, and the newly made protein is released. The small and large ribosomal subunits dissociate from the mRNA and from each other; they are recruited almost immediately into another translation initiation complex. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

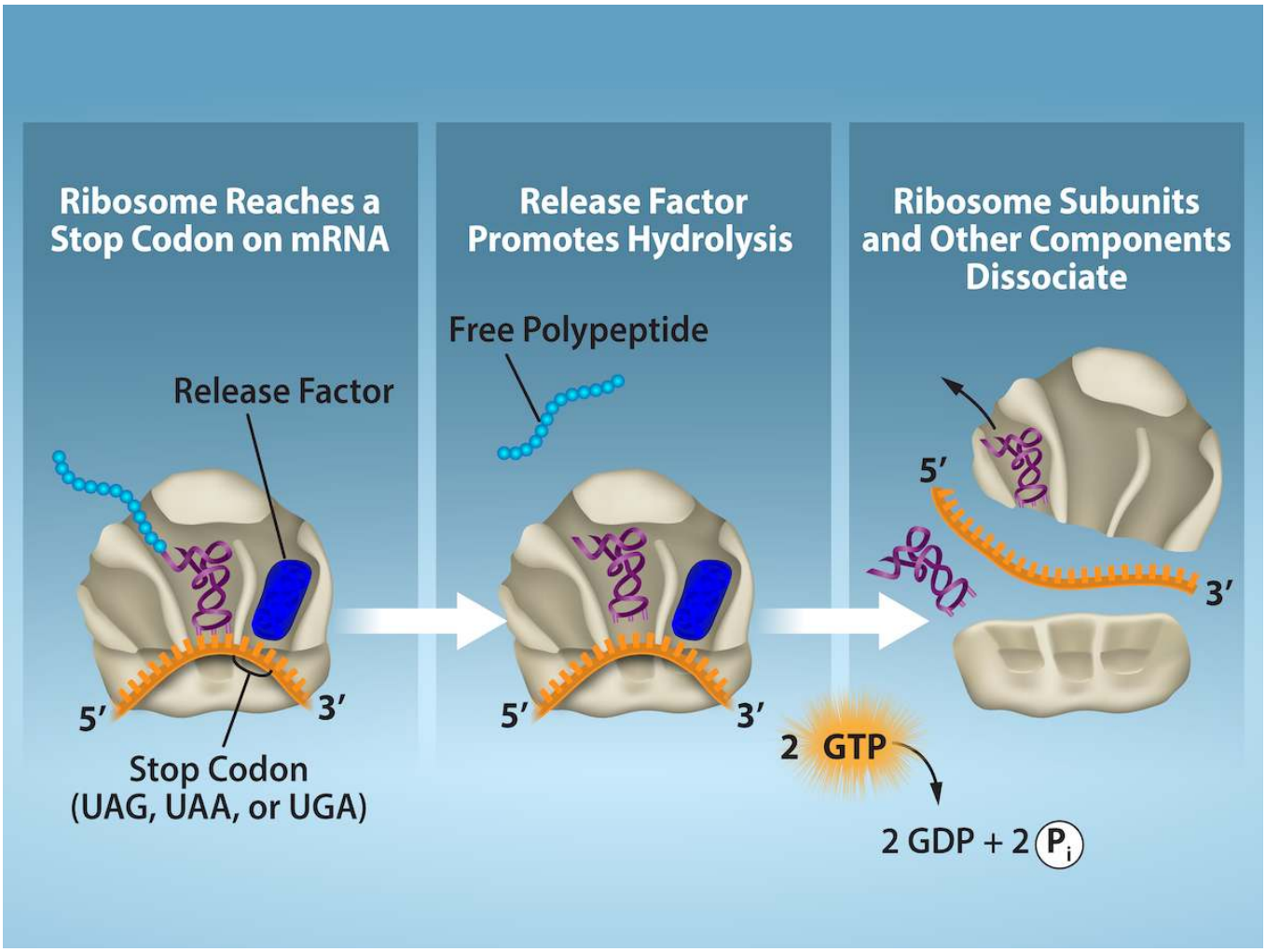


Figure 15.19 Translation Termination is Active. Translation is terminated when a STOP codon is in the A-site of the ribosome. Since there are no tRNAs corresponding to the STOP codons, the Release Factor protein enters and catalyzes the hydrolysis between the last amino acid and its tRNA. This hydrolysis releases the free carboxyl terminus (C-term) of the protein. Additional factors use the energy in GTP hydrolysis to disassemble the large and small ribosomal subunits and mRNA. Credit: Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

PROTEIN FOLDING, MODIFICATION, AND TARGETING

During and after translation, individual amino acids may be chemically modified, signal sequences appended, and the new protein "folded" into a distinct three-dimensional structure as a result of intramolecular interactions. A signal sequence is a short sequence at the amino end of a protein that directs it to a specific cellular compartment. These sequences can be thought of as the protein's "train ticket" to its ultimate destination, and are recognized by signal-recognition proteins that act as conductors. For instance, a specific signal sequence terminus will direct a protein to the mitochondria or chloroplasts (in plants). Once the protein reaches its cellular destination, the signal sequence is usually clipped off.

Many proteins fold spontaneously, but some proteins require helper molecules, called *chaperones*, to prevent them from aggregating during the complicated process of folding. Even if a protein is properly specified by its corresponding mRNA, it could take on a completely dysfunctional shape if abnormal temperature or pH conditions prevent it from folding correctly.

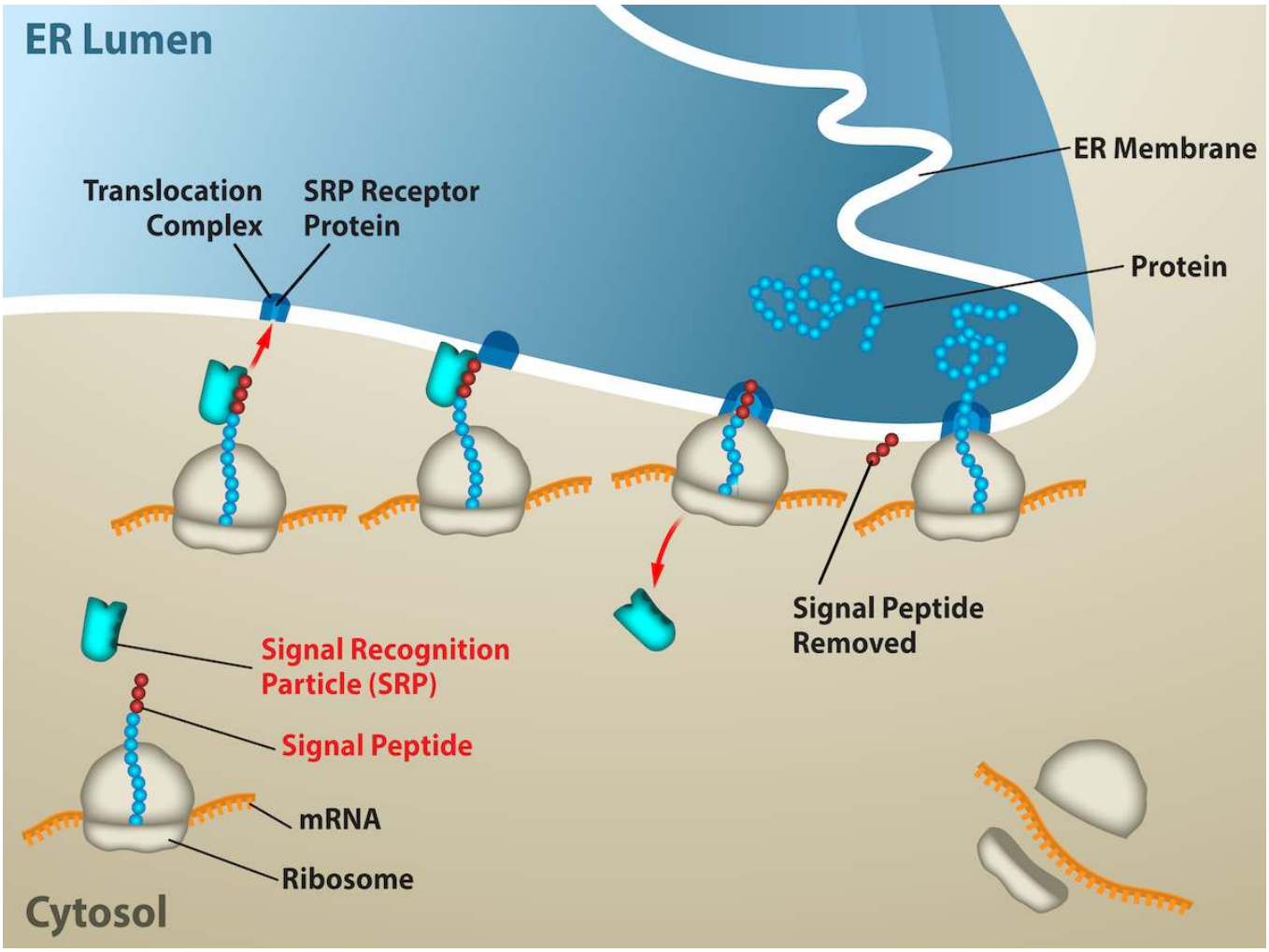


Figure 15.20 Proteins are co-translationally targeted into the ER for secretion. Proteins that will be secreted from the cell will contain a signal sequence at the N-terminus. The signal will be recognized by SRP as soon as the amino acids emerge from the ribosome, and the ribosome will be targeted to the translocation channel in the ER membrane. The rest of the protein will go directly from the ribosome, across the ER membrane and into the ER lumen. From the ER, proteins can be secreted from the cell via vesicle trafficking. Credit: Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

This page titled 8.6: Ribosomes and Protein Synthesis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

SECTION OVERVIEW

8.7: GENE EXPRESSION

Whereas each cell shares the same genome and DNA sequence, each cell does not turn on, or express, the same set of genes. Each cell type needs a different set of proteins to perform its function. Therefore, only a small subset of proteins is expressed in a cell. For the proteins to be expressed, the DNA must be transcribed into RNA and the RNA must be translated into protein. In a given cell type, not all genes encoded in the DNA are transcribed into RNA or translated into protein because specific cells in our body have specific functions. Specialized proteins that make up the eye (iris, lens, and cornea) are only expressed in the eye, whereas the specialized proteins in the heart (pacemaker cells, heart muscle, and valves) are only expressed in the heart. At any given time, only a subset of all of the genes encoded by our DNA are expressed and translated into proteins. The expression of specific genes is a highly regulated process with many levels and stages of control. This complexity ensures the proper expression in the proper cell at the proper time.

8.7.1: INTRODUCTION

8.7.2: REGULATION OF GENE EXPRESSION

8.7.3: PROKARYOTIC GENE REGULATION

8.7.4: EUKARYOTIC EPIGENETIC GENE REGULATION

8.7.5: EUKARYOTIC TRANSCRIPTION GENE REGULATION

8.7.6: EUKARYOTIC POST-TRANSCRIPTIONAL GENE REGULATION

8.7.7: EUKARYOTIC TRANSLATIONAL AND POST-TRANSLATIONAL GENE REGULATION

8.7.8: CANCER AND GENE REGULATION

8.7.9: KEY TERMS

8.7.10: CHAPTER SUMMARY

8.7.11: VISUAL CONNECTION QUESTIONS

8.7.12: REVIEW QUESTIONS

8.7.13: CRITICAL THINKING QUESTIONS

This page titled 8.7: Gene Expression is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.1: INTRODUCTION

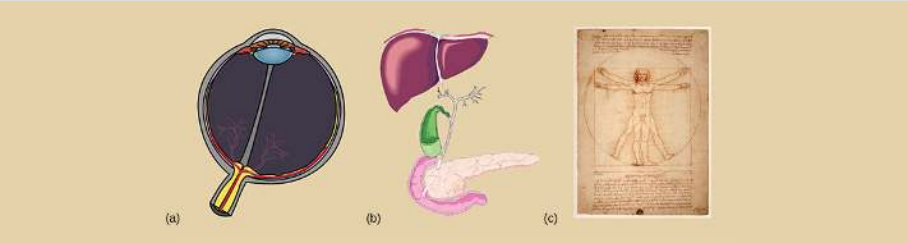


Figure 16.1 The genetic content of each somatic cell in an organism is the same, but not all genes are expressed in every cell. The control of which genes are expressed dictates whether a cell is, for example, (a) an eye cell or (b) a liver cell. It is the differential gene expression patterns that arise in different cells that give rise to (c) a complete organism.

CHAPTER OUTLINE

16.1 Regulation of Gene Expression
16.2 Prokaryotic Gene Regulation
16.3 Eukaryotic Epigenetic Gene Regulation
16.4 Eukaryotic Transcription Gene Regulation
16.5 Eukaryotic Post-transcriptional Gene Regulation
16.6 Eukaryotic Translational and Post-translational Gene Regulation
16.7 Cancer and Gene Regulation

Each somatic cell in the body generally contains the same DNA. A few exceptions include red blood cells, which contain no DNA in their mature state, and some immune system cells that rearrange their DNA while producing antibodies. In general, however, the genes that determine whether you have green eyes, brown hair, and how fast you metabolize food are the same in the cells in your eyes and your liver, even though these organs function quite differently. If each cell has the same DNA, how is it that cells or organs are different? Why do cells in the eye differ so dramatically from cells in the liver?

Whereas each cell shares the same genome and DNA sequence, each cell does not turn on, or express, the same set of genes. Each cell type needs a different set of proteins to perform its function. Therefore, only a small subset of proteins is expressed in a cell. For the proteins to be expressed, the DNA must be transcribed into RNA and the RNA must be translated into protein. In a given cell type, not all genes encoded in the DNA are transcribed into RNA or translated into protein because specific cells in our body have specific functions. Specialized proteins that make up the eye (iris, lens, and cornea) are only expressed in the eye, whereas the specialized proteins in the heart (pacemaker cells, heart muscle, and valves) are only expressed in the heart. At any given time, only a subset of all of the genes encoded by our DNA are expressed and translated into proteins. The expression of specific genes is a highly regulated process with many levels and stages of control. This complexity ensures the proper expression in the proper cell at the proper time.

This page titled 8.7.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.2: REGULATION OF GENE EXPRESSION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Discuss why every cell does not express all of its genes all of the time
- Describe how prokaryotic gene regulation occurs at the transcriptional level
- Discuss how eukaryotic gene regulation occurs at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels

For a cell to function properly, necessary proteins must be synthesized at the proper time and place. All cells control or regulate the synthesis of proteins from information encoded in their DNA. The process of turning on a gene to produce RNA and protein is called gene expression. Whether in a simple unicellular organism or a complex multi-cellular organism, each cell controls when and how its genes are expressed. For this to occur, there must be internal chemical mechanisms that control when a gene is expressed to make RNA and protein, how much of the protein is made, and when it is time to stop making that protein because it is no longer needed.

The regulation of gene expression conserves energy and space. It would require a significant amount of energy for an organism to express every gene at all times, so it is more energy efficient to turn on the genes only when they are required. In addition, only expressing a subset of genes in each cell saves space because DNA must be unwound from its tightly coiled structure to transcribe and translate the DNA. Cells would have to be enormous if every protein were expressed in every cell all the time.

The control of gene expression is extremely complex. Malfunctions in this process are detrimental to the cell and can lead to the development of many diseases, including cancer.

PROKARYOTIC VERSUS EUKARYOTIC GENE EXPRESSION

To understand how gene expression is regulated, we must first understand how a gene codes for a functional protein in a cell. The process occurs in both prokaryotic and eukaryotic cells, just in slightly different manners.

Prokaryotic organisms are single-celled organisms that lack a cell nucleus, and their DNA therefore floats freely in the cell cytoplasm. To synthesize a protein, the processes of transcription and translation occur almost simultaneously. When the resulting protein is no longer needed, transcription stops. As a result, the primary method to control what type of protein and how much of each protein is expressed in a prokaryotic cell is the regulation of DNA transcription. All of the subsequent steps occur automatically. When more protein is required, more transcription occurs. Therefore, in prokaryotic cells, the control of gene expression is mostly at the transcriptional level.

Eukaryotic cells, in contrast, have intracellular organelles that add to their complexity. In eukaryotic cells, the DNA is contained inside the cell's nucleus and there it is transcribed into RNA. The newly synthesized RNA is then transported out of the nucleus into the cytoplasm, where ribosomes translate the RNA into protein. The processes of transcription and translation are *physically separated* by the nuclear membrane; transcription occurs only within the nucleus, and translation occurs only outside the nucleus in the cytoplasm. The regulation of gene expression can occur at all stages of the process (Figure 16.3). Regulation may occur when the DNA is uncoiled and loosened from nucleosomes to bind transcription factors (epigenetic level), when the RNA is transcribed (transcriptional level), when the RNA is processed and exported to the cytoplasm after it is transcribed (post-transcriptional level), when the RNA is translated into protein (translational level), or after the protein has been made (post-translational level).

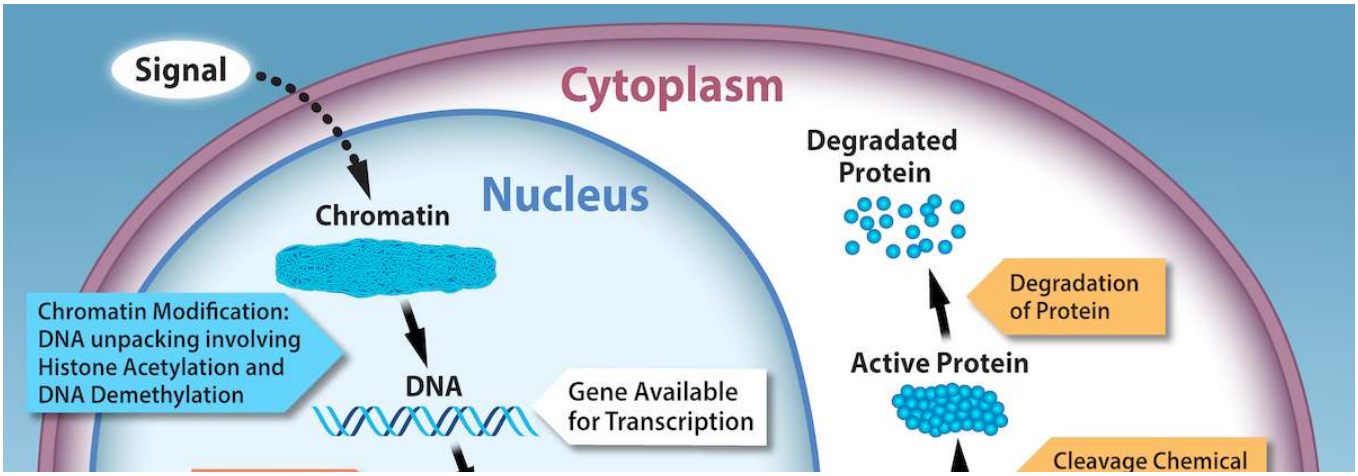


Figure 16.2 Locations of gene regulation. The regulation of gene expression occurs at multiple steps going from DNA to the functional gene product, usually a protein. It begins with chromatin structure making the DNA more or less accessible for transcription by RNA polymerase. In eukaryotes, the primary mRNA transcript must be processed before it can be translated in the cytoplasm. The final level of active protein in the cell depends not only on the rate of synthesis, but also on the rate of degradation of mRNA and protein. Credit: Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

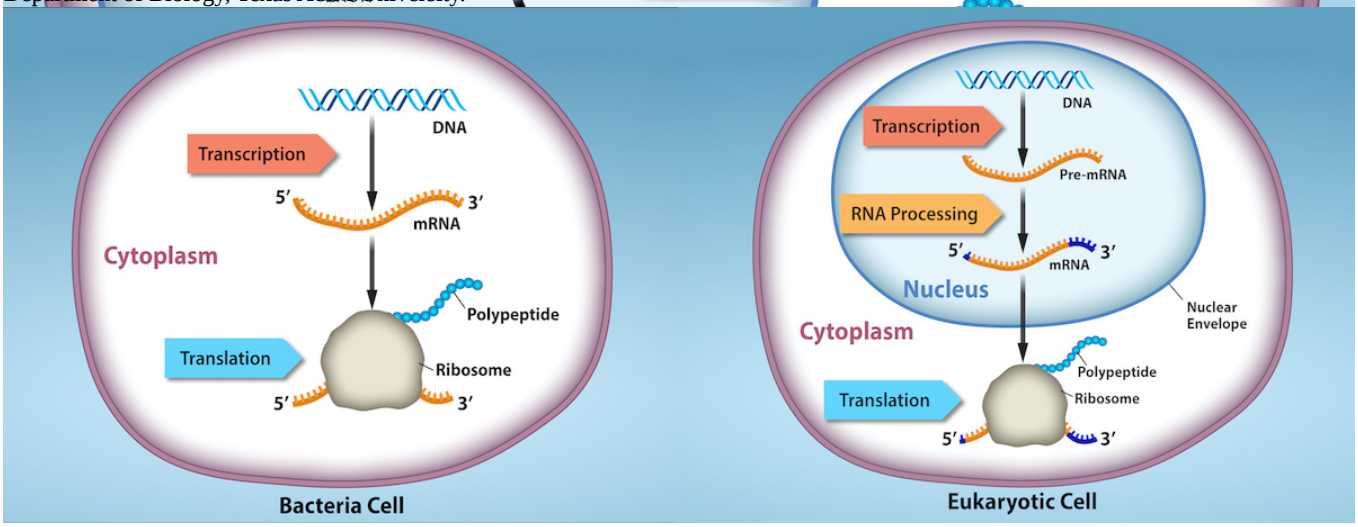


Figure 16.3 Regulation in prokaryotes and eukaryotes. A. Prokaryotic transcription and translation occur simultaneously in the cytoplasm, and regulation occurs primarily at the transcriptional level. B. Eukaryotic gene expression is regulated during transcription and RNA processing, which take place in the nucleus, and during protein translation, which takes place in the cytoplasm. Further regulation may occur through post-translational modifications of proteins in both prokaryotes and eukaryotes. Credit: Rao, A., Ryan, K. Fletcher, S. and Tag, A. Department of Biology, Texas A&M University.

The differences in the regulation of gene expression between prokaryotes and eukaryotes are summarized in Table 16.1. The regulation of gene expression is discussed in detail in subsequent modules.

Differences in the Regulation of Gene Expression of Prokaryotic and Eukaryotic Organisms

Prokaryotic organisms	Eukaryotic organisms
Lack a membrane-bound nucleus	Contain nucleus
DNA is found in the cytoplasm	DNA is confined to the nuclear compartment
RNA transcription and protein formation occur almost simultaneously	RNA transcription occurs prior to protein formation, and it takes place in the nucleus. Translation of RNA to protein occurs in the cytoplasm.
Gene expression is regulated primarily at the transcriptional level	Gene expression is regulated at many levels (epigenetic, transcriptional, nuclear shuttling, post-transcriptional, translational, and post-translational)

Table 16.1

EVOLUTION CONNECTION

EVOLUTION CONNECTION

Evolution of Gene RegulationProkaryotic cells can only regulate gene expression by controlling the amount of transcription. As eukaryotic cells evolved, the complexity of the control of gene expression increased. For example, with the evolution of eukaryotic cells came compartmentalization of important cellular components and cellular processes. A nuclear region that contains the DNA was formed. Transcription and translation were physically separated into two different cellular compartments. It therefore became possible to control gene expression by regulating transcription in the nucleus, and also by controlling the RNA levels and protein translation present outside the nucleus.

Most gene regulation is done to conserve cell resources. However, other regulatory processes may be defensive. Cellular processes such as gene silencing developed to protect the cell from viral or parasitic infections. If the cell could quickly shut off gene expression for a short period of time, it would be able to survive an infection when other organisms could not. Therefore, the organism evolved a new process that helped it survive, and it was able to pass this new development to offspring.

This page titled 8.7.2: Regulation of Gene Expression is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.3: PROKARYOTIC GENE REGULATION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Describe the steps involved in prokaryotic gene regulation
- Explain the roles of activators, inducers, and repressors in gene regulation

The DNA of prokaryotes is organized into a circular chromosome, supercoiled within the nucleoid region of the cell cytoplasm. Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are encoded together in blocks called operons. For example, all of the genes needed to use lactose as an energy source are coded next to each other in the lactose (or *lac*) operon, and transcribed into a single mRNA.

In prokaryotic cells, there are three types of regulatory molecules that can affect the expression of operons: repressors, activators, and inducers. Repressors and activators are proteins produced in the cell. Both repressors and activators regulate gene expression by binding to specific DNA sites *adjacent* to the genes they control. *In general, activators bind to the promoter site, while repressors bind to operator regions*. Repressors prevent transcription of a gene in response to an external stimulus, whereas activators increase the transcription of a gene in response to an external stimulus. Inducers are small molecules that may be produced by the cell or that are in the cell's environment. Inducers either activate or repress transcription depending on the needs of the cell and the availability of substrate.

THE *TRP* OPERON: A REPRESSIBLE OPERON

Bacteria such as *Escherichia coli* need amino acids to survive, and are able to synthesize many of them. Tryptophan is one such amino acid that *E. coli* can either ingest from the environment or synthesize using enzymes that are encoded by five genes. These five genes are next to each other in what is called the tryptophan (*trp*) operon (Figure 16.4). The genes are transcribed into a single mRNA, which is then translated to produce all five enzymes. If tryptophan is present in the environment, then *E. coli* does not need to synthesize it and the *trp* operon is switched off. However, when tryptophan availability is low, the switch controlling the operon is turned on, the mRNA is transcribed, the enzyme proteins are translated, and tryptophan is synthesized.

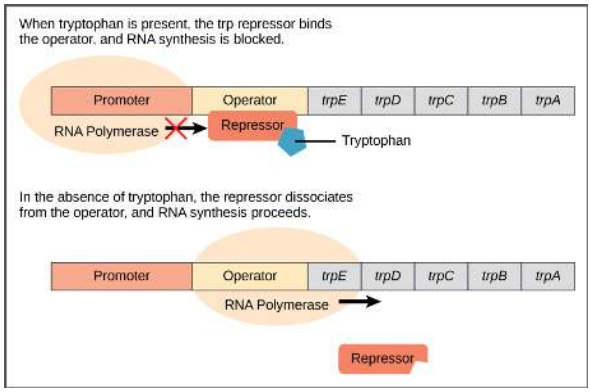


Figure 16.4 The tryptophan operon. The five genes that are needed to synthesize tryptophan in *E. coli* are located next to each other in the *trp* operon. When tryptophan is plentiful, two tryptophan molecules bind the repressor protein at the operator sequence. This physically blocks the RNA polymerase from transcribing the tryptophan genes. When tryptophan is absent, the repressor protein does not bind to the operator and the genes are transcribed.

The *trp* operon includes three important regions: the coding region, the *trp* operator and the *trp* promoter. The coding region includes the genes for the five tryptophan biosynthesis enzymes. Just before the coding region is the transcriptional start site. The promoter sequence, to which RNA polymerase binds to initiate transcription, is before or "upstream" of the transcriptional start site. Between the promoter and the transcriptional start site is the operator region.

The *trp* operator contains the DNA code to which the *trp* repressor protein can bind. However, the repressor alone cannot bind to the operator. When tryptophan is present in the cell, two tryptophan molecules bind to the *trp* repressor, which changes the shape of the repressor protein to a form that can bind to the *trp* operator. Binding of the tryptophan–repressor complex at the operator physically prevents the RNA polymerase from binding to the promoter and transcribing the downstream genes.

When tryptophan is not present in the cell, the repressor by itself does not bind to the operator, the polymerase can transcribe the enzyme genes, and tryptophan is synthesized. Because the repressor protein actively binds to the operator to keep the genes turned off, the *trp*

operon is said to be *negatively regulated* and the proteins that bind to the operator to silence *trp* expression are negative regulators.

LINK TO LEARNING

LINK TO LEARNING

Watch this video to learn more about the *trp* operon.

Click to view content

CATABOLITE ACTIVATOR PROTEIN (CAP): A TRANSCRIPTIONAL ACTIVATOR

Just as the *trp* operon is negatively regulated by tryptophan molecules, there are proteins that bind to the promoter sequences that act as positive regulators to turn genes on and activate them. For example, when glucose is scarce, *E. coli* bacteria can turn to other sugar sources for fuel. To do this, new genes to process these alternate sugars must be transcribed. When glucose levels drop, cyclic AMP (cAMP) begins to accumulate in the cell. The cAMP molecule is a signaling molecule that is involved in glucose and energy metabolism in *E. coli*. Accumulating cAMP binds to the positive regulator catabolite activator protein (CAP), a protein that binds to the promoters of operons which control the processing of alternative sugars. When cAMP binds to CAP, the complex then binds to the promoter region of the genes that are needed to use the alternate sugar sources (Figure 16.5). In these operons, a CAP-binding site is located upstream of the RNA-polymerase-binding site in the promoter. CAP binding stabilizes the binding of RNA polymerase to the promoter region and increases transcription of the associated protein-coding genes.

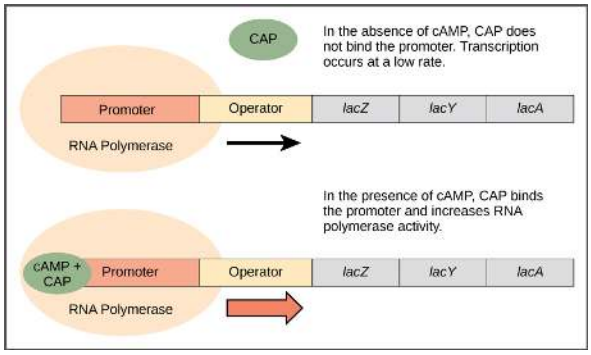


Figure 16.5 Transcriptional activation by the CAP protein. When glucose levels fall, *E. coli* may use other sugars for fuel but must transcribe new genes to do so. As glucose supplies become limited, cAMP levels increase. This cAMP binds to the CAP protein, a positive regulator that binds to a promoter region upstream of the genes required to use other sugar sources.

THE *LAC* OPERON: AN INDUCIBLE OPERON

The third type of gene regulation in prokaryotic cells occurs through *inducible operons*, which have proteins that bind to activate or repress transcription depending on the local environment and the needs of the cell. The *lac* operon is a typical inducible operon. As mentioned previously, *E. coli* is able to use other sugars as energy sources when glucose concentrations are low. One such sugar source is lactose. The *lac* operon encodes the genes necessary to acquire and process the lactose from the local environment. The Z gene of the *lac* operon encodes beta-galactosidase, which breaks lactose down to glucose and galactose.

However, for the *lac* operon to be activated, two conditions must be met. First, the level of glucose must be very low or non-existent. Second, lactose must be present. Only when glucose is absent and lactose is present will the *lac* operon be transcribed (Figure 16.6). In the absence of glucose, the binding of the CAP protein makes transcription of the *lac* operon more effective. When lactose is present, its metabolite, allolactose, binds to the *lac* repressor and changes its shape so that it cannot bind to the *lac* operator to prevent transcription. This combination of conditions makes sense for the cell, because it would be energetically wasteful to synthesize the enzymes to process lactose if glucose was plentiful or lactose was not available. It should be mentioned that the lac operon is transcribed at a very low rate even when glucose is present and lactose absent.

VISUAL CONNECTION

VISUAL CONNECTION

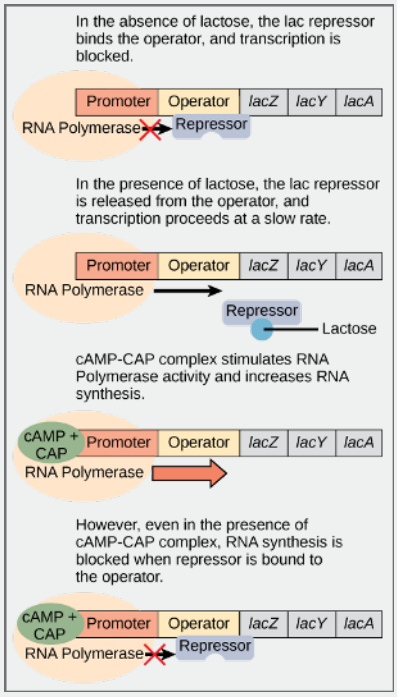


Figure 16.6 Regulation of the *lac* operon. Transcription of the *lac* operon is carefully regulated so that its expression only occurs when glucose is limited and lactose is present to serve as an alternative fuel source.

In *E. coli*, the *trp* operon is on by default, while the *lac* operon is off. Why do you think this is the case?

If glucose is present, then CAP fails to bind to the promoter sequence to activate transcription. If lactose is absent, then the repressor binds to the operator to prevent transcription. If either of these conditions is met, then transcription remains off. Only when glucose is absent and lactose is present is the *lac* operon transcribed (Table 16.2).

Signals that Induce or Repress Transcription of the *lac* Operon

Glucose	CAP binds	Lactose	Repressor binds	Transcription
+	-	-	+	No
+	-	+	-	Some
-	+	-	+	No
-	+	+	-	Yes

Table 16.2

LINK TO LEARNING

LINK TO LEARNING

Watch an animated tutorial about the workings of *lac* operon here.

Click to view content

This page titled 8.7.3: Prokaryotic Gene Regulation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.4: EUKARYOTIC EPIGENETIC GENE REGULATION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Explain how chromatin remodeling controls transcriptional access
- Describe how access to DNA is controlled by histone modification
- Describe how DNA methylation is related to epigenetic gene changes

Eukaryotic gene expression is more complex than prokaryotic gene expression because the processes of transcription and translation are physically separated. Unlike prokaryotic cells, eukaryotic cells can regulate gene expression at many different levels. Epigenetic changes are inheritable changes in gene expression that do not result from changes in the DNA sequence. Eukaryotic gene expression begins with control of access to the DNA. Transcriptional access to the DNA can be controlled in two general ways: chromatin remodeling and DNA methylation. Chromatin remodeling changes the way that DNA is associated with chromosomal histones. DNA methylation is associated with developmental changes and gene silencing.

EPIGENETIC CONTROL: REGULATING ACCESS TO GENES WITHIN THE CHROMOSOME

The human genome encodes over 20,000 genes, with hundreds to thousands of genes on each of the 23 human chromosomes. The DNA in the nucleus is precisely wound, folded, and compacted into chromosomes so that it will fit into the nucleus. It is also organized so that specific segments can be accessed as needed by a specific cell type.

The first level of organization, or packing, is the winding of DNA strands around histone proteins. Histones package and order DNA into structural units called nucleosome complexes, which can control the access of proteins to the DNA regions (Figure 16.7a). Under the electron microscope, this winding of DNA around histone proteins to form nucleosomes looks like small beads on a string (Figure 16.7b).

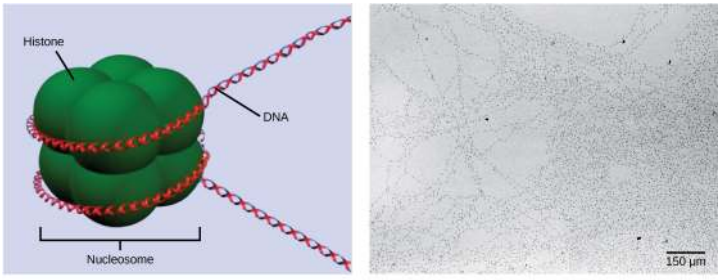


Figure 16.7 DNA is folded around histone proteins to create (a) nucleosome complexes. These nucleosomes control the access of proteins to the underlying DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (credit “micrograph”: modification of work by Chris Woodcock)

These beads (histone proteins) can move along the string (DNA) to expose different sections of the molecule. If DNA encoding a specific gene is to be transcribed into RNA, the nucleosomes surrounding that region of DNA can slide down the DNA to open that specific chromosomal region and allow for the transcriptional machinery (RNA polymerase) to initiate transcription (Figure 16.8).

VISUAL CONNECTION

VISUAL CONNECTION

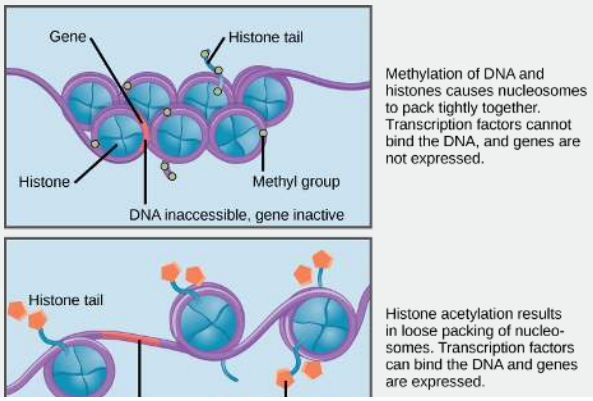


Figure 16.8 Nucleosomes can slide along DNA. When nucleosomes are spaced closely together (top), transcription factors cannot bind and gene expression is turned off. When the nucleosomes are spaced far apart (bottom), the DNA is exposed. Transcription factors can bind, allowing gene expression to occur. Modifications to the histones and DNA affect nucleosome spacing.

In females, one of the two X chromosomes is inactivated during embryonic development because of epigenetic changes to the chromatin. What impact do you think these changes would have on nucleosome packing?

How closely the histone proteins associate with the DNA is regulated by signals found on both the histone proteins and on the DNA. These signals are functional groups added to histone proteins or to DNA and determine whether a chromosomal region should be open or closed (Figure 16.9 depicts modifications to histone proteins and DNA). These tags are not permanent, but may be added or removed as needed. Some chemical groups (phosphate, methyl, or acetyl groups) are attached to specific amino acids in histone "tails" at the N-terminus of the protein. These groups do not alter the DNA base sequence, but they do alter how tightly wound the DNA is around the histone proteins. DNA is a negatively charged molecule and unmodified histones are positively charged; therefore, changes in the charge of the histone will change how tightly wound the DNA molecule will be. By adding chemical modifications like acetyl groups, the charge becomes less positive, and the binding of DNA to the histones is relaxed. Altering the location of nucleosomes and the tightness of histone binding opens some regions of chromatin to transcription and closes others.

The DNA molecule itself can also be modified by methylation. DNA methylation occurs within very specific regions called CpG islands. These are stretches with a high frequency of cytosine and guanine dinucleotide DNA pairs (CG) found in the promoter regions of genes. The cytosine member of the CG pair can be methylated (a methyl group is added). Methylated genes are usually silenced, although methylation may have other regulatory effects. In some cases, genes that are silenced during the development of the gametes of one parent are transmitted in their silenced condition to the offspring. Such genes are said to be imprinted. Parental diet or other environmental conditions may also affect the methylation patterns of genes, which in turn modifies gene expression. Changes in chromatin organization interact with DNA methylation. DNA methyltransferases appear to be attracted to chromatin regions with specific histone modifications. Highly methylated (*hypermethylated*) DNA regions with deacetylated histones are tightly coiled and transcriptionally inactive.

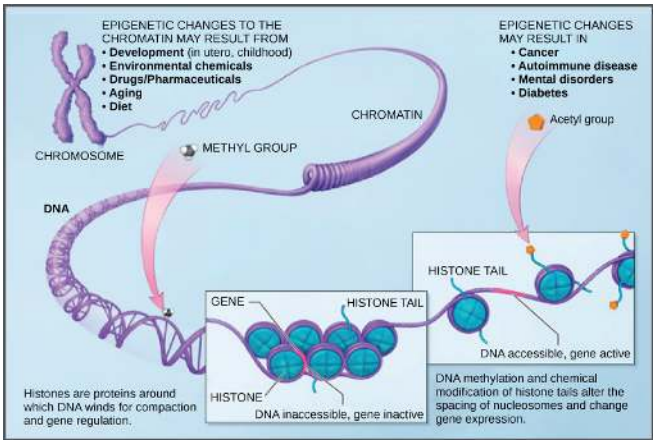


Figure 16.9 Histone proteins and DNA nucleotides can be modified chemically. Modifications affect nucleosome spacing and gene expression. (credit: modification of work by NIH)

Epigenetic changes are not permanent, although they often persist through multiple rounds of cell division and may even cross generational lines. Chromatin remodeling alters the chromosomal structure (open or closed) as needed. If a gene is to be transcribed, the histone proteins and DNA in the chromosomal region encoding that gene are modified in a way that opens the promoter region to allow RNA polymerase and other proteins, called transcription factors, to bind and initiate transcription. If a gene is to remain turned off, or silenced, the histone proteins and DNA have different modifications that signal a closed chromosomal configuration. In this closed configuration, the RNA polymerase and transcription factors do not have access to the DNA and transcription cannot occur (Figure 16.9).

LINK TO LEARNING

LINK TO LEARNING

View this video that describes how epigenetic regulation controls gene expression.

Click to view content

This page titled 8.7.4: Eukaryotic Epigenetic Gene Regulation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.5: EUKARYOTIC TRANSCRIPTION GENE REGULATION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Discuss the role of transcription factors in gene regulation
- Explain how enhancers and repressors regulate gene expression

Like prokaryotic cells, the transcription of genes in eukaryotes requires the action of an RNA polymerase to bind to a DNA sequence upstream of a gene in order to initiate transcription. However, unlike prokaryotic cells, the eukaryotic RNA polymerase requires other proteins, or transcription factors, to facilitate transcription initiation. RNA polymerase by itself cannot initiate transcription in eukaryotic cells. There are two types of transcription factors that regulate eukaryotic transcription: *General (or basal) transcription factors* bind to the core promoter region to assist with the binding of RNA polymerase. *Specific transcription factors* bind to various regions outside of the core promoter region and interact with the proteins at the core promoter to enhance or repress the activity of the polymerase.

LINK TO LEARNING

LINK TO LEARNING

View the process of transcription—the making of RNA from a DNA template.

Click to view content

THE PROMOTER AND THE TRANSCRIPTION MACHINERY

Genes are organized to make the control of gene expression easier. The promoter region is immediately upstream of the coding sequence. This region can be short (only a few nucleotides in length) or quite long (hundreds of nucleotides long). The longer the promoter, the more available space for proteins to bind. This also adds more control to the transcription process. The length of the promoter is gene-specific and can differ dramatically between genes. Consequently, the level of control of gene expression can also differ quite dramatically between genes. The purpose of the promoter is to bind transcription factors that control the initiation of transcription.

Within the core promoter region, 25 to 35 bases upstream of the transcriptional start site, resides the TATA box. The TATA box has the consensus sequence of 5’-TATAAA-3’. The TATA box is the binding site for a protein complex called TFIID, which contains a TATA-

binding protein. Binding of TFIID recruits other transcription factors, including TFIIB, TFIIE, TFIIF, and TFIIH. Some of these transcription factors help to bind the RNA polymerase to the promoter, and others help to activate the transcription initiation complex.

In addition to the TATA box, other binding sites are found in some promoters. Some biologists prefer to restrict the range of the eukaryotic promoter to the core promoter, or polymerase binding site, and refer to these additional sites as promoter-proximal elements, because they are usually found within a few hundred base pairs upstream of the transcriptional start site. Examples of these elements are the CAAT box, with the consensus sequence 5'-CCAAT-3' and the GC box, with the consensus sequence 5'-GGGCGG-3'. Specific transcription factors can bind to these promoter-proximal elements to regulate gene transcription. A given gene may have its own combination of these specific transcription-factor binding sites. There are hundreds of transcription factors in a cell, each of which binds specifically to a particular DNA sequence motif. When transcription factors bind to the promoter just upstream of the encoded gene, it is referred to as a cis-acting element, because it is on the same chromosome just next to the gene. Transcription factors respond to environmental stimuli that cause the proteins to find their binding sites and initiate transcription of the gene that is needed.

ENHANCERS AND TRANSCRIPTION

In some eukaryotic genes, there are additional regions that help increase or enhance transcription. These regions, called enhancers, are not necessarily close to the genes they enhance. They can be located upstream of a gene, within the coding region of the gene, downstream of a gene, or may be thousands of nucleotides away.

Enhancer regions are binding sequences, or sites, for specific transcription factors. When a protein transcription factor binds to its enhancer sequence, the shape of the protein changes, allowing it to interact with proteins at the promoter site. However, since the enhancer region may be distant from the promoter, the DNA must bend to allow the proteins at the two sites to come into contact. DNA bending proteins help to bend the DNA and bring the enhancer and promoter regions together (Figure 16.10). This shape change allows for the interaction of the specific activator proteins bound to the enhancers with the general transcription factors bound to the promoter region and the RNA polymerase.

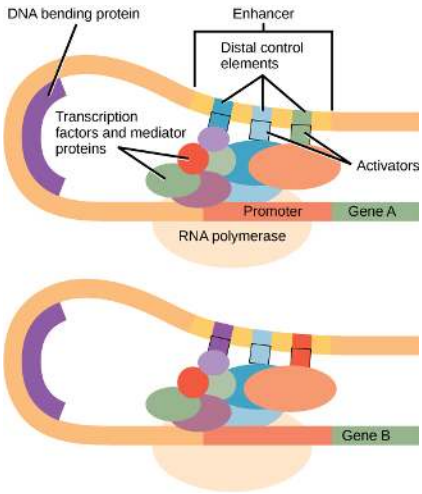


Figure 16.10 Interaction between proteins at the promoter and enhancer sites. An enhancer is a DNA sequence that promotes transcription. Each enhancer is made up of short DNA sequences called distal control elements. Activators bound to the distal control elements interact with mediator proteins and transcription factors. Two different genes may have the same promoter but different distal control elements, enabling differential gene expression.

TURNING GENES OFF: TRANSCRIPTIONAL REPRESSORS

Like prokaryotic cells, eukaryotic cells also have mechanisms to prevent transcription. Transcriptional repressors can bind to promoter or enhancer regions and block transcription. Like the transcriptional activators, repressors respond to external stimuli to prevent the binding of activating transcription factors.

This page titled 8.7.5: Eukaryotic Transcription Gene Regulation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.6: EUKARYOTIC POST-TRANSCRIPTIONAL GENE REGULATION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Understand RNA splicing and explain its role in regulating gene expression
- Describe the importance of RNA stability in gene regulation

RNA is transcribed, but must be processed into a mature form before translation can begin. This processing that takes place after an RNA molecule has been transcribed, but before it is translated into a protein, is called *post-transcriptional modification*. As with the epigenetic and transcriptional stages of processing, this post-transcriptional step can also be regulated to control gene expression in the cell. If the RNA is not processed, shuttled, or translated, then no protein will be synthesized.

RNA SPLICING, THE FIRST STAGE OF POST-TRANSCRIPTIONAL CONTROL

In eukaryotic cells, the RNA transcript often contains regions, called introns, that are removed prior to translation. The regions of RNA that code for protein are called exons. (Figure 16.11). After an RNA molecule has been transcribed, but prior to its departure from the nucleus to be translated, the RNA is processed and the introns are removed by splicing. Splicing is done by spliceosomes, ribonucleoprotein complexes that can recognize the two ends of the intron, cut the transcript at those two points, and bring the exons together for ligation.

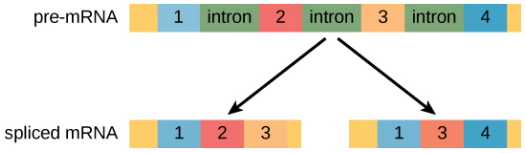


Figure 16.11 Pre-mRNA can be alternatively spliced to create different proteins.

EVOLUTION CONNECTION

EVOLUTION CONNECTION

Alternative RNA SplicingIn the 1970s, genes were first observed that exhibited alternative RNA splicing. Alternative RNA splicing is a mechanism that allows different protein products to be produced from one gene when different combinations of exons are combined to form the mRNA (Figure 16.12). This alternative splicing can be haphazard, but more often it is controlled and acts as a mechanism of *gene regulation*, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells or at different stages of development. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; according to one estimate, 70 percent of genes in humans are expressed as multiple proteins through alternative splicing. Although there are multiple ways to alternatively splice RNA transcripts, the original 5'-3' order of the exons is *always conserved*. That is, a transcript with exons 1 2 3 4 5 6 7 might be spliced 1 2 4 5 6 7 or 1 2 3 6 7, but never 1 2 5 4 3 6 7.

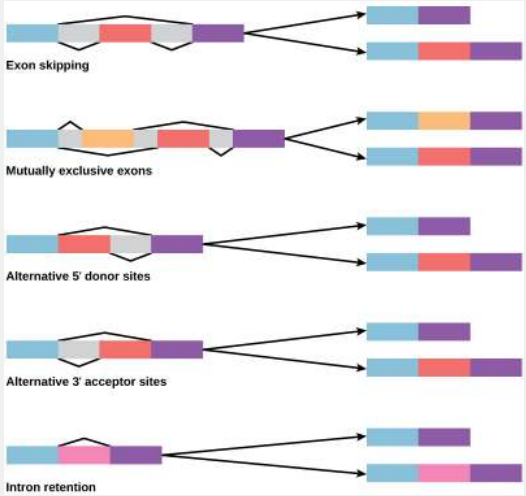


Figure 16.12 There are five basic modes of alternative splicing.

How could alternative splicing evolve? Introns have a beginning- and ending-recognition sequence; it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and instead find the end of the next intron, thus removing two introns and the intervening exon. In fact, there are mechanisms in place to prevent such intron skipping, but mutations are likely to lead to their failure. Such “mistakes” would more than likely produce a nonfunctional protein. Indeed, the cause of many genetic diseases is abnormal splicing rather than mutations in a coding sequence. However, alternative splicing could possibly create a protein variant without the loss of the original protein, opening up possibilities for adaptation of the new variant to new functions. Gene duplication has played an important role in the evolution of new functions in a similar way by providing genes that may evolve without eliminating the original, functional protein.

Question: In the corn snake *Pantherophis guttatus*, there are several different color variants, including amelanistic snakes whose skin patterns display only red and yellow pigments. The cause of amelanism in these snakes was recently identified as the insertion of a transposable element into an intron in the OCA2 (oculocutaneous albinism) gene. How might the insertion of extra genetic material into an intron lead to a nonfunctional protein?

LINK TO LEARNING

LINK TO LEARNING

Visualize how mRNA splicing happens by watching the process in action in this video.

Click to view content

CONTROL OF RNA STABILITY

Before the mRNA leaves the nucleus, it is given two protective "caps" that prevent the ends of the strand from degrading during its journey. 5' and 3' exonucleases can degrade unprotected RNAs. The 5' cap, which is placed on the 5' end of the mRNA, is usually composed of a methylated guanosine triphosphate molecule (GTP). The GTP is placed "backward" on the 5' end of the mRNA, so that the 5' carbons of the GTP and the terminal nucleotide are linked through three phosphates. The poly-A tail, which is attached to the 3' end, is usually composed of a long chain of adenine nucleotides. These changes protect the two ends of the RNA from exonuclease attack.

Once the RNA is transported to the cytoplasm, the length of time that the RNA resides there can be controlled. Each RNA molecule has a defined lifespan and decays at a specific rate. This rate of decay can influence how much protein is in the cell. If the decay rate is increased, the RNA will not exist in the cytoplasm as long, shortening the time available for translation of the mRNA to occur. Conversely, if the rate of decay is decreased, the mRNA molecule will reside in the cytoplasm longer and more protein can be translated. This rate of decay is referred to as the RNA stability. If the RNA is stable, it will be detected for longer periods of time in the cytoplasm.

Binding of proteins to the RNA can also influence its stability. Proteins called RNA-binding proteins, or RBPs, can bind to the regions of the mRNA just upstream or downstream of the protein-coding region. These regions in the RNA that are not translated into protein are

called the untranslated regions, or UTRs. They are not introns (those have been removed in the nucleus). Rather, these are regions that regulate mRNA localization, stability, and protein translation. The region just before the protein-coding region is called the 5' UTR, whereas the region after the coding region is called the 3' UTR (Figure 16.13). The binding of RBPs to these regions can increase or decrease the stability of an RNA molecule, depending on the specific RBP that binds.

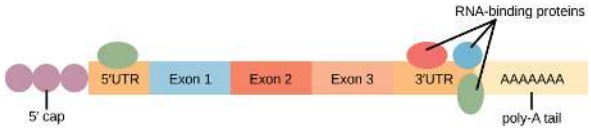


Figure 16.13 RNA-binding proteins. The protein-coding region of this processed mRNA is flanked by 5' and 3' untranslated regions (UTRs). The presence of RNA-binding proteins at the 5' or 3' UTR influences the stability of the RNA molecule.

RNA STABILITY AND MICRORNAS

In addition to RBPs that bind to and control (increase or decrease) RNA stability, other elements called microRNAs can bind to the RNA molecule. These microRNAs, or miRNAs, are short RNA molecules that are only 21 to 24 nucleotides in length. The miRNAs are made in the nucleus as longer pre-miRNAs. These pre-miRNAs are chopped into mature miRNAs by a protein called *Dicer*. Like transcription factors and RBPs, mature miRNAs recognize a specific sequence and bind to the RNA; however, miRNAs also associate with a ribonucleoprotein complex called the RNA-induced silencing complex (RISC). The RNA component of the RISC base-pairs with complementary sequences on an mRNA and either impede translation of the message or lead to the degradation of the mRNA.

This page titled 8.7.6: Eukaryotic Post-transcriptional Gene Regulation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.7: EUKARYOTIC TRANSLATIONAL AND POST-TRANSLATIONAL GENE REGULATION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Understand the process of translation and discuss its key factors
- Describe how the initiation complex controls translation
- Explain the different ways in which the post-translational control of gene expression takes place

After RNA has been transported to the cytoplasm, it is translated into protein. Control of this process is largely dependent on the RNA molecule. As previously discussed, the stability of the RNA will have a large impact on its translation into a protein. As the stability changes, the amount of time that it is available for translation also changes.

THE INITIATION COMPLEX AND TRANSLATION RATE

Like transcription, translation is controlled by proteins that bind and initiate the process. In translation, the complex that assembles to start the process is referred to as the translation initiation complex. In eukaryotes, translation is initiated by binding the initiating met-tRNAi to the 40S ribosome. This tRNA is brought to the 40S ribosome by a protein initiation factor, eukaryotic initiation factor-2 (eIF-2). The eIF-2 protein binds to the high-energy molecule guanosine triphosphate (GTP). The tRNA-eIF2-GTP complex then binds to the 40S ribosome. A second complex forms on the mRNA. Several different initiation factors recognize the 5' cap of the mRNA and proteins bound to the poly-A tail of the same mRNA, forming the mRNA into a loop. The cap-binding protein eIF4F brings the mRNA complex together with the 40S ribosome complex. The ribosome then scans along the mRNA until it finds a start codon AUG. When the anticodon of the initiator tRNA and the start codon are aligned, the GTP is hydrolyzed, the initiation factors are released, and the large 60S ribosomal subunit binds to form the translation complex. The binding of eIF-2 to the RNA is controlled by phosphorylation. If eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP. Therefore, the initiation complex cannot form properly and translation is impeded (Figure 16.14). When eIF-2 remains unphosphorylated, the initiation complex can form normally and translation can proceed.

VISUAL CONNECTION

VISUAL CONNECTION

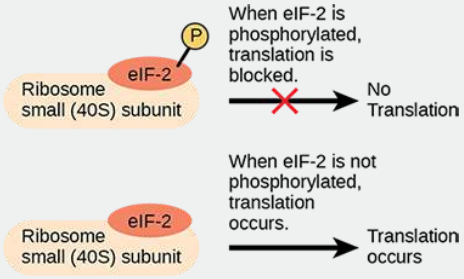


Figure 16.14 Gene expression can be controlled by factors that bind the translation initiation complex.

An increase in phosphorylation levels of eIF-2 has been observed in patients with neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's. What impact do you think this might have on protein synthesis?

CHEMICAL MODIFICATIONS, PROTEIN ACTIVITY, AND LONGEVITY

Proteins can be chemically modified with the addition of groups including methyl, phosphate, acetyl, and ubiquitin groups. The addition or removal of these groups from proteins regulates their activity or the length of time they exist in the cell. Sometimes these modifications can regulate where a protein is found in the cell—for example, in the nucleus, in the cytoplasm, or attached to the plasma membrane.

Chemical modifications occur in response to external stimuli such as stress, the lack of nutrients, heat, or ultraviolet light exposure. These changes can alter epigenetic accessibility, transcription, mRNA stability, or translation—all resulting in changes in expression of various genes. This is an efficient way for the cell to rapidly change the levels of specific proteins in response to the environment. Because proteins are involved in every stage of gene regulation, the phosphorylation of a protein (depending on the protein that is modified) can alter accessibility to the chromosome, can alter translation (by altering transcription factor binding or function), can change nuclear shuttling (by influencing modifications to the nuclear pore complex), can alter RNA stability (by binding or not binding to the RNA to regulate its

stability), can modify translation (increase or decrease), or can change post-translational modifications (add or remove phosphates or other chemical modifications).

The addition of a ubiquitin group to a protein marks that protein for degradation. Ubiquitin acts like a flag indicating that the protein lifespan is complete. These proteins are moved to the proteasome, an organelle that functions to remove proteins, to be degraded (Figure 16.15). One way to control gene expression, therefore, is to alter the longevity of the protein.

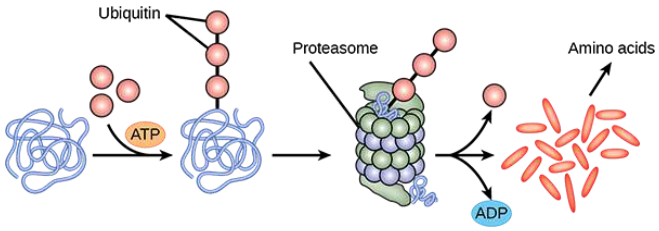


Figure 16.15 Proteins with ubiquitin tags are marked for degradation within the proteasome.

This page titled 8.7.7: Eukaryotic Translational and Post-translational Gene Regulation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.8: CANCER AND GENE REGULATION

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Describe how changes to gene expression can cause cancer
- Explain how changes to gene expression at different levels can disrupt the cell cycle
- Discuss how understanding regulation of gene expression can lead to better drug design

Cancer is not a single disease but includes many different diseases. In cancer cells, mutations modify cell-cycle control and cells don't stop growing as they normally would. Mutations can also alter the growth rate or the progression of the cell through the cell cycle. One example of a gene modification that alters the growth rate is increased phosphorylation of cyclin B, a protein that controls the progression of a cell through the cell cycle and serves as a cell-cycle checkpoint protein.

For cells to move through each phase of the cell cycle, the cell must pass through checkpoints. This ensures that the cell has properly completed the step and has not encountered any mutation that will alter its function. Many proteins, including cyclin B, control these checkpoints. The phosphorylation of cyclin B, a post-translational event, alters its function. As a result, cells can progress through the cell cycle unimpeded, even if mutations exist in the cell and its growth should be terminated. This post-translational change of cyclin B prevents it from controlling the cell cycle and contributes to the development of cancer.

CANCER: DISEASE OF ALTERED GENE EXPRESSION

Cancer can be described as a disease of altered gene expression. There are many proteins that are turned on or off (gene activation or gene silencing) that dramatically alter the overall activity of the cell. A gene that is not normally expressed in that cell can be switched on and expressed at high levels. This can be the result of gene mutation or changes in gene regulation (epigenetic, transcription, post-transcription, translation, or post-translation).

Changes in epigenetic regulation, transcription, RNA stability, protein translation, and post-translational control can be detected in cancer. While these changes don't occur simultaneously in one cancer, changes at each of these levels can be detected when observing cancer at different sites in different individuals. Therefore, changes in histone acetylation (epigenetic modification that leads to gene expression), activation of transcription factors by phosphorylation, increased RNA stability, increased translational control, and protein modification can all be detected at some point in various cancer cells. Scientists are working to understand the common changes that give rise to certain types of cancer or how a modification might be exploited to destroy a tumor cell.

TUMOR SUPPRESSOR GENES, ONCOGENES, AND CANCER

In normal cells, some genes function to prevent excess, inappropriate cell growth. These are tumor-suppressor genes, which are active in normal cells to prevent uncontrolled cell growth. There are many tumor-suppressor genes in cells. The most studied tumor-suppressor gene is p53, which is mutated in over 50 percent of all cancer types. The p53 protein itself functions as a transcription factor. It can bind to sites in the promoters of genes to initiate transcription. Therefore, the mutation of p53 in cancer will dramatically alter the transcriptional activity of its target genes.

LINK TO LEARNING

LINK TO LEARNING

Watch this animation to learn more about the use of p53 in fighting cancer.

Proto-oncogenes are positive cell-cycle regulators. When mutated, proto-oncogenes can become oncogenes and cause cancer. Overexpression of the oncogene can lead to uncontrolled cell growth. This is because oncogenes can alter transcriptional activity, stability, or protein translation of another gene that directly or indirectly controls cell growth. An example of an oncogene involved in cancer is a protein called myc. Myc is a transcription factor that is aberrantly activated in Burkett's Lymphoma, a cancer of the lymph system. Overexpression of myc transforms normal B cells into cancerous cells that continue to grow uncontrollably. High B-cell numbers can result in tumors that can interfere with normal bodily function. Patients with Burkett's lymphoma can develop tumors on their jaw or in their mouth that interfere with the ability to eat.

CANCER AND EPIGENETIC ALTERATIONS

Silencing genes through epigenetic mechanisms is also very common in cancer cells. There are characteristic modifications to histone proteins and DNA that are associated with silenced genes. In cancer cells, the DNA in the promoter region of silenced genes is methylated

on cytosine DNA residues in CpG islands. Histone proteins that surround that region lack the acetylation modification that is present when the genes are expressed in normal cells. This combination of DNA methylation and histone deacetylation (epigenetic modifications that lead to gene silencing) is commonly found in cancer. When these modifications occur, the gene present in that chromosomal region is silenced. Increasingly, scientists understand how epigenetic changes are altered in cancer. Because these changes are temporary and can be reversed—for example, by preventing the action of the histone deacetylase protein that removes acetyl groups, or by DNA methyl transferase enzymes that add methyl groups to cytosines in DNA—it is possible to design new drugs and new therapies to take advantage of the reversible nature of these processes. Indeed, many researchers are testing how a silenced gene can be switched back on in a cancer cell to help re-establish normal growth patterns.

Genes involved in the development of many other illnesses, ranging from allergies to inflammation to autism, are thought to be regulated by epigenetic mechanisms. As our knowledge of how genes are controlled deepens, new ways to treat diseases like cancer will emerge.

CANCER AND TRANSCRIPTIONAL CONTROL

Alterations in cells that give rise to cancer can affect the transcriptional control of gene expression. Mutations that activate transcription factors, such as increased phosphorylation, can increase the binding of a transcription factor to its binding site in a promoter. This could lead to increased transcriptional activation of that gene that results in modified cell growth. Alternatively, a mutation in the DNA of a promoter or enhancer region can increase the binding ability of a transcription factor. This could also lead to the increased transcription and aberrant gene expression that is seen in cancer cells.

Researchers have been investigating how to control the transcriptional activation of gene expression in cancer. Identifying how a transcription factor binds, or a pathway that activates where a gene can be turned off, has led to new drugs and new ways to treat cancer. In breast cancer, for example, many proteins are overexpressed. This can lead to increased phosphorylation of key transcription factors that increase transcription. One such example is the overexpression of the epidermal growth-factor receptor (EGFR) in a subset of breast cancers. The EGFR pathway activates many protein kinases that, in turn, activate many transcription factors which control genes involved in cell growth. New drugs that prevent the activation of EGFR have been developed and are used to treat these cancers.

CANCER AND POST-TRANSCRIPTIONAL CONTROL

Changes in the post-transcriptional control of a gene can also result in cancer. Recently, several groups of researchers have shown that specific cancers have altered expression of miRNAs. Because miRNAs bind to the 3' UTR of RNA molecules to degrade them, overexpression of these miRNAs could be detrimental to normal cellular activity. Too many miRNAs could dramatically decrease the RNA population, leading to a decrease in protein expression. Several studies have demonstrated a change in the miRNA population in specific cancer types. It appears that the subset of miRNAs expressed in breast cancer cells is quite different from the subset expressed in lung cancer cells or even from normal breast cells. This suggests that alterations in miRNA activity can contribute to the growth of breast cancer cells. These types of studies also suggest that if some miRNAs are specifically expressed only in cancer cells, they could be potential drug targets. It would, therefore, be conceivable that new drugs that turn off miRNA expression in cancer could be an effective method to treat cancer.

CANCER AND TRANSLATIONAL/POST-TRANSLATIONAL CONTROL

There are many examples of how translational or post-translational modifications of proteins arise in cancer. Modifications are found in cancer cells from the increased translation of a protein to changes in protein phosphorylation to alternative splice variants of a protein. An example of how the expression of an alternative form of a protein can have dramatically different outcomes is seen in colon cancer cells. The c-Flip protein, a protein involved in mediating the cell-death pathway, comes in two forms: long (c-FLIPL) and short (c-FLIPS). Both forms appear to be involved in initiating controlled cell-death mechanisms in normal cells. However, in colon cancer cells, expression of the long form results in increased cell growth instead of cell death. Clearly, the expression of the wrong protein dramatically alters cell function and contributes to the development of cancer.

NEW DRUGS TO COMBAT CANCER: TARGETED THERAPIES

Scientists are using what is known about the regulation of gene expression in disease states, including cancer, to develop new ways to treat and prevent disease development. Many scientists are designing drugs on the basis of the gene expression patterns within individual tumors. This idea, that therapy and medicines can be tailored to an individual, has given rise to the field of personalized medicine. With an increased understanding of gene regulation and gene function, medicines can be designed to specifically target diseased cells without harming healthy cells. Some new medicines, called targeted therapies, have exploited the overexpression of a specific protein or the mutation of a gene to develop a new medication to treat disease. One such example is the use of anti-EGF receptor medications to treat the subset of breast cancer tumors that have very high levels of the EGF protein. Undoubtedly, more targeted therapies will be developed as scientists learn more about how gene expression changes can cause cancer.

CAREER CONNECTION

CAREER CONNECTION

Clinical Trial CoordinatorA clinical trial coordinator is the person managing the proceedings of the clinical trial. This job includes coordinating patient schedules and appointments, maintaining detailed notes, building the database to track patients (especially for long-term follow-up studies), ensuring proper documentation has been acquired and accepted, and working with the nurses and doctors to facilitate the trial and publication of the results. A clinical trial coordinator may have a science background, like a nursing degree, or other certification. People who have worked in science labs or in clinical offices are also qualified to become a clinical trial coordinator. These jobs are generally in hospitals; however, some clinics and doctor's offices also conduct clinical trials and may hire a coordinator.

This page titled 8.7.8: Cancer and Gene Regulation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.9: KEY TERMS

3' UTR
3' untranslated region; region just downstream of the protein-coding region in an RNA molecule that is not translated

5' cap
a methylated guanosine triphosphate (GTP) molecule that is attached to the 5' end of a messenger RNA to protect the end from degradation

5' UTR
5' untranslated region; region just upstream of the protein-coding region in an RNA molecule that is not translated

activator
protein that binds to prokaryotic operators to increase transcription

catabolite activator protein (CAP)
protein that complexes with cAMP to bind to the promoter sequences of operons which control sugar processing when glucose is not available

***cis*-acting element**
transcription factor binding sites within the promoter that regulate the transcription of a gene adjacent to it

Dicer
enzyme that chops the pre-miRNA into the mature form of the miRNA

DNA methylation
epigenetic modification that leads to gene silencing; a process involving adding a methyl group to the DNA molecule

enhancer
segment of DNA that is upstream, downstream, perhaps thousands of nucleotides away, or on another chromosome that influence the transcription of a specific gene

epigenetic
heritable changes that do not involve changes in the DNA sequence

eukaryotic initiation factor-2 (eIF-2)
protein that binds first to an mRNA to initiate translation

gene expression
processes that control the turning on or turning off of a gene

guanine diphosphate (GDP)
molecule that is left after the energy is used to start translation

guanine triphosphate (GTP)
energy-providing molecule that binds to eIF-2 and is needed for translation

histone acetylation
epigenetic modification that leads to gene expression; a process involving adding or removing an acetyl functional group

inducible operon
operon that can be activated or repressed depending on cellular needs and the surrounding environment

initiation complex
protein complex containing eIF-2 that starts translation

***lac* operon**

operon in prokaryotic cells that encodes genes required for processing and intake of lactose

large 60S ribosomal subunit
second, larger ribosomal subunit that binds to the RNA to translate it into protein

microRNA (miRNA)
small RNA molecules (approximately 21 nucleotides in length) that bind to RNA molecules to degrade them

myc
oncogene that causes cancer in many cancer cells

negative regulator
protein that prevents transcription

operator
region of DNA outside of the promoter region that binds activators or repressors that control gene expression in prokaryotic cells

operon
collection of genes involved in a pathway that are transcribed together as a single mRNA in prokaryotic cells

poly-A tail
a series of adenine nucleotides that are attached to the 3' end of an mRNA to protect the end from degradation

positive regulator
protein that increases transcription

post-transcriptional
control of gene expression after the RNA molecule has been created but before it is translated into protein

post-translational
control of gene expression after a protein has been created

proteasome
organelle that degrades proteins

repressor
protein that binds to the operator of prokaryotic genes to prevent transcription

RISC
protein complex that binds along with the miRNA to the RNA to degrade it

RNA stability
how long an RNA molecule will remain intact in the cytoplasm

RNA-binding protein (RBP)
protein that binds to the 3' or 5' UTR to increase or decrease the RNA stability

small 40S ribosomal subunit
ribosomal subunit that binds to the RNA to translate it into protein

***trans*-acting element**
transcription factor binding site found outside the promoter or on another chromosome that influences the transcription of a particular gene

transcription factor
protein that binds to the DNA at the promoter or enhancer region and that influences transcription of a gene

transcription factor binding site

sequence of DNA to which a transcription factor binds

transcriptional start site

site at which transcription begins

***trp* operon**

series of genes necessary to synthesize tryptophan in prokaryotic cells

tryptophan

amino acid that can be synthesized by prokaryotic cells when necessary

untranslated region

segment of the RNA molecule that is not translated into protein. These regions lie before (upstream or 5') and after (downstream or 3') the protein-coding region

---

This page titled 8.7.9: Key Terms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.7.10: CHAPTER SUMMARY

16.1 REGULATION OF GENE EXPRESSION

While all somatic cells within an organism contain the same DNA, not all cells within that organism express the same proteins. Prokaryotic organisms express most of their genes most of the time. However, some genes are expressed only when they are needed. Eukaryotic organisms, on the other hand, express only a subset of their genes in any given cell. To express a protein, the DNA is first transcribed into RNA, which is then translated into proteins, which are then targeted to specific cellular locations. In prokaryotic cells, transcription and translation occur almost simultaneously. In eukaryotic cells, transcription occurs in the nucleus and is separate from the translation that occurs in the cytoplasm. Gene expression in prokaryotes is mostly regulated at the transcriptional level (some epigenetic and post-translational regulation is also present), whereas in eukaryotic cells, gene expression is regulated at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels.

16.2 PROKARYOTIC GENE REGULATION

The regulation of gene expression in prokaryotic cells occurs at the transcriptional level. There are two majors kinds of proteins that control prokaryotic transcription: repressors and activators. Repressors bind to an operator region to block the action of RNA polymerase. Activators bind to the promoter to enhance the binding of RNA polymerase. Inducer molecules can increase transcription either by inactivating repressors or by activating activator proteins. In the *trp* operon, the *trp* repressor is itself activated by binding to tryptophan. Therefore, if tryptophan is not needed, the repressor is bound to the operator and transcription remains off. The *lac* operon is activated by the CAP (catabolite activator protein), which binds to the promoter to stabilize RNA polymerase binding. CAP is itself activated by cAMP, whose concentration rises as the concentration of glucose falls. However, the *lac* operon also requires the presence of lactose for transcription to occur. Lactose inactivates the *lac* repressor, and prevents the repressor protein from binding to the *lac* operator. With the repressor inactivated, transcription may proceed. Therefore glucose must be absent and lactose must be present for effective transcription of the *lac* operon.

16.3 EUKARYOTIC EPIGENETIC GENE REGULATION

In eukaryotic cells, the first stage of gene-expression control occurs at the epigenetic level. Epigenetic mechanisms control access to the chromosomal region to allow genes to be turned on or off. Chromatin remodeling controls how DNA is packed into the nucleus by regulating how tightly the DNA is wound around histone proteins. The DNA itself may be methylated to selectively silence genes. The addition or removal of chemical modifications (or flags) to histone proteins or DNA signals the cell to open or close a chromosomal region. Therefore, eukaryotic cells can control whether a gene is expressed by controlling accessibility to the binding of RNA polymerase and its transcription factors.

16.4 EUKARYOTIC TRANSCRIPTION GENE REGULATION

To start transcription, general transcription factors, such as TFIID, TFIIB, and others, must first bind to the TATA box and recruit RNA polymerase to that location. Additional transcription factors may also bind to other regulatory elements at the promoter to increase or prevent transcription. In addition to promoter sequences, enhancer regions help augment transcription. Enhancers can be upstream, downstream, within a gene itself, or on other chromosomes. Specific transcription factors bound to enhancer regions may either increase or prevent transcription.

16.5 EUKARYOTIC POST-TRANSCRIPTIONAL GENE REGULATION

Post-transcriptional control can occur at any stage after transcription, including RNA splicing and RNA stability. Once RNA is transcribed, it must be processed to create a mature RNA that is ready to be translated. This involves the removal of introns that do not code for protein. Spliceosomes bind to the signals that mark the exon/intron border to remove the introns and ligate the exons together. Once this occurs, the RNA is mature and can be translated. Alternative splicing can produce more than one mRNA from a given transcript. Different splicing variants may be produced under different conditions.

RNA is created and spliced in the nucleus, but needs to be transported to the cytoplasm to be translated. RNA is transported to the cytoplasm through the nuclear pore complex. Once the RNA is in the cytoplasm, the length of time it resides there before being degraded, called RNA stability, can also be altered to control the overall amount of protein that is synthesized. The RNA stability can be increased, leading to longer residency time in the cytoplasm, or decreased, leading to shortened time and less protein synthesis. RNA stability is controlled by RNA-binding proteins (RPBs) and microRNAs (miRNAs). These RPBs and miRNAs bind to the 5' UTR or the 3' UTR of the RNA to increase or decrease RNA stability. MicroRNAs associated with RISC complexes may repress translation or lead to mRNA breakdown.

16.6 EUKARYOTIC TRANSLATIONAL AND POST-TRANSLATIONAL GENE REGULATION

Changing the status of the RNA or the protein itself can affect the amount of protein, the function of the protein, or how long it is found in the cell. To translate the protein, a protein initiator complex must assemble on the RNA. Modifications (such as phosphorylation) of proteins in this complex can prevent proper translation from occurring. Once a protein has been synthesized, it can be modified (phosphorylated, acetylated, methylated, or ubiquitinated). These post-translational modifications can greatly impact the stability, degradation, or function of the protein.

16.7 CANCER AND GENE REGULATION

Cancer can be described as a disease of altered gene expression. Changes at every level of eukaryotic gene expression can be detected in some form of cancer at some point in time. In order to understand how changes to gene expression can cause cancer, it is critical to understand how each stage of gene regulation works in normal cells. By understanding the mechanisms of control in normal, non-diseased cells, it will be easier for scientists to understand what goes wrong in disease states including complex ones like cancer.

This page titled 8.7.10: Chapter Summary is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

8.7.11: VISUAL CONNECTION QUESTIONS

1.

Figure 16.6 In *E. coli*, the *trp* operon is on by default, while the *lac* operon is off. Why do you think that this is the case?

2.

Figure 16.8 In females, one of the two X chromosomes is inactivated during embryonic development because of epigenetic changes to the chromatin. What impact do you think these changes would have on nucleosome packing?

3.

Figure 16.14 An increase in phosphorylation levels of eIF-2 has been observed in patients with neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's. What impact do you think this might have on protein synthesis?

This page titled 8.7.11: Visual Connection Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

8.7.12: REVIEW QUESTIONS

4.

Control of gene expression in eukaryotic cells occurs at which level(s)?

1. only the transcriptional level
2. epigenetic and transcriptional levels
3. epigenetic, transcriptional, and translational levels
4. epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels

5.

Post-translational control refers to:

1. regulation of gene expression after transcription
2. regulation of gene expression after translation
3. control of epigenetic activation
4. period between transcription and translation

6.

How does the regulation of gene expression support continued evolution of more complex organisms?

1. Cells can become specialized within a multicellular organism.
2. Organisms can conserve energy and resources.
3. Cells grow larger to accommodate protein production.
4. Both A and B.

7.

If glucose is absent, but so is lactose, the *lac* operon will be _____.

1. activated
2. repressed
3. activated, but only partially
4. mutated

8.

Prokaryotic cells lack a nucleus. Therefore, the genes in prokaryotic cells are:

1. all expressed, all of the time
2. transcribed and translated almost simultaneously
3. transcriptionally controlled because translation begins before transcription ends
4. b and c are both true

9.

The *ara* operon is an inducible operon that controls the breakdown of the sugar arabinose. When arabinose is present in a bacterium it binds to the protein AraC, and the complex binds to the initiator site to promote transcription. In this scenario, AraC is a(n) _____.

1. activator
2. inducer
3. repressor
4. operator

10.

What are epigenetic modifications?

1. the addition of reversible changes to histone proteins and DNA
2. the removal of nucleosomes from the DNA
3. the addition of more nucleosomes to the DNA
4. mutation of the DNA sequence

11.

Which of the following are true of epigenetic changes?

1. allow DNA to be transcribed

2. move histones to open or close a chromosomal region
3. are temporary
4. all of the above

12.

The binding of _____ is required for transcription to start.

1. a protein
2. DNA polymerase
3. RNA polymerase
4. a transcription factor

13.

What will result from the binding of a transcription factor to an enhancer region?

1. decreased transcription of an adjacent gene
2. increased transcription of a distant gene
3. alteration of the translation of an adjacent gene
4. initiation of the recruitment of RNA polymerase

14.

A scientist compares the promoter regions of two genes. Gene A's core promoter plus proximal promoter elements encompasses 70bp. Gene B's core promoter plus proximal promoter elements encompasses 250bp. Which of the scientist's hypotheses is most likely to be correct?

1. More transcripts will be made from Gene B.
2. Transcription of Gene A involves fewer transcription factors.
3. Enhancers control Gene B's transcription.
4. Transcription of Gene A is more controlled than transcription of Gene B.

15.

Which of the following are involved in post-transcriptional control?

1. control of RNA splicing
2. control of RNA shuttling
3. control of RNA stability
4. all of the above

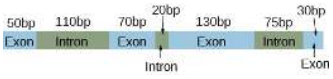
16.

Binding of an RNA binding protein will _____ the stability of the RNA molecule.

1. increase
2. decrease
3. neither increase nor decrease
4. either increase or decrease

17.

An unprocessed pre-mRNA has the following structure.



Which of the following is not a possible size (in bp) of the mature mRNA?

1. 205bp
2. 180bp
3. 150bp
4. 100bp

18.

Alternative splicing has been estimated to occur in more than 95% of multi-exon genes. Which of the following is not an evolutionary advantage of alternative splicing?

1. Alternative splicing increases diversity without increasing genome size.
2. Different gene isoforms can be expressed in different tissues.

3. Alternative splicing creates shorter mRNA transcripts.
4. Different gene isoforms can be expressed during different stages of development.

19.

Post-translational modifications of proteins can affect which of the following?

1. protein function
2. transcriptional regulation
3. chromatin modification
4. all of the above

20.

A scientist mutates eIF-2 to eliminate its GTP hydrolysis capability. How would this mutated form of eIF-2 alter translation?

1. Initiation factors would not be able to bind to mRNA.
2. The large ribosomal subunit would not be able to interact with mRNA transcripts.
3. tRNAi-Met would not scan mRNA transcripts for the start codon.
4. eIF-2 would not be able to interact with the small ribosomal subunit.

21.

Cancer causing genes are called _____.

1. transformation genes
2. tumor suppressor genes
3. oncogenes
4. mutated genes

22.

Targeted therapies are used in patients with a set gene expression pattern. A targeted therapy that prevents the activation of the estrogen receptor in breast cancer would be beneficial to which type of patient?

1. patients who express the EGFR receptor in normal cells
2. patients with a mutation that inactivates the estrogen receptor
3. patients with lots of the estrogen receptor expressed in their tumor
4. patients that have no estrogen receptor expressed in their tumor

---

This page titled 8.7.12: Review Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

8.7.13: CRITICAL THINKING QUESTIONS

23.

Name two differences between prokaryotic and eukaryotic cells and how these differences benefit multicellular organisms.

24.

Describe how controlling gene expression will alter the overall protein levels in the cell.

25.

Describe how transcription in prokaryotic cells can be altered by external stimulation such as excess lactose in the environment.

26.

What is the difference between a repressible and an inducible operon?

27.

In cancer cells, alteration to epigenetic modifications turns off genes that are normally expressed. Hypothetically, how could you reverse this process to turn these genes back on?

28.

A scientific study demonstrated that rat mothering behavior impacts the stress response in their pups. Rats that were born and grew up with attentive mothers showed low activation of stress-response genes later in life, while rats with inattentive mothers had high activation of stress-response genes in the same situation. An additional study that swapped the pups at birth (i.e., rats born to inattentive mothers grew up with attentive mothers and vice versa) showed the same positive effect of attentive mothering. How do genetics and/or epigenetics explain the results of this study?

29.

Some autoimmune diseases show a positive correlation with dramatically decreased expression of histone deacetylase 9 (HDAC9, an enzyme that removes acetyl groups from histones). Why would the decreased expression of HDAC9 cause immune cells to produce inflammatory genes at inappropriate times?

30.

A mutation within the promoter region can alter transcription of a gene. Describe how this can happen.

31.

What could happen if a cell had too much of an activating transcription factor present?

32.

A scientist identifies a potential transcription regulation site 300bp downstream of a gene and hypothesizes that it is a repressor. What experiment (with results) could he perform to support this hypothesis?

33.

Describe how RBPs can prevent miRNAs from degrading an RNA molecule.

34.

How can external stimuli alter post-transcriptional control of gene expression?

35.

Protein modification can alter gene expression in many ways. Describe how phosphorylation of proteins can alter gene expression.

36.

Alternative forms of a protein can be beneficial or harmful to a cell. What do you think would happen if too much of an alternative protein bound to the 3' UTR of an RNA and caused it to degrade?

37.

Changes in epigenetic modifications alter the accessibility and transcription of DNA. Describe how environmental stimuli, such as ultraviolet light exposure, could modify gene expression.

38.

A scientist discovers a virus encoding a Protein X that degrades a subunit of the eIF4F complex. Knowing that this virus transcribes its own mRNAs in the cytoplasm of human cells, why would Protein X be an effective virulence factor?

39.

New drugs are being developed that decrease DNA methylation and prevent the removal of acetyl groups from histone proteins. Explain how these drugs could affect gene expression to help kill tumor cells.

40

10.
How can understanding the gene expression pattern in a cancer cell tell you something about that specific form of cancer?

This page titled 8.7.13: Critical Thinking Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

8.8: KEY TERMS

7-methylguanosine cap
modification added to the 5' end of pre-mRNAs to protect mRNA from degradation and assist translation

aminoacyl tRNA synthetase
enzyme that “charges” tRNA molecules by catalyzing a bond between the tRNA and a corresponding amino acid

anticodon
three-nucleotide sequence in a tRNA molecule that corresponds to an mRNA codon

CAAT box
(GGCCAATCT) essential eukaryotic promoter sequence involved in binding transcription factors

central dogma
states that genes specify the sequence of mRNAs, which in turn specify the sequence of proteins

codon
three consecutive nucleotides in mRNA that specify the insertion of an amino acid or the release of a polypeptide chain during translation

colinear
in terms of RNA and protein, three “units” of RNA (nucleotides) specify one “unit” of protein (amino acid) in a consecutive fashion

consensus
DNA sequence that is used by many species to perform the same or similar functions

core enzyme
prokaryotic RNA polymerase consisting of α, α, β, and β' but missing σ; this complex performs elongation

degeneracy
(of the genetic code) describes that a given amino acid can be encoded by more than one nucleotide triplet; the code is degenerate, but not ambiguous

downstream
nucleotides following the initiation site in the direction of mRNA transcription; in general, sequences that are toward the 3' end relative to a site on the mRNA

exon
sequence present in protein-coding mRNA after completion of pre-mRNA splicing

FACT
complex that “facilitates chromatin transcription” by disassembling nucleosomes ahead of a transcribing RNA polymerase II and reassembling them after the polymerase passes by

GC-rich box
(GGCG) nonessential eukaryotic promoter sequence that binds cellular factors to increase the efficiency of transcription; may be present several times in a promoter

hairpin
structure of RNA when it folds back on itself and forms intramolecular hydrogen bonds between complementary nucleotides

holoenzyme
prokaryotic RNA polymerase consisting of α, α, β, β', and σ; this complex is responsible for transcription initiation

initiation site

nucleotide from which mRNA synthesis proceeds in the 5' to 3' direction; denoted with a “+1”

initiator tRNA

in prokaryotes, called $tRNA_f^{Met}$; in eukaryotes, called $tRNA_i$; a tRNA that interacts with a start codon, binds directly to the ribosome P site, and links to a special methionine to begin a polypeptide chain

intron

non–protein-coding intervening sequences that are spliced from mRNA during processing

Kozak’s rules

determines the correct initiation AUG in a eukaryotic mRNA; the following consensus sequence must appear around the AUG: 5’-GCC(**purine**)CCAUG**G**-3’; the bolded bases are most important

nonsense codon

one of the three mRNA codons that specifies termination of translation

nontemplate strand

strand of DNA that is not used to transcribe mRNA; this strand is identical to the mRNA except that T nucleotides in the DNA are replaced by U nucleotides in the mRNA

Octamer box

(ATTTGCAT) nonessential eukaryotic promoter sequence that binds cellular factors to increase the efficiency of transcription; may be present several times in a promoter

peptidyl transferase

RNA-based enzyme that is integrated into the 50S ribosomal subunit and catalyzes the formation of peptide bonds

plasmid

extrachromosomal, covalently closed, circular DNA molecule that may only contain one or a few genes; common in prokaryotes

poly-A tail

modification added to the 3' end of pre-mRNAs to protect mRNA from degradation and assist mRNA export from the nucleus

polysome

mRNA molecule simultaneously being translated by many ribosomes all going in the same direction

preinitiation complex

cluster of transcription factors and other proteins that recruit RNA polymerase II for transcription of a DNA template

promoter

DNA sequence to which RNA polymerase and associated factors bind and initiate transcription

reading frame

sequence of triplet codons in mRNA that specify a particular protein; a ribosome shift of one or two nucleotides in either direction completely abolishes synthesis of that protein

rho-dependent termination

in prokaryotes, termination of transcription by an interaction between RNA polymerase and the rho protein at a run of G nucleotides on the DNA template

rho-independent

termination sequence-dependent termination of prokaryotic mRNA synthesis; caused by hairpin formation in the mRNA that stalls the polymerase

RNA editing

direct alteration of one or more nucleotides in an mRNA that has already been synthesized

Shine-Dalgarno sequence
(AGGAGG); initiates prokaryotic translation by interacting with rRNA molecules comprising the 30S ribosome

signal sequence
short tail of amino acids that directs a protein to a specific cellular compartment

small nuclear RNA
molecules synthesized by RNA polymerase III that have a variety of functions, including splicing pre-mRNAs and regulating transcription factors

splicing
process of removing introns and reconnecting exons in a pre-mRNA

start codon
AUG (or rarely, GUG) on an mRNA from which translation begins; always specifies methionine

TATA box
conserved promoter sequence in eukaryotes and prokaryotes that helps to establish the initiation site for transcription

template strand
strand of DNA that specifies the complementary mRNA molecule

transcription bubble
region of locally unwound DNA that allows for transcription of mRNA

upstream
nucleotides preceding the initiation site; in general, sequences toward the 5' end relative to a site on the mRNA

This page titled 8.8: Key Terms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.9: CHAPTER SUMMARY

15.1 THE GENETIC CODE

The genetic code refers to the DNA alphabet (A, T, C, G), the RNA alphabet (A, U, C, G), and the polypeptide alphabet (20 amino acids). The central dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three nonsense codons. Most amino acids have several similar codons. Almost every species on the planet uses the same genetic code.

15.2 PROKARYOTIC TRANSCRIPTION

In prokaryotes, mRNA synthesis is initiated at a promoter sequence on the DNA template comprising two consensus sequences that recruit RNA polymerase. The prokaryotic polymerase consists of a core enzyme of four protein subunits and a σ protein that assists only with initiation. Elongation synthesizes mRNA in the 5' to 3' direction at a rate of 40 nucleotides per second. Termination liberates the mRNA and occurs either by rho protein interaction or by the formation of an mRNA hairpin.

15.3 EUKARYOTIC TRANSCRIPTION

Transcription in eukaryotes involves one of three types of polymerases, depending on the gene being transcribed. RNA polymerase II transcribes all of the protein-coding genes, whereas RNA polymerase I transcribes the tandemly duplicated rRNA genes, and RNA polymerase III transcribes various small RNAs, like the 5S rRNA, tRNA, and small nuclear RNA genes. The initiation of transcription in eukaryotes involves the binding of several transcription factors to complex promoter sequences that are usually located upstream of the gene being copied. The mRNA is synthesized in the 5' to 3' direction, and the FACT complex moves and reassembles nucleosomes as the polymerase passes by. Whereas RNA polymerases I and III terminate transcription by protein- or RNA hairpin-dependent methods, RNA polymerase II transcribes for 1,000 or more nucleotides beyond the gene template and cleaves the excess during pre-mRNA processing.

15.4 RNA PROCESSING IN EUKARYOTES

Eukaryotic pre-mRNAs are modified with a 5' methylguanosine cap and a poly-A tail. These structures protect the mature mRNA from degradation and help export it from the nucleus. Pre-mRNAs also undergo splicing, in which introns are removed and exons are reconnected with single-nucleotide accuracy. Only finished mRNAs that have undergone 5' capping, 3' polyadenylation, and intron splicing are exported from the nucleus to the cytoplasm. Pre-rRNAs and pre-tRNAs may be processed by intramolecular cleavage, splicing, methylation, and chemical conversion of nucleotides. Rarely, RNA editing is also performed to insert missing bases after an mRNA has been synthesized.

15.5 RIBOSOMES AND PROTEIN SYNTHESIS

The players in translation include the mRNA template, ribosomes, tRNAs, and various enzymatic factors. The small ribosomal subunit binds to the mRNA template either at the Shine-Dalgarno sequence (prokaryotes) or the 5' cap (eukaryotes). Translation begins at the initiating AUG on the mRNA, specifying methionine. The formation of peptide bonds occurs between sequential amino acids matched to the mRNA template by their tRNAs according to the genetic code. Charged tRNAs enter the ribosomal A site, and their amino acid bonds with the amino acid at the P site. The entire mRNA is translated in three-nucleotide “steps” of the ribosome. When a nonsense codon is encountered, a release factor binds and dissociates the components and frees the new protein. Folding of the protein occurs during and after translation.

This page titled 8.9: Chapter Summary is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

8.10: VISUAL CONNECTION QUESTIONS

1.

Figure 15.11 A scientist splices a eukaryotic promoter in front of a bacterial gene and inserts the gene in a bacterial chromosome. Would you expect the bacteria to transcribe the gene?

2.

Figure 15.13 Errors in splicing are implicated in cancers and other human diseases. What kinds of mutations might lead to splicing errors? Think of different possible outcomes if splicing errors occur.

3.

Figure 15.18 Many antibiotics inhibit bacterial protein synthesis. For example, tetracycline blocks the A site on the bacterial ribosome, and chloramphenicol blocks peptidyl transfer. What specific effect would you expect each of these antibiotics to have on protein synthesis?

Tetracycline would directly affect:

1. tRNA binding to the ribosome
2. ribosome assembly
3. growth of the protein chain

Chloramphenicol would directly affect

1. tRNA binding to the ribosome
2. ribosome assembly
3. growth of the protein chain

This page titled 8.10: Visual Connection Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

SECTION OVERVIEW

8.11: Regulation of Gene Expression

Within most multicellular organisms, every cell contains essentially the same genomic sequence. How then do cells develop and function differently from each other? The answer lies in the regulation of **gene expression**. Only a subset of all the genes is expressed (i.e. are functionally active) in any given cell participating in a particular biological process. Gene expression is regulated at many different steps along the process that converts DNA information into active proteins. In the first stage, transcript abundance can be controlled by regulating the rate of transcription initiation and processing, as well as the degradation of transcripts. In many cases, higher abundance of a gene's transcripts is correlated with its increased expression. In this chapter, we will focus on **transcriptional regulation**. Be aware, however, that cells also regulate the overall activity of genes in other ways. For example, by controlling the rate of mRNA translation, processing, and degradation, as well as the post-translational modification of proteins and protein complexes.



Figure 12.1: *The stickleback is an example of an organism in which mutations cause changes in the regulation of gene expression. These mutations confer a selective advantage in some environments. Natural selection acts on mutations altering gene expression as well as those changing the coding regions of genes. (Flickr-frequency-CC:AND)*

8.11.1: The lac Operon

8.11.2: The Use of Mutants to Study the lac Operon

8.11.3: Eukaryotic Gene Regulation

8.11.4: Regulatory Elements in Evolution

8.11.5: Additional Levels of Regulating Transcription

8.11.6: Epigenetics

8.11.7: Regulation of Gene Expression (Exercises)

8.11.8: Gene Expression Essentials

8.11.S: Regulation of Gene Expression (Summary)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 8.11: Regulation of Gene Expression is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.11.1: The lac Operon

Early insights into mechanisms of transcriptional regulation came from studies of *E. coli* by researchers Francois Jacob & Jacques Monod. In *E. coli*, and many other bacteria, genes encoding several different proteins may be located on a single transcription unit called an **operon**. The genes in an operon share the same transcriptional regulation, but are translated individually. Eukaryotes generally do not group genes together as operons (exception is *C. elegans* and a few other species).

Basic *lac* Operon structure

E. coli encounters many different sugars in its environment. These sugars, such as **lactose** and **glucose**, require different enzymes for their metabolism. Three of the enzymes for lactose metabolism are grouped in the ***lac* operon**: ***lacZ***, ***lacY***, and ***lacA*** (Figure 8.11.1.1). *LacZ* encodes an enzyme called **β-galactosidase**, which digests lactose into its two constituent sugars: glucose and galactose. *lacY* is a **permease** that helps to transfer lactose into the cell. Finally, *lacA* is a **trans-acetylase**; the relevance of which in lactose metabolism is not entirely clear. Transcription of the *lac* operon normally occurs only when lactose is available for it to digest. Presumably, this avoids wasting energy in the synthesis of enzymes for which no substrate is present. A single mRNA transcript includes all three enzyme-coding sequences and is called polycistronic. A cistron is equivalent to a gene.

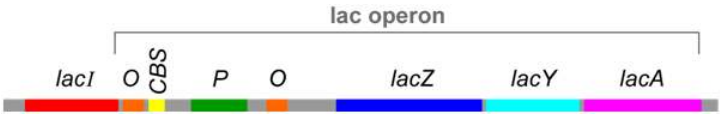


Figure 8.11.1.1: Diagram of a segment of an E. coli chromosome containing the lac operon, as well as the lacI coding region. The various genes and cis-elements are not drawn to scale. (Originall-Deyholos-CC:AN)

cis- and *trans*Regulators

In addition to the three protein-coding genes, the *lac* operon contains short DNA sequences that do not encode proteins, but are instead binding sites for proteins involved in transcriptional regulation of the operon. In the *lac* operon, these sequences are called **P (promoter)**, **O (operator)**, and **CBS (CAP-binding site)**. Collectively, sequence elements such as these are called ***cis*-elements** because they must be located on the same piece of DNA as the genes they regulate. On the other hand, the proteins that bind to these *cis*-elements are called ***trans*-regulators** because (as diffusible molecules) they do not necessarily need to be encoded on the same piece of DNA as the genes they regulate.

lacI is an allosterically regulated repressor

One of the major *trans*-regulators of the *lac* operon is encoded by ***lacI***. Four identical molecules of *lacI* proteins assemble together to form a **homotetramer** called a **repressor** (Figure 8.11.1.2). This repressor binds to two operator sequences adjacent to the promoter of the *lac* operon. Binding of the repressor prevents RNA polymerase from binding to the promoter (Figure 8.11.1.3). Therefore, the operon will not be transcribed when the operator is occupied by a repressor.

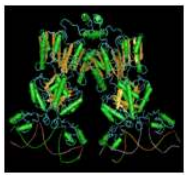


Figure 8.11.1.2: Structure of lacI homotetramer bound to DNA (Origianl-Deyholos-CC:AN)

Besides its ability to bind to specific DNA sequences at the operator, another important property of the *lacI* protein is its ability to bind to lactose. When lactose is bound to *lacI*, the shape of the protein changes in a way that prevents it from binding to the operator. Therefore, in the presence of lactose, RNA polymerase is able to bind to the promoter and transcribe the *lac* operon, leading to a moderate level of expression of the *lacZ*, *lacY*, and *lacA* genes. Proteins such as *lacI* that change their shape and functional properties after binding to a ligand are said to be regulated through an **allosteric** mechanism. The role of *lacI* in regulating the *lac* operon is summarized in Figure 8.11.1.4.

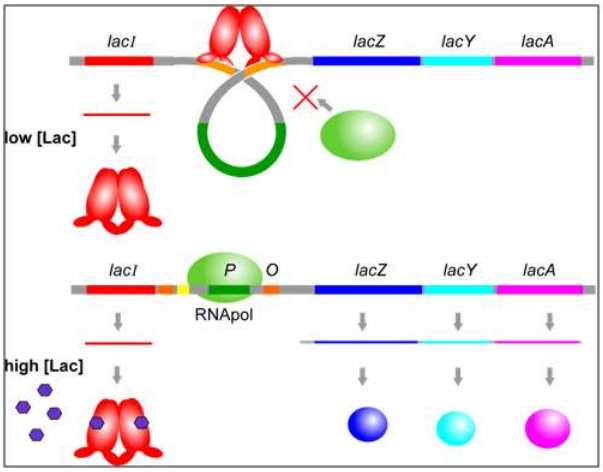


Figure 8.11.1.3: When the concentration of lactose [Lac] is low, lacI tetramers bind to operator sequences (O), thereby blocking binding of RNApol (green) to the promoter (P). Alternatively, when [Lac] is high, lactose binds to lacI, preventing the repressor from binding to O, and allowing transcription by RNApol. (Origianl-Deyholos-CC:AN)

CAP is an allosteric activator of the *lac* operon

A second aspect of *lac* operon regulation is conferred by a *trans*-factor called **cAMP binding protein** (**CAP**, Figure 8.11.1.4). CAP is another example of an allosterically regulated *trans*-factor. Only when the CAP protein is bound to cAMP can another part of the protein bind to a specific *cis*-element within the *lac* promoter called the **CAP binding sequence (CBS)**. CBS is located very close to the promoter (P). When CAP is bound to at CBS, RNA polymerase is better able to bind to the promoter and initiate transcription. Thus, the presence of cAMP ultimately leads to a further increase in *lac* operon transcription.

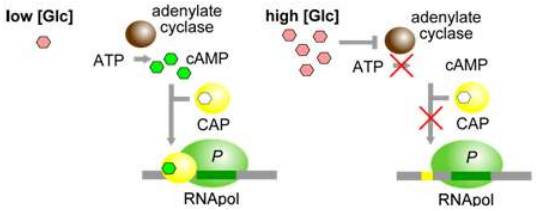


Figure 8.11.1.4: CAP, when bound to cAMP, helps RNApol to bind to the lac operon. cAMP is produced only when glucose [Glc] is low. (Origianl-Deyholos-CC:AN)

The physiological significance of regulation by cAMP becomes more obvious in the context of the following information. The concentration of cAMP is inversely proportional to the abundance of glucose: when glucose concentrations are low, an enzyme called **adenylate cyclase** is able to produce cAMP from ATP. Evidently, *E. coli* prefers glucose over lactose, and so expresses the *lac* operon at high levels only when glucose is absent and lactose is present. This provides another layer of logical control of *lac* operon expression: only in the presence of lactose, and in the absence of glucose is the operon expressed at its highest levels.

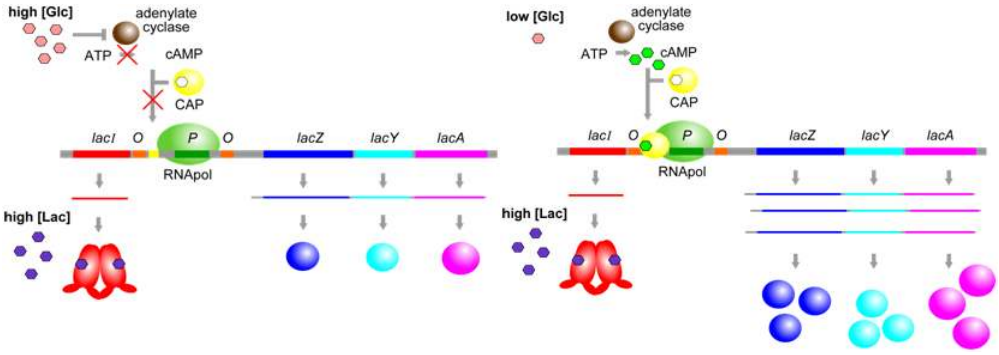


Figure 8.11.1.6: When glucose [Glc] and lactose [Lac] are both high, the lac operon is transcribed at a moderate level, because CAP (in the absence of cAMP) is unable to bind to its corresponding cis-element (yellow) and therefore cannot help to stabilize binding of RNApol at the promoter. Alternatively, when [Glc] is low, and [Lac] is high, CAP and cAMP can bind near the promoter and increase further the transcription of the lac operon. (Origianl-Deyholos-CC:AN)

This page titled 8.11.1: The lac Operon is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **12.1: The lac Operon** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.11.2: The Use of Mutants to Study the lac Operon

Single mutants of the *lac* operon

The *lac* operon and its regulators were first characterized by studying mutants of *E. coli* that exhibited various abnormalities in lactose metabolism. Some mutants expressed the *lac* operon genes constitutively, meaning the operon was expressed whether or not lactose was present in the medium. Such mutant are called **constitutive** mutants.

The operator locus (*lacO*) - One example is $\boldsymbol{O^c}$, in which a mutation in an operator sequence and reduces or precludes the repressor (the *lacI* gene product) from recognizing and binding to the operator sequence. Thus, in O^c mutants, *lacZ*, *lacY*, and *lacA* are expressed whether or not lactose is present.

The *lacI* locus – One type of mutant allele of *lacI* (callled $\boldsymbol{I^-}$) prevents either the production of a repressor polypeptide or produces a polypeptide that cannot bind to the operator sequence. This is also a constitutive expresser of the *lac* operon because absence of repressor binding permits transcription.

Another type of mutant of *lacI* called $\boldsymbol{I^s}$ prevents the repressor polypeptide from binding lactose, and thus will bind to the operator and be non-inducible.. This mutant constitutively represses the *lac* operon whether lactose is present or not. The lac operon is not expressed and this mutant is called a “super-suppressor”.

The F-factor and two lac operons in a single cell – partial diploid in *E.coli*

More can be learned about the regulation of the *lac* operon when two different copies are present in one cell. This can be accomplished by using the **F-factor** to carry one copy, while the other is on the genomic *E. coli* chromosome. This results in a partial diploid in *E. coli.*

The F-factor is an **episome** that is capable of being either a free plasmid or integrated into the host bacterial chromosome. This switching is accomplished by IS elements where unequal crossing over can recombine the F-factor and adjacent DNA sequences (genes) in and out of the host chromosome. Researchers have used this genetic tool to create partial diploids (merozygotes) that allow them to test the regulation with different combinations of different mutations in one cell. For example, the F-factor copy may have a I^S mutation while the genomic copy might have an O^C mutation. How would this cell respond to the presence/absence of lactose (or glucose)? This partial diploid can be used to determine that I^S is dominant to I^+, which in turn is dominant to I^-. It can also be used to show the O^C mutation only acts in *cis*- while the *lacI* mutation can act in *trans*- .

This page titled 8.11.2: The Use of Mutants to Study the lac Operon is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **12.2: The Use of Mutants to Study the lac Operon** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.11.3: Eukaryotic Gene Regulation

Like prokaryotes, transcriptional regulation in eukaryotes involves both *cis*-elements and *trans*-factors, only there are more of them and they interact in a more complex way. A diagram of a typical eukaryotic gene, including several types of *cis*-elements, is shown in Figure 8.11.3.7.

Proximal Regulatory Sequences

As in prokaryotes the RNA polymerase binds to the gene at its promoter to begin transcription. In eukaryotes, however, RNApol is part of a large protein complex that includes additional proteins that bind to one or more specific *cis*-elements in the promoter region, including **GC-rich boxes**, **CAAT boxes**, **and TATA boxes**. High levels of transcription require both the presence of this protein complex at the promoter, as well as their interaction with other *trans*-factors described below. The approximate position of these elements relative to the **transcription start site** ($^{+}1$) is shown in Figure 8.11.3.7, but it should be emphasized that the distance between any of these elements and the transcription start site can vary, but are typically within ~200 base pairs of the start of transcription. This contrasts the next set of elements.

Distal Regulatory elements

Even more variation is observed in the position and orientation of the second major type of *cis*-regulatory element in eukaryotes, which are called **enhancer elements**. Regulatory *trans*-factor proteins called **transcription factors** bind to enhancer sequences, then, while still bound to DNA, these proteins interact with RNApol and other proteins at the promoter to enhance the rate of transcription. There are a wide variety of different transcription factors and each recognizes a specific DNA sequence (enhancer element) to promote gene expression in the adjacent gene under specific circumstances. Because DNA is a flexible molecule, enhancers can be located near (~100s of bp) or far (~10K of bp), and either upstream or downstream, from the promoter (Figures 8.11.3.7 and 8.11.3.8).

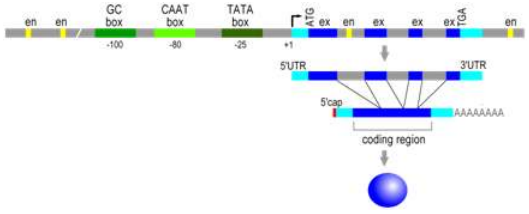


Figure 8.11.3.7: Structure of a typical eukaryotic gene. RNA polymerase binding may involve one or more cis-elements within the proximal region of a promoter (green boxes). Enhancers (yellow boxes) may be located any distance upstream or downstream of the promoter and are also involved in regulating gene expression. The processing of a primary transcript to a mature mRNA is also shown. (Origianl-Deyholos-CC:AN)

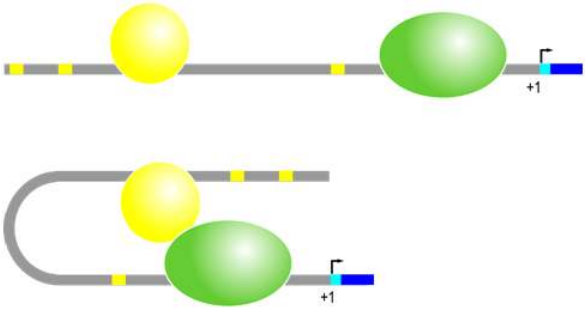


Figure 8.11.3.8: A transcription factor (yellow) bound to an enhancer that is located far from a promoter. Because of the flexibility of the DNA molecule, the transcription factor and RNApol are able to interact physically, even though the cis-elements to which they are bound are located far apart. In eukaryotic cells, RNApol is actually part of a large complex of proteins (not shown here) that assembles at the promoter. (Origianl-Deyholos-CC:AN)

Example 1: Drosophila yellow gene

The *yellow* gene of Drosophila provides an example of the modular nature of enhancers. This gene encodes an enzyme in the pathway that produces a dark pigment in the insect's exoskeleton. Mutants have a yellow cuticle rather than the wild type darker pigmented cuticle. (Why call the gene "yellow": recall that genes are often named after their mutant phenotype.) Figure 8.11.3.9

shows three enhancer elements (left side - wing, body, mouth), each binds a different tissue specific transcription factor to enhancer transcription of *yellow*$^{+}$ in that tissue and makes the pigment. So, the wing cells will have a transcription factor that binds to the wing enhancer to drive expression; likewise in the body and mouth cells. Thus, specific combinations of *cis*-elements and *trans*-factors control the differential, tissue-specific expression of genes. This type of combinatorial action of enhancers is typical of the transcriptional activation of most eukaryotic genes: specific transcription factors activate the transcription of target genes under specific conditions.

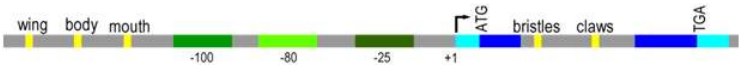


Figure 8.11.3.9: Tissue-specific cis-regulatory elements within a simplified representation of the yellow gene of Drosophila. (Origianl-Deyholos-CC:AN)

While enhancer sequences promote expression, there is an oppositely acting type of element, called **silencers**. These elements function in much the same manner, with transcription factors that bind to DNA sequences, but they act to silence or reduce transcription from the adjacent gene.

Again, a gene's expression profile (transcription level, tissue specific, temporal specific) is a combination of various enhancer and silencer elements.

Example 2: Gal4-UAS system from yeast – a genetic tool

Genetic researchers have taken advantage of a yeast distal enhancer sequence to make the **GAL4-UAS system**, a powerful technique for studying the expression of genes in other eukaryotes. It relies on two parts: a "**driver**" and a "**responder**" (Figure 8.11.3.10). The driver part is a gene encoding a yeast transcriptional activator protein called Gal4. It is separate from the responder part, which contains the enhancer sequence, or upstream activation sequence (UAS, as it is called in yeast) to which the Gal4 protein specifically binds. This UAS is placed upstream (using genetic engineering) from a promoter transcribing a reporter gene, or other gene of interest, such as GFP (green fluorescent protein).

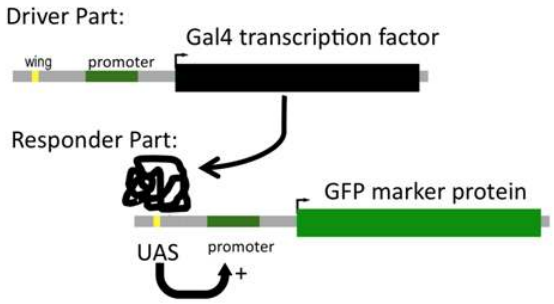


Figure 8.11.3.10: The GAL4-UAS system. The driver, with a wing enhancers, expresses the Gal4 protein that then binds to the UAS element upstream of a marker gene, GFP. This would express the GFP in the wing tissues. The modular aspect of this system would let the wing enhancer be replaced by any other enhancer and the GFP marker replaced with any other gene. (Origianl-Locke-CC:AN)

Both parts must be present in the same cell for the system to express the responder gene. If the driver is absent, the responder product will not be expressed. However, both are in the same cell (or organism) the pattern of expression of the driver part will induce the responder part's expression in the same pattern.

This system works is a variety of eukaryotes, including humans. It has been especially well exploited in Drosophila where 100's (1,000's ?) of differently expressing driver lines are available. These lines permit the tissue specific expression of any responder gene to examine its effect on development.

This page titled 8.11.3: Eukaryotic Gene Regulation is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **12.3: Eukaryotic Gene Regulation** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.11.4: Regulatory Elements in Evolution

Mutations can occur in both *cis*-elements and *trans*-factors; both can result in altered patterns of gene expression. If an altered pattern of gene expression results in a selective advantage (or at least do not produce a major disadvantage), they may be selected and maintained in future populations. They may even contribute to the evolution of new species. An example of a sequence change in an enhancer is found in the *Pitx* gene.

Example: *Pitx* expression in Stickleback

The three-spined **stickleback** (Figure 8.11.4.11) provides an example of natural selection of a mutation in a *cis*-regulatory element. This fish occurs in two forms: (1) populations that inhabit deep, open water and have a spiny pelvic fin that deters larger predator fish from feeding on them; (2) populations from shallow water environments and lack this spiny pelvic fin. In shallow water, it appears that a long, spiny pelvic fin would be a disadvantage because it frequently contacts the sediment at the bottom of the pond and allows parasitic insects in the sediment to invade the stickleback. Researchers compared gene sequences of individuals from both deep and shallow water environments as shown in Figure 8.11.4.11. They observed that in embryos from the deep-water population, a gene called *Pitx* was expressed in several groups of cells, including those that developed into the pelvic fin. Embryos from the shallow-water population expressed *Pitx* in the same groups of cells as the other population, with an important exception: *Pitx* was not expressed in the pelvic fin **primordium** in the shallow-water population. Further genetic analysis showed that the absence of *Pitx* gene expression from the developing pelvic fin of shallow-water stickleback was due to the absence (mutation) of a particular enhancer element upstream of *Pitx*.

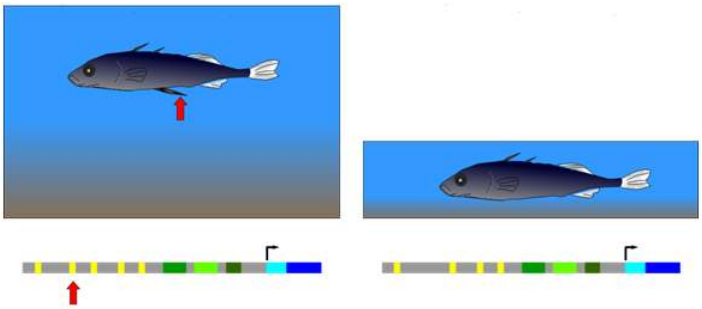


Figure 8.11.4.11: Development of a large, spiny pelvic fin in deep-water stickleback (left) depends on the presence of a particular enhancer element upstream of a gene called Pitx. Mutants lacking this element, and therefore the large pelvic fin (right), have been selected for in shallow-water environments. (Wikipedia-Richard Wheeler-GFDL)

Example: Hemoglobin expression in placental mammals.

Hemoglobin is the oxygen-carrying component of red blood cells (erythrocytes). Hemoglobin usually exists as tetramers of four non-covalently bound hemoglobin molecules (Fig 12.12). Each hemoglobin molecule consists of a **globin** polypeptide with a covalently attached heme molecule. Heme is made through a specialized metabolic pathway and is then bound to globin polypeptide through **post-translational modification**.

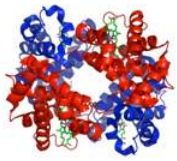


Figure 8.11.4.12: A tetramer of human hemoglobin, type $\alpha 2\beta 2$. The α chains are labeled red, and the β chains are labeled blue. Heme groups are green. (Wikipedia-from Wood, W.G. 1976 Br. Med. Bull. 32, 282-CC:AS)

The composition of the tetramers changes during development (Figure 8.11.4.13). From early childhood onward, most tetramers are of the type $\alpha_2\beta_2$, which means they contain of two copies of each of two slightly different globin proteins named α and β. A small amount of adult hemoglobin is $\alpha_2\delta_2$, which has δ globin instead of the more common β globin. Other tetrameric combinations predominate before birth: $\zeta_2\varepsilon_2$ is most abundant in embryos, and $\alpha_2\gamma_2$ is most abundant in fetuses. Although the six globin proteins (α = alpha, β = beta , γ = gamma, δ =delta, ε =epsilon , ζ = zeta) are very similar to each other, they do have slightly different

functional properties. For example, fetal hemoglobin has a higher oxygen affinity than adult hemoglobin, allowing the fetus to more effectively extract oxygen from maternal blood. The specialized γ globin genes that are characteristic of fetal hemoglobin are found only in placental mammals.

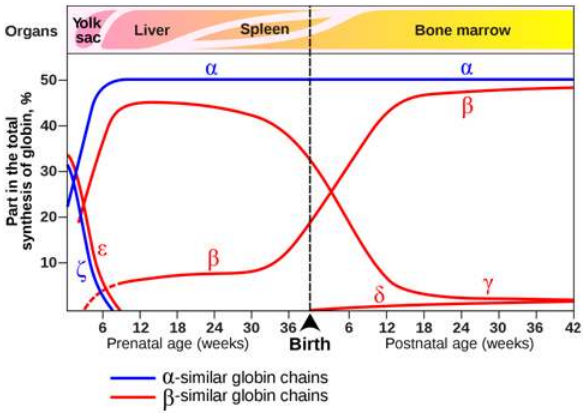


Figure 8.11.4.13: Expression of globin genes during prenatal and postnatal development in humans. The organs in which globin genes are primarily expressed at each developmental stage are also indicated. (Origianl-Deyholos-CC:AN)

Each of these globin polypeptides is encoded by a different gene. In humans, globin genes are located in clusters on two chromosomes (Figure 8.11.4.14). We can infer that these clusters arose through a series of duplications of an ancestral globin gene. **Gene duplication** events can occur through rare errors in processes such as DNA replication, meiosis, or transposition. The duplicated genes can accumulate mutations independently of each other. Mutations can occur in either the regulatory regions (e.g. promoter regions), or in the coding regions, or both. In this way, the promoters of globin genes have evolved to be expressed at different phases of development, and to produce proteins optimized for the prenatal environment.

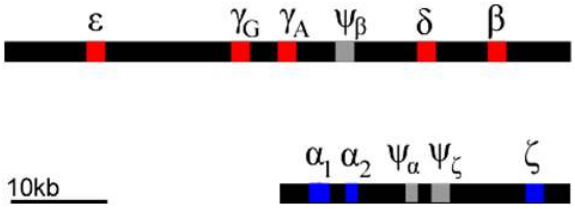


Figure 8.11.4.14: Fragments of human chromosome 11 and human chromosome 16 on which are located clusters of β-like and α-like goblin genes, respectively. Additional globin genes (θ, μ) have also been described by some researchers, but are not shown here. (Origianl-Deyholos-CC:AN)

Of course, not all mutations are beneficial: some mutations can lead to inactivation of one or more of the products of a gene duplication. This can produce what is called a **pseudogene**. Examples of pseudogenes (ψ) are also found in the globin clusters. Pseudogenes have mutations that prevent them from being expressed at all. The globin genes provide an example of how gene duplication and mutation, followed by selection, allows genes to evolve specialized expression patterns and functions. Many genes have evolved as **gene families** in this way, although they are not always clustered together as are the globins.

This page titled 8.11.4: Regulatory Elements in Evolution is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **12.4: Regulatory Elements in Evolution** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.11.5: Additional Levels of Regulating Transcription

Eukaryotes regulate transcription via promoter sequences close to the transcription unit (as in prokaryotes) and also use more distant enhancer sequences to provide more variation in the timing, level, and location of transcription, however, there are still additional levels of genetic control. This consists of two major mechanism: (1) large-scale changes in chromatin structure, and (2) modification of bases in the DNA sequence. These two are often inter-connected.

Chromatin Dynamics

Despite the simplified way in which we often represent DNA in figures such as those in this chapter, DNA is almost always associated with various chromatin proteins. For example, histones remain associated with the DNA even during transcription. Thus the rate of transcription is also controlled by the accessibility of DNA to RNApol and regulatory proteins. So, in regions were the chromatin is highly compacted, it is unlikely that any gene will be transcribed, even if all the necessary *cis*- and *trans*- factors are present in the nucleus. The extent of chromatin compaction in various regions is regulated through the action of **chromatin remodeling** proteins. These protein complexes include enzymes that add or remove chemical tags, such as methyl or acetyl groups, to various DNA bound proteins. These modifications alter the local chromatin density and thus the availability for transcription. **Acetylated** histones, for example, tend to be associated with actively transcribed genes, whereas **deacetylated** histone are associated with genes that are silenced (Figure 8.11.5.15).

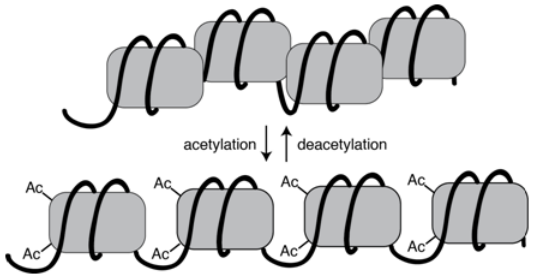


Figure 8.11.5.15: Acetylation of histone proteins is associated with more a more open chromatin configuration. Acetylation is a reversible process. (Origianl-Deyholos-CC:AN)

Likewise, **methylation** of DNA itself is also associated with transcription regulation. Cytosine bases, particularly when followed by a guanine (**CpG sites**) are important targets for DNA methylation (Figure 8.11.5.16). Methylated cytosine within clusters of CpG sites is often associated with transcriptionally inactive DNA.

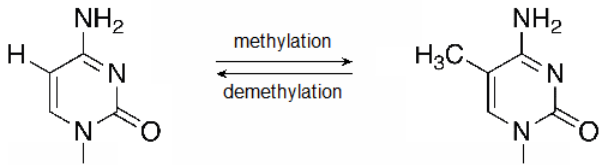


Figure 8.11.5.16: The methylation reaction shown here produces 5-methylcytosine (5mC). Methyl groups may also be removed by various processes. (flickr-Beardy Git-CC:AND)

The modification of DNA and its associated proteins is enzymatically reversible (acetylation/deacetylation; methylation/demethylation) and thus a cyclical activity. Regulation of this provides another layer through which eukaryotic cells control the transcription of specific genes.

This page titled 8.11.5: Additional Levels of Regulating Transcription is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **12.5: Additional Levels of Regulating Transcription** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.11.6: Epigenetics

The word "**epigenetics**" has become popular in the last decade and its meaning has become confused. The term **epigenetics** describes any heritable change in phenotype that is not associated with a change the chromosomal DNA sequence.

Originally it meant the processes through which the genes were expressed to give the phenotype; that is, the changes in gene expression that occur during normal development of multicellular organisms. This includes the change in transcriptional state of a DNA sequence (gene) via DNA or chromatin protein reversible modifications. Thus, DNA methylation and chromatin protein methylation, phosphorylation, and acetylation have been targeted as mechanisms for "heritable" changes in cells as they grow from a single cell (zygote) and differentiate to a multicellular organism. Here, dividing cells commit to differentiate into different tissues such as muscle, neuron, and fibroblast due to the genes that they express or silence. Some genes are irreversibly silenced, through epigenetic mechanisms, in some cell types, but not in others. This doesn't involve any change in DNA sequence.

Remember, these epigenetic effects are not permanent changes and thus cannot be selectable in an evolutionary context. However, mutations in the genes that regulate the epigenetic effect can be selected.

Definition: Epigenetics

Epigenetics describes any heritable change in phenotype that is not associated with a change the chromosomal DNA sequence.

Some heritable information can be passed on independent of the DNA sequence

More recently however, researchers have found many cases of environmentally induced changes in gene expression that can be passed on to subsequent generations – a multi-generational effect. These cases have also been called "**epigenetics**", and probably involve similar reversible changes to the DNA and chromatin proteins.These altered expression patterns represent the diversity of expression for a genome. This "extended" phenotype, the ability to influence traits in the next generation, is a topic of current research and only some examples will be discussed here.

One example comes from the grandchildren of famine victims are known to have lower birth weight than children without a family history of famine. This heritability of altered state of gene expression is surprising, since it appears not to involve typical changes in the sequence of DNA.The term **epigenetics** is applied here since the apparently heritable change in phenotype is associated with something other than chromosomal DNA sequence.

This change is inherited from one generation to the next and is thus **transgenerational.** In develpmental epigenetics, the expression state (developmentally differentiated state) is conserved only from one mitosis to the next, but is erased or rest at meiosis (the boundary of one generation to the next). The basis of at least some types of epigenetic inheritance appears to be replication of patterns of histone and DNA methylation that occurs in parallel with the replication of the primary DNA sequence. It is becoming clear that epigenetics is an important part of biology, and can serve as a type of cellular memory, sometimes within an individual, or sometimes across a few generations, at least.

The permanence of this "change" is not the same as changes in the DNA sequence itself. What is clear is that epigenetics is an important part of regulating gene expression, and can serve as a type of cellular memory, certainly within an individual, or across a few generations in some cases.

Imprinting and Parent-of-Origin Effects

For some genes, the allele inherited from the female parent is expressed differently than the allele that is inherited from the male parent. This is distinct from sex-linkage and is true even if both alleles are wild-type and autosomal. During gamete development (**gametogenesis**), each parent **imprints** epigenetic information on some genes that will affect the activity of the gene in the offspring. Imprinting does not change the DNA sequence, but does involve methylation of DNA and histones, and generally silences the expression of one of the parent's alleles. In humans, some genes are expressed only from the paternal allele, and other genes are expressed only from the maternal allele. The imprinting marks are reprogrammed before the next generation of gametes are formed. Thus, although a male inherits epigenetic information from both his mother and father, this information is erased before sperm development, and he passes only one pattern of imprinting to both his sons and daughters. Most examples of imprinting come from placental mammals, and many imprinted genes control growth rate, such as ***IGF2*** (insulin-like growth factor 2).

Imprinting appears to explain many different **parent-of-origin** effects. For example, **Prader-Willi Syndrome** (**PWS**) and **Angelman Syndrome** (**AS**) are two phenotypically different conditions in humans that result from deletion of a specific region of chromosome 15, which contains several genes. Whether the deletion results in PWS or in AS depends on the parent-of-origin. If the deletion is inherited from the father, PWS results.Conversely, if the deletion is inherited from the mother, AS is the result. The gene(s) involved in PWS is maternally silenced by imprinting, therefore the deletion of its paternally-inherited allele results in a complete deficiency of a required protein.On the other hand, the paternal allele of the gene involved in AS is silenced by imprinting, so deletion of the maternal allele results in deficiency of the protein encoded by that gene.

Transgenerational inheritance of nutritional influences

Nutrition is one aspect of the environment that has been particularly well-studied from an epigenetic perspective in both mice and humans. People alive today who experienced the Dutch famine of 1944-1945 as fetuses have *IGF2* genes that are less-methylated than their siblings. Methylation of *IGF2* (and birth rate) is also lower in children of mothers who do not take folic acid supplements as compared those who do.Furthermore, an individual's phenotype can be influenced by the nutrition of parents or even grandparents.This transgenerational inheritance of nutritional effects appears to involve epigenetic mechanisms.

The mouse ***agouti*** gene produces a signaling molecule that regulates pigment-producing cells and brain cells that affect feeding and body weight. Normally, *agouti* is silenced by methylation, and these mice are brown and have a normal weight. When *agouti* is demethylated by feeding certain chemicals or by mutating a gene that controls methylation, some mice become yellow and overweight, although their DNA sequence remains unchanged.Methylation of *agouti* and normal weight and pigmentation of offspring can be restored if their mothers are fed folic acid and other vitamins during pregnancy.

A study of an isolated Swedish village called Överkalix provides an example of transgenerational inheritance of nutritional factors. Detailed historical records allowed researchers to infer the nutritional status of villagers going back to 1890.The researchers then studied the health of two generations of these villagers' offspring, using medical records. A significant correlation was found between the mortality risk of grandsons and the food availability of their paternal grandfathers.This effect was not seen in the granddaughters. Furthermore, the nutrition of paternal grandmothers, or either of the maternal grandparents did not affect the health of the grandsons. It was therefore proposed that epigenetic information affecting health (specifically diabetes and heart disease) was passed from the grandfathers, to the grandsons, through the male line.

Vernalization as an example of epigenetics

Many plant species in temperate regions are **winter annuals**, meaning that their seeds germinate in the late summer, and grow vegetatively through early fall before entering a dormant phase during the winter, often under a cover of snow. In the spring, the plant resumes growth and is able to produce seeds before other species that germinated in the spring. In order for this life strategy to work, the winter annual must not resume growth or start flower production until winter has ended. **Vernalization** is the name given to the requirement to experience a long period of cold temperatures prior to flowering.

How does a plant sense that winter has passed? The signal for resuming growth cannot simply be warm air temperature, since occasional warm days, followed by long periods of freezing, are common in temperate climates. Researchers have discovered that winter annuals use epigenetic mechanisms to sense and "remember" that winter has occurred



Figure 8.11.6.17: A winter wheat crop (green) in early spring in the English countryside. (Flickr-Beardy Git-CC:AND)

Fortunately for the researchers who were interested in vernalization, some varieties of Arabidopsis are winter annuals. Through mutational analysis of Arabidopsis, researchers found that a gene called *FLC* (*FLOWERING LOCUS C*) encodes a transcription repressor acting on several of the genes involved in early stages of flowering (Figure 8.11.6.16).In the fall and under other warm conditions, the histones associated with *FLC* are acetylated and so *FLC* is transcribed at high levels; expression of flowering genes is therefore entirely repressed. However, in response to cold temperatures, enzymes gradually deacetylate the histones associated with *FLC*. The longer the cold temperatures persist, the more acetyl groups are removed from the *FLC*-associated histones, until

finally the *FLC* locus is no longer transcribed and the flowering genes are free to respond to other environmental and hormonal signals that induce flowering later in the spring. Because the deacetylated state of *FLC* is inherited as cells divide and the plant grows in the early spring, this is an example of a type of cellular memory mediated by an epigenetic mechanism.

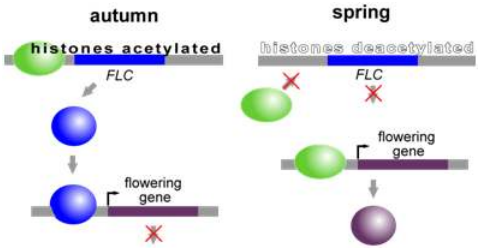


Figure 8.11.6.18: In the autumn, histones associated with FLC are acetylated, allowing this repressor of flowering genes to be expressed. During winter, enzymes progressive deacetylate FLC, preventing it from being expressed, and therefore allowing flowering genes to respond to other signals that induce flowering. (Origianl-Deyholos-CC:AN)

This page titled 8.11.6: Epigenetics is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **12.6: Epigenetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.11.7: Regulation of Gene Expression (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions:

12.1 List all the mechanisms that can be used to regulate gene expression in eukaryotes.

12.2 With respect to the expression of β-galactosidase, what would be the phenotype of each of the following strains of *E. coli*?

a) I^+, O^+, Z^+, Y^+ (no glucose, no lactose)

b) I^+, O^+, Z^+, Y^+ (no glucose, high lactose)

c) I^+, O^+, Z^+, Y^+ (high glucose, no lactose)

d) I^+, O^+, Z^+, Y^+ (high glucose, high lactose)

e) I^+, O^+, Z^-, Y^+ (no glucose, no lactose)

f) I^+, O^+, Z^-, Y^+ (high glucose, high lactose)

g) I^+, O^+, Z^+, Y^- (high glucose, high lactose)

h) I^+, Oc, Z^+, Y^+ (no glucose, no lactose)

i) I^+, Oc,Z^+, Y^+ (no glucose, high lactose)

j) I^+, Oc, Z^+, Y^+ (high glucose, no lactose)

k) I^+, Oc, Z^+, Y^+ (high glucose, high lactose)

l) I^-, O^+, Z^+, Y^+ (no glucose, no lactose)

m) I^-, O^+, Z^+, Y^+ (no glucose, high lactose)

n) I^-, O^+, Z^+, Y^+ (high glucose, no lactose)

o) I^-, O^+, Z^+, Y^+ (high glucose, high lactose)

p) I^s, O^+, Z^+, Y^+ (no glucose, no lactose)

q) I^s, O^+, Z^+, Y^+ (no glucose, high lactose)

r) I^s, O^+, Z^+, Y^+ (high glucose, no lactose)

s) I^s, O^+, Z^+, Y^+ (high glucose, high lactose)

12.3 In the *E. coli* strains listed below, some genes are present on both the chromosome, and the extrachromosomal F^-factor episome. The genotypes of the chromosome and episome are separated by a slash. What will be the β^-galactosidase phenotype of these strains? All of the strains are grown in media that lacks glucose.

a) I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^- (high lactose)

b) I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^- (no lactose)

c) I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+ (high lactose)

d) I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+ (no lactose)

e) I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+ (high lactose)

f) I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+ (no lactose)

g) I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (high lactose)

h) I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (no lactose)

i) I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (high lactose)

j) I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (no lactose)

k) I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+ (high lactose)

l) I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+ (no lactose)

m) I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+ (high lactose)

n) I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+ (no lactose)

o) I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (high lactose)

p) I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (no lactose)

12.1 Transcriptional: initiation, processing & splicing, degradation

Translational: initiation, processing, degradation

Post-translational: modifications (e.g. phosphorylation), localization

Others: histone modification, other chromatin remodeling, DNA methylation

12.2 Legend:

+++ Lots of β-galactosidase activity

\+ Moderate β-galactosidase activity

-- No β-galactosidase activity

-- a) I^+, O^+, Z^+, Y^+ (no glucose, no lactose)

+++ b) I^+, O^+, Z^+, Y^+ (no glucose, high lactose)

-- c) I^+, O^+, Z^+, Y^+ (high glucose, no lactose)

\+ d) I^+, O^+, Z^+, Y^+ (high glucose, high lactose)

-- e) I^+, O^+, Z^-, Y^+ (no glucose, no lactose)

-- f) I^+, O^+, Z^-, Y^+ (high glucose, high lactose)

\+ g) I^+, O^+, Z^+, Y^- (high glucose, high lactose)

+++ h) I^+, Oc, Z^+, Y^+ (no glucose, no lactose)

+++ i) I^+, Oc,Z^+, Y^+ (no glucose, high lactose)

+ j) I^+, Oc, Z^+, Y^+ (high glucose, no lactose)

+ k) I^+, Oc, Z^+, Y^+ (high glucose, high lactose)

+++ l) I^-, O^+, Z^+, Y^+ (no glucose, no lactose)

+++ m) I^-, O^+, Z^+, Y^+ (no glucose, high lactose)

+ n) I^-, O^+, Z^+, Y^+ (high glucose, no lactose)

+ o) I^-, O^+, Z^+, Y^+ (high glucose, high lactose)

-- p) I^s, O^+, Z^+, Y^+ (no glucose, no lactose)

-- q) I^s, O^+, Z^+, Y^+ (no glucose, high lactose)

-- r) I^s, O^+, Z^+, Y^+ (high glucose, no lactose)

-- s) I^s, O^+, Z^+, Y^+ (high glucose, high lactose)

12.3 Legend:

+++ Lots of β-galactosidase activity

+ Moderate β-galactosidase activity

-- No β-galactosidase activity

+++ a) I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^- (high lactose)

-- b) I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^- (no lactose)

+++ c) I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+ (high lactose)

+ d) I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+ (no lactose)

+++ e) I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+ (high lactose)

-- f) I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+ (no lactose)

+++ g) I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (high lactose)

-- h) I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (no lactose)

+++ i) I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (high lactose)

+++ j) I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (no lactose)

+++ k) I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+ (high lactose)

+++ l) I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+ (no lactose)

-- m) I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+ (high lactose)

-- n) I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+ (no lactose)

-- o) I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (high lactose)

-- p) I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+ (no lactose)

12.4 You could demonstrate this with just $I^+O^cZ^-/I^+O^+Z^+$. The fact that this does not have constitutive lactose expression shows that the operator only acts on the same piece of DNA on which it is located. There are also other possible answers.

12.5 You could also demonstrate this with just $I^+O^+Z^-/I^-O^+Z^+$. The fact that this has the same lactose-inducible phenotype as wild-type hows that a functional lacI gene can act on operators on both the same piece of DNA from which it is transcribed, or on a different piece of DNA. There are also other possible answers.

12.6 For all of these, the answer is the same: The l*ac*operon would be inducible by lactose, but only moderate expression of the lac operon would be possible, even in the absence of glucose

a) loss-of-function of adenylate cyclase

b) loss of DNA binding ability of CAP

c) loss of cAMP binding ability of CAP

d) mutation of CAP binding site (CBS) *cis*-element so that CAP could not bind

12.7 Both involve *trans*-factors binding to corresponding *cis*-elements to regulate the initiation of transcription by recruiting or stabilizing the binding of RNApol and related transcriptional proteins at the promoter. In prokaryotes, genes may be regulated as a single operon. In eukaryotes, enhancers may be located much further from the promoter than in prokaryotes.

12.8 These fish would all have spiny tales like the deep-water population.

12.9 These could have arisen from loss-of-function mutation in *FLC*, or in the cis-element to which *FLC* normally binds.

12.10 If there was no deacetylation of *FLC* by HDAC, transcription of FLC might continue constantly, leading to constant suppression of flowering, even after winter.

This page titled 8.11.7: Regulation of Gene Expression (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

8.11.8: Gene Expression Essentials



This page titled 8.11.8: Gene Expression Essentials is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

8.11.S: Regulation of Gene Expression (Summary)

- Regulation of gene expression is essential to the normal development and efficient functioning of cells
- Gene expression may be regulated by many mechanisms, including those affecting transcript abundance, protein abundance, and post-translational modifications
- Regulation of transcript abundance may involve controlling the rate of initiation and elongation of transcription, as well as transcript splicing, stability, and turnover
- The rate of initiation of transcription is related to the presence of RNA polymerase and associated proteins at the promoter.
- RNApol may be blocked from the promoter by repressors, or may be recruited or stabilized at the promoter by other proteins including transcription factors
- The *lac* operon is a classic, fundamental paradigm demonstrating both positive and negative regulation through allosteric effects on *trans*-factors.
- In eukaryotes, *cis*-elements that are usually called enhancers bind to specific *trans*-factors to regulate transcriptional initiation.
- Enhancers may be modular, with each enhancer and its transcription factor regulating a distinct component of a gene's expression pattern, as in the *yellow* gene.
- Sticklebacks provide examples of recent evolutionary events in which mutation of an enhancer produced a change in morphology and a selective advantage.
- Chromatin structure, including reversible modifications such as acetylation of histones, and methylation DNA CpG sites also regulates the initiation of transcription.
- Chromatin modifications or DNA methylation of some genes are heritable over many mitotic, and sometimes even meiotic divisions.
- Heritable changes in phenotype that do not result from a change in DNA sequence are called epigenetic. Many epigenetic phenomena involve regulation of gene expression by chromatin modification and/or DNA methylation.

Key Terms:

gene expression
transcriptional regulation
operon
lactose
glucose
lac operon
lacZ
lacY
lacA
galactosidase
permease
trans-acetylase
P / promoter
O / operator
CBS
CAP-binding site
cis-elements
trans-regulators
lacI
homotetramer
repressor
allosteric
cAMP binding protein
CAP
CAP binding sequence
CBS
adenylate cyclase
constitutive
O^c / I^- / I^s
F-factor / episome
GC boxes
CAAT boxes
TATA boxes
GAL4-UAS
Driver/responder
transcription start site
enhancers/silencers
transcription factors
hemoglobin/heme/globin
pseudogene
gene families
stickleback
primordium
chromatin remodeling
acetylation/deacetylation
methylation/demethylation
CpG sites
epigenetics
winter annual
vernalization
FLC

This page titled 8.11.S: Regulation of Gene Expression (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **12.S: Regulation of Gene Expression (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.12: REVIEW QUESTIONS

4.

The AUC and AUA codons in mRNA both specify isoleucine. What feature of the genetic code explains this?

1. complementarity
2. nonsense codons
3. universality
4. degeneracy

5.

How many nucleotides are in 12 mRNA codons?

1. 12
2. 24
3. 36
4. 48

6.

Which event contradicts the central dogma of molecular biology?

1. Poly-A polymerase enzymes process mRNA in the nucleus.
2. Endonuclease enzymes splice out and repair damaged DNA.
3. Scientists use reverse transcriptase enzymes to make DNA from RNA.
4. Codons specifying amino acids are degenerate and universal.

7.

Which subunit of the *E. coli* polymerase confers specificity to transcription?

1. α
2. β
3. β'
4. σ

8.

The -10 and -35 regions of prokaryotic promoters are called consensus sequences because _____.

1. they are identical in all bacterial species
2. they are similar in all bacterial species
3. they exist in all organisms
4. they have the same function in all organisms

9.

Three different bacteria species have the following consensus sequences upstream of a conserved gene.

	Species A	Species B	Species C
-10	TAATAAT	TTTAAT	TATATT
-35	TTGACA	TTGGCC	TTGAAA

Table 15.2

Order the bacteria from most to least efficient initiation of gene transcription.

1. A > B > C
2. B > C > A
3. C > B > A
4. A > C > B

10.

Which feature of promoters can be found in both prokaryotes and eukaryotes?

1. GC box
2. TATA box

3. octamer box
4. -10 and -35 sequences

11.

What transcripts will be most affected by low levels of α-amanitin?

1. 18S and 28S rRNAs
2. pre-mRNAs
3. 5S rRNAs and tRNAs
4. other small nuclear RNAs

12.

How do enhancers and promoters differ?

1. Enhancers bind transcription factors to silence gene expression, while promoters activate transcription.
2. Enhancers increase the efficiency of gene expression, but are not essential for transcription. Promoter recognition is essential to transcription initiation.
3. Promoters bind transcription factors to increase the efficiency of transcription. Enhancers bind RNA polymerases to initiate transcription.
4. There is no difference. Both are transcription factor-binding sequences in DNA.

13.

Which pre-mRNA processing step is important for initiating translation?

1. poly-A tail
2. RNA editing
3. splicing
4. 7-methylguanosine cap

14.

What processing step enhances the stability of pre-tRNAs and pre-rRNAs?

1. methylation
2. nucleotide modification
3. cleavage
4. splicing

15.

A scientist identifies a pre-mRNA with the following structure.



What is the predicted size of the corresponding mature mRNA in base pairs (bp), excluding the 5’ cap and 3’ poly-A tail?

1. 220bp
2. 295bp
3. 140bp
4. 435bp

16.

The RNA components of ribosomes are synthesized in the _____.

1. cytoplasm
2. nucleus
3. nucleolus
4. endoplasmic reticulum

17.

In any given species, there are at least how many types of aminoacyl tRNA synthetases?

1. 20
2. 40
3. 100
4. 200

18.

A scientist introduces a mutation that makes the 60S ribosomal subunit nonfunctional in a human cell line. What would be the predicted effect on translation?

1. Translation stalls after the initiation AUG codon is identified.
2. The ribosome cannot catalyze the formation of peptide bonds between the tRNAs in the A and P sites.
3. The ribosome cannot interact with mRNAs.
4. tRNAs cannot exit the E site of the ribosome.

This page titled 8.12: Review Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

8.13: CRITICAL THINKING QUESTIONS

19.

Imagine if there were 200 commonly occurring amino acids instead of 20. Given what you know about the genetic code, what would be the shortest possible codon length? Explain.

20.

Discuss how degeneracy of the genetic code makes cells more robust to mutations.

21.

A scientist sequencing mRNA identifies the following strand: CUAUGUGUCGUAACAGCCGAUGACCCG

What is the sequence of the amino acid chain this mRNA makes when it is translated?

22.

If mRNA is complementary to the DNA template strand and the DNA template strand is complementary to the DNA nontemplate strand, then why are base sequences of mRNA and the DNA nontemplate strand not identical? Could they ever be?

23.

In your own words, describe the difference between rho-dependent and rho-independent termination of transcription in prokaryotes.

24.

A fragment of bacterial DNA reads:

3’ –TACCTATAATCTCAATTGATAGAAGCACTCTAC– 5’

Assuming that this fragment is the template strand, what is the sequence of mRNA that would be transcribed? (Hint: Be sure to identify the initiation site.)

25.

A scientist observes that a cell has an RNA polymerase deficiency that prevents it from making proteins. Describe three additional observations that would together support the conclusion that a defect in RNA polymerase I activity, and not problems with the other polymerases, causes the defect.

26.

Chronic lymphocytic leukemia patients often harbor nonsense mutations in their spliceosome machinery. Describe how this mutation of the spliceosome would change the final location and sequence of a pre-mRNA.

27.

Transcribe and translate the following DNA sequence (nontemplate strand): 5'-ATGGCCGGTTATTAAGCA-3'

28.

Explain how single nucleotide changes can have vastly different effects on protein function.

29.

A normal mRNA that reads 5’ – UGCCAUGGUAAUAACACAUGAGGCCUGAAC– 3’ has an insertion mutation that changes the sequence to 5’ -UGCCAUGG**U**UUAAUAACACAUGAGGCCUGAAC– 3’. Translate the original mRNA and the mutated mRNA, and explain how insertion mutations can have dramatic effects on proteins. (Hint: Be sure to find the initiation site.)

This page titled 8.13: Critical Thinking Questions is shared under a not declared license and was authored, remixed, and/or curated by Ying Liu.

CHAPTER OVERVIEW

9: Mutation and Variation

The techniques of genetic analysis discussed in the previous chapters depend on the availability of two or more alleles for a gene of interest. Where do these alleles come from? The short answer is **mutation**. Humans have an interesting relationship with mutation. From our perspective, mutations can be extraordinarily useful, since mutations are need for evolution to occur.Mutation is also essential for the domestication and improvement of almost all of our food. On the other hand, mutations are the cause of many cancers and other diseases, and can be devastating to individuals. Yet, the vast majority of mutations probably go undetected.In this section, we will examine some of the causes and effects of mutations.

- 9.1: Mutation and Polymorphism
- 9.2: Origins of Mutations
- 9.3: Genetic screening for mutations- Forward genetics
- 9.4: Types of Mutations
- 9.5: Some mutations may not have detectable phenotypes
- 9.6: Complementation tests and Allelism
- 9.7: Example of Human Mutations
- 9.8: Linkage and Mapping
  - 9.8.1: Linkage
  - 9.8.2: Recombination
  - 9.8.3: Linkage Reduces Recombination Frequency
  - 9.8.4: Crossovers Allow Recombination of Linked Loci
  - 9.8.5: Inferring Recombination From Genetic Data
  - 9.8.6: Genetic Mapping
  - 9.8.7: Mapping With Three-Point Crosses
  - 9.8.E: Linkage and Mapping (Exercises)
  - 9.8.S: Linkage and Mapping (Summary)
- 9.E: Mutation and Variation (Exercises)
- 9.S: Mutation and Variation (Summary)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

Thumbnail: The difference in appearance between pigmented and white peacocks is due to mutation. (Flickr-ecstaticist-CCANS)

This page titled 9: Mutation and Variation is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Mutation and Polymorphism

We have previously noted that an important property of DNA is its fidelity: most of the time it accurately passes the same information from one generation to the next. However, DNA sequences can also change. Changes in DNA sequences are called **mutations**. If a mutation changes the phenotype of an individual, the individual is said to be a **mutant**. Naturally occurring, but rare, sequence variants that are clearly different from a normal, wild-type sequence are also called mutations.On the other hand, many naturally occurring variants exist for traits for which no clearly normal type can be defined; thus, we use the term **polymorphism** to refer to variants of DNA sequences (and other phenotypes) that co-exist in a population at relatively high frequencies (>1%).

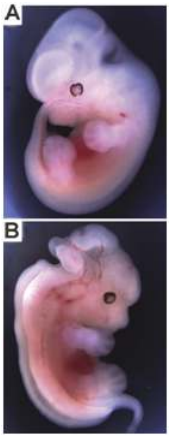


Figure 9.1.1: Examples of wild-type (A) and mutant (B) mouse embryos observed while screening for genes affecting cranium development.(PLoS Biology- Zarbalis, K. et al (2004) PLoS Biology-PD)

Polymorphisms and mutations arise through similar biochemical processes, but the use of the word “polymorphism” avoids implying that any particular allele is more normal or abnormal. For example, a change in a person’s DNA sequence that leads to a disease such as cancer is appropriately called a mutation, but a difference in DNA sequence that explains whether a person has red hair rather than brown or black hair is an example of polymorphism. Molecular markers, which we will discuss in Chapter 9, are a particularly useful type of polymorphism for some areas of genetics research.

This page titled 9.1: Mutation and Polymorphism is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.1: Mutation and Polymorphism** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.2: Origins of Mutations

Types of mutations

Mutations may involve the loss (**deletion**), gain (**insertion**) of one or more base pairs, or else the **substitution** of one or more base pairs with another DNA sequence of equal length. These changes in DNA sequence can arise in many ways, some of which are **spontaneous** and due to natural processes, while others are **induced** by humans intentionally (or unintentionally) using mutagens. There are many ways to classify **mutagens**, which are the agents or processes that cause mutation or increase the frequency of mutations. We will classify mutagens here as being (1) biological, (2) chemical, or (3) physical.

Mutations of biological origin

A major source of spontaneous mutation is errors that arise during DNA replication. DNA polymerases are usually very accurate in adding a base to the growing strand that is the exact complement of the base on the template strand. However, occasionally, an incorrect base is inserted. Usually, the machinery of DNA replication will recognize and repair mispaired bases, but nevertheless, some errors become permanently incorporated in a daughter strand, and so become mutations that will be inherited by the cell's descendents (Figure 9.2.1).

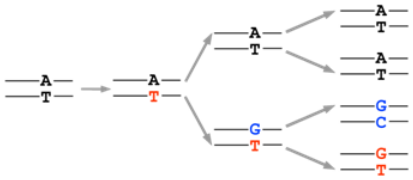


Figure 9.2.1: Mispairing of bases (e.g. G with T) can occur due to tautomerism, alkylating agents, or other effects. As a result, in this example the AT base pair in the original DNA strand will become permanently substituted by a GC based pair in some progeny. The mispaired GT basepair will likely be repaired or eliminated before further rounds of replication. (Original-Deyholos-CC:AN)

Another type of error introduced during replication is caused by a rare, temporary misalignment of a few bases between the template strand and daughter strand (Figure 9.2.2). This **strand-slippage** causes one or more bases on either strand to be temporarily displaced in a **loop** that is not paired with the opposite strand. If this loop forms on the template strand, the bases in the loop may not be replicated, and a **deletion** will be introduced in the growing daughter strand. Conversely, if a region of the daughter strand that has just been replicated becomes displaced in a loop, this region may be replicated again, leading to an **insertion** of additional sequence in the daughter strand, as compared to the template strand.

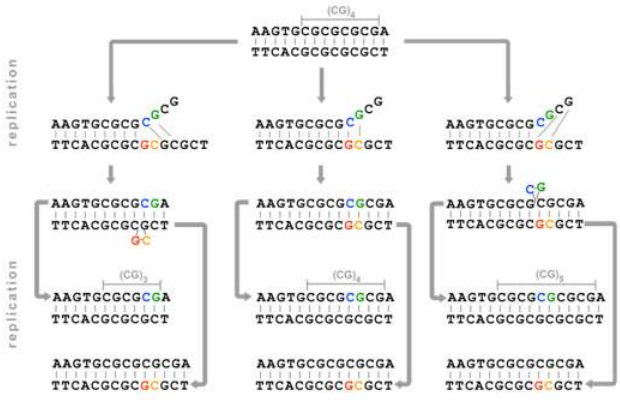


Figure 9.2.2: Strand-slippage can occur occasionally during replication, especially in regions with short, repeated sequences. This can lead to either deletion (left) or insertion (right) of sequences compared to the products of normal replication (center), depending on whether the template strand or daughter strand "loops-out" during replication. (Original-Deyholos-CC:AN)

Consequences: Regions of DNA that have several repeats of the same few nucleotides in a row are especially prone to this type of error during replication. Thus regions with **short-sequence repeats** (**SSRs**) are tend to be highly polymorphic, and are therefore particularly useful in genetics. They are called **microsatellites**.

Mutations can also be caused by the insertion of viruses, transposable elements (transposons), see below, and other types of DNA that are naturally added at more or less random positions in chromosomes. The insertion may disrupt the coding or regulatory sequence of a gene, including the fusion of part of one gene with another. These insertions can occur spontaneously, or they may also be intentionally stimulated in the laboratory as a method of mutagenesis called **transposon-tagging**. For example, a type of transposable element called a ***P* element** is widely used in Drosophila as a biological mutagen. **T-DNA**, which is an insertional element modified from a bacterial pathogen, is used as a mutagen in some plant species.

Mutations due to Transposable Elements

Transposable elements (TEs) are also known as **mobile genetic elements**, or more informally as **jumping genes**. They are present throughout the chromosomes of almost all organisms. These DNA sequences have a unique ability to be cut or copied from their original location and inserted into new locations in the genome. This is called transposition. These insert locations are not entirely random, but TEs can, in principle, be inserted into almost any region of the genome. TEs can therefore insert into genes, disrupting its function and causing a mutation. Researchers have developed methods of artificially increasing the rate of transposition, which makes some TEs a useful type of mutagen. However, the biological importance of TEs extends far beyond their use in mutant screening. TEs are also important causes of disease and phenotypic instability, and they are a major mutational force in evolution.

There are two major classes of TEs in eukaryotes (Figure 9.2.3).

- **Class I** elements include **retrotransposons**; these transpose by means of an RNA intermediate. The TE transcript is reverse transcribed into DNA before being inserted elsewhere in the genome through the action of enzymes such as **integrase**.
- **Class II** elements are known also as **transposons**. They do not use reverse transcriptase or an RNA intermediate for transposition. Instead, they use an enzyme called **transposase** to cut DNA from the original location and then this excised dsDNA fragment is inserted into a new location. Note that the name transposon is sometimes used incorrectly to refer to any type of TEs, but in this book we use transposon to refer specifically to Class II elements.

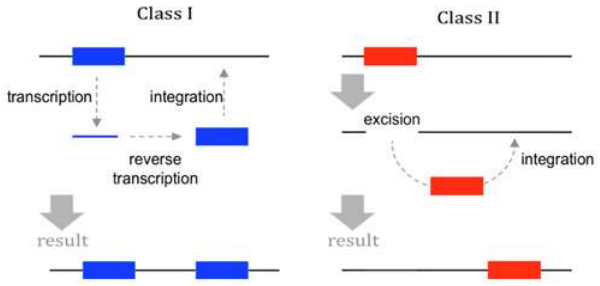


Figure 9.2.3**:** Diagrams of the two main types of transposable elements. (TEs) Class I elements transpose via an ssRNA intermediate, which is reverse transcribed to dsDNA prior to insertion of this copy in a new site in the genome. Class II elements do not involve an RNA intermediate; most Class II elements are cut from their original location as dsDNA, prior to being inserted into a new site in the genome. Although the diagram shows TEs being inserted on the same chromosome as they originated from, TEs can also move to other chromosomes within the same cell.

TEs are relatively short DNA sequences (100-10,000 bp), and encode no more than a few proteins (if any). Normally, the protein-coding genes within a TE are all related to the TE's own transposition functions. These proteins may include **reverse transcriptase**, transposase, and integrase. However, some TEs (of either Class I or II) do not encode any proteins at all. These **non-autonomous** TEs can only transpose if they are supplied with enzymes produced by other, **autonomous** TEs located elsewhere in the genome. In all cases, enzymes for transposition recognize conserved nucleotide sequences within the TE, which dictate where the enzymes begin cutting or copying.

The human genome consists of nearly 45% TEs, the vast majority of which are families of Class I elements called **LINEs** and **SINEs**. The short, ***Alu*** type of SINE occurs in more than one million copies in the human genome (compare this to the approximately 21,000, non-TE, protein-coding genes in humans). Indeed, TEs make up a significant portion of the genomes of almost all eukaryotes. Class I elements, which usually transpose via an RNA **copy-and-paste** mechanism, tend to be more abundant than Class II elements, which mostly use a **cut-and-paste** mechanism. But even the cut-paste mechanism can lead to an increase in TE copy number. For example, if the site vacated by an excised transposon is repaired with a DNA template from a

homologous chromosome that itself contains a copy of a transposon, then the total number of transposons in the genome will increase.

Besides greatly expanding the overall DNA content of genomes, TEs contribute to genome evolution in many other ways. As already mentioned, they may disrupt gene function by insertion into a gene's coding region or regulatory region. More interestingly adjacent regions of chromosomal DNA are sometimes mistakenly transposed along with the TE; this can lead to gene duplication. The duplicated genes are then free to evolve independently, leading in some cases to the development of new functions. The breakage of strands by TE excision and integration can disrupt genes, and can lead to chromosome rearrangement or deletion if errors are made during strand rejoining. Furthermore, having so many similar TE sequences distributed throughout a chromosome sometimes allows mispairing of regions of homologous chromosomes at meiosis, which can cause unequal crossing-over, resulting in deletion or duplication of large segments of chromosomes. Thus, TEs are a potentially important evolutionary force, and may not be included as merely "junk DNA", as they once were.

This page titled 9.2: Origins of Mutations is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.2: Origins of Mutations** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.3: Genetic screening for mutations- Forward genetics

One way to identify genes that affect a particular biological process is to induce random mutations in a large population, and then look for mutants with phenotypes that might be caused by a disruption of a particular biochemical pathway. This strategy of **mutant screening** has been used very effectively to identify and understand the molecular components of hundreds of different biological processes. For example, to find the basic biological processes of memory and learning, researchers have screened mutagenized populations of *Drosophila* to recover flies (or larvae) that lack the normal ability to learn. They lack the ability to associate a particular odor with an electric shock. Because of the similarity of biology among all organisms, some of the genes identified by this mutant screen of a model organism may be relevant to learning and memory in humans, including conditions such as Alzheimer's disease.

– Genetic Screens

In a typical mutant screen, researchers treat a parental population with a mutagen. This may involve soaking seeds in EMS, or mixing a mutagen with the food fed to flies. Usually, no phenotypes are visible among the individuals that are directly exposed to the mutagen because in all the cells every strand of DNA will be affected independently. Thus the induced mutations will be heterozygous and limited to single cells. However, what is most important to geneticists are the mutations in the germline of the mutagenized individuals. The **germline** is defined as the gametes and any of their developmental precursors, and is therefore distinct from the **somatic cells** (i.e. non-reproductive cells) of the body. Because most induced mutations are recessive, the progeny of mutagenized individuals must be mated in a way that allows the new mutations to become homozygous (or hemizygous). Strategies for doing this vary between organisms. In any case, the generation in which induced mutations are expected to occur can be examined for the presence of novel phenotypes. Once a relevant mutant has been identified, geneticists can begin to make inferences about what the normal function of the mutated gene is, based on its mutant phenotype. This can then be investigated further with molecular genetic techniques.

Exposure of an organism to a mutagen causes mutations in essentially random positions along the chromosomes. Most of the mutant phenotypes recovered from a genetic screen are caused by **loss-of-function** mutations. These alleles are due to changes in the DNA sequence that cause it to no longer produce the same level of active protein as the wild-type allele. Loss-of-function alleles tend to be recessive because the wild-type allele is haplosufficient (see Chapter 3). A loss-of-function allele that produces no active protein is called an **amorph**, or **null**. On the other hand, alleles with only a partial loss-of-function are called **hypomorphic**. More rarely, a mutant allele may have a **gain-of-function**, producing either more of the active protein (**hypermorph**) or producing an active protein with a new function (**neomorph**). Finally, **antimorph** alleles have an activity that is dominant and opposite to the wild-type function; antimorphs are also known as **dominant negative** mutations.

This page titled 9.3: Genetic screening for mutations- Forward genetics is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.3: Genetic screening for mutations- Forward genetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.4: Types of Mutations

Mutations (changes in a gene sequence) can result in mutant alleles that no longer produce the same level or type of active product as the wild-type allele. Any mutant allele can be classified into one of five types: (1) amorph, (2) hypomorph, (3) hypermorph, (4) neomorph, and (5) antimorph.

- **Amorph** alleles are complete loss-of-function. They make no active product – zero function. The absence of function can be due to a lack of transcription (gene regulation mutation) or due to the production of a malfunctioning (protein coding mutation) product. These are also sometimes referred to as a **Null** allele.
- **Hypomorph** alleles are only a partial loss-of-function. They make an incompletely functioning product. This could occur via reduced transcription or via the production of a product that lacks complete activity. These alleles are sometimes referred to as **Leaky** mutations, because they provide some function, but not complete function.

Both amorphs and hypomorphs tend to be recessive to wild type because the wild type allele is usually able to supply sufficient product to produce a wild type phenotype (called haplo-sufficient - see Chapter 6). If the mutant allele is not haplo-sufficient, then it will be dominant to the wild type.

While the first two classes involve a **loss-of-function**, the next two involve a **gain-of-function** – quantity or quality. Gain-of-function alleles are almost always dominant to the wild type allele.

- **Hypermorph** alleles produce more of the same, active product. This can occur via increased transcription or by changing the product to make it more efficient/effective at its function.
- **Neomorph** alleles produce an active product with a new, different function, something that the wild type allele doesn't do. It can be either new expression (new tissue or time) or a mutation in the product to create a new function (additional substrate or new binding site), not present in the wild type product.

Antimorph alleles are relatively rare, and have an activity that is dominant and opposite to the wild-type function. These alleles usually have no normal function of their own and they interfere with the function from the wild type allele. Thus, when an antimorph allele is heterozygous with wild type, the wild type allele function is reduced. While at the molecular level there are many ways this can happen, the simplest model to explain antimorph effect is that the product acts as a dimer (or any multimer) and one mutant subunit poisons the whole complex. Antimorphs are also known as dominant negative mutations.

Identifying Muller's Morphs - All mutations can be sorted into one of the five morphs base on how they behave when heterozygous with other alleles – deletion alleles (zero function), wild type alleles (normal function), and duplication alleles (double normal function).

This page titled 9.4: Types of Mutations is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.4: Types of Mutations** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.5: Some mutations may not have detectable phenotypes

Silent Changes

After mutagen treatment, the vast majority of base pair changes (especially substitutions) have no effect on the phenotype. Often, this is because the change occurs in the DNA sequence of a non-coding region of the DNA, such as in **inter-genic regions** (between genes) or within an intron region. Also, even if the change occurs in a base within a codon, it may not change the amino acid that it encodes (recall that the genetic code is degenerate; for example, GCT, GCC, GCA, and GCG all encode alanine) and is referred to as a **silent** mutation. Additionally, the base substitution may change an amino acid, but this doesn't alter the function of the product, so no phenotypic change would occur.

Environment and Genetic Redundancy

There are also situations where a mutation can cause a complete loss-of-function of a gene, yet not produce a change in the phenotype, even when the mutant allele is homozygous. The lack of a phenotypic change can be due to environmental effects: the loss of that gene product may not be apparent in that environment, but might in another. Alternatively, the lack of a phenotype might be attributed to genetic **redundancy**, i.e. the encoding of similarly functioning genes at more than one locus in the genome. Thus the loss of one gene is compensated by another. This important limitation of mutational analysis should be remembered: genes with redundant functions cannot be easily identified by mutant screening.

Essential Genes and Lethal Alleles

Some phenotypes require individuals to reach a particular developmental stage before they can be scored. For example, flower color can only be scored in plants that are mature enough to make flowers, and eye color can only be scored in flies that have developed eyes. However, some alleles may not develop sufficiently to be included among the progeny that are scored for a particular phenotype. Mutations in **essential genes** create **recessive lethal alleles** that arrest the development of an individual at an embryonic stage. This type of mutation may therefore go unnoticed in a typical mutant screen because they are absent from the progeny being screened. Furthermore, the progeny of a monohybrid cross involving an embryonic lethal recessive allele may therefore all be of a single phenotypic class, giving a phenotypic ratio of 1:0 (which is the same as 3:0). In this case the mutation may not be detected.

Naming Genes

Many genes have been first identified in mutant screens, and so they tend to be named after their mutant phenotypes, not the normal function or phenotype. This can cause some confusion for students of genetics. For example, we have already encountered an X-linked gene named *white* in fruit flies. Null mutants of the *white* gene have white eyes, but the normal *white*$^{+}$ allele has red eyes. This tells us that the wild type (normal) function of this gene is actually to help make red eyes. Its product is a protein that imports a pigment precursor into developing cells of the eye. Why don't we call it the "red" gene, since that is what its product does? Because there are more than one-dozen genes that when mutant alter the eye colour; e.g. *violet*, *cinnabar*, *brown*, *scarlet*, etc. For all these genes their function is also needed to make the eye wild type red and not the mutant colour. If we used the name "red" for all these genes it would be confusing, so we use the distinctive mutant phenotype as the gene name. However, this can be problematic, as with the "lethal" mutations described above. This problem is usually handled by giving numbers or locations to the gene name, or making up names that describe how they die (e.g. *even-skipped*, *hunchback*, *hairy*, *runt*, etc.) .

This page titled 9.5: Some mutations may not have detectable phenotypes is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.5: Some mutations may not have detectable phenotypes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.6: Complementation tests and Allelism

– One, or more than one gene?

As explained earlier in this chapter, mutant screening is one of the beginning steps geneticists use to investigate biological processes. When geneticists obtain two independently derived mutants (either from natural populations or during a mutant screen) with similar phenotypes, an immediate question is whether or not the mutant phenotype is due to a loss of function in the same gene, or are they mutant in different genes that both affect the same phenotype (e.g., in the same pathway). That is, are they **allelic mutations**, or **non-allelic mutations**, respectively? This question can be resolved using **complementation tests**, which bring together, or combine, the two mutations under consideration into the same organism to assess the combined phenotype.

hypothetical example of purple flowers

The easiest way to understand a complementation test is by example (Fig.4.9). The pigment in a purple flower could depend on a biochemical pathway much like the **biochemical pathways** leading to the production of arginine in Neurospora (review in Chapter 1). A plant that lacks the function of gene A (genotype *aa*) would produce mutant, white flowers that looked just like the flowers of a plant that lacked the function of gene B (genotype *bb*). (The genetics of two loci are discussed more in the following chapters.) Both A and B are enzymes in the same pathway that leads from a colorless compound#1, thorough colorless compound#2, to the purple pigment. Blocks at either step will result in a mutant white, not wild type purple, flower.

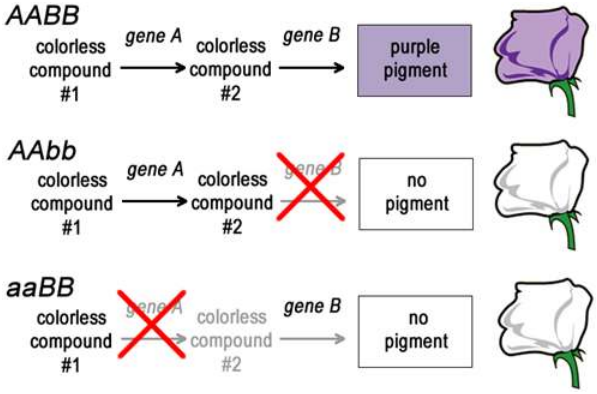


Figure 9.6.9: In this simplified biochemical pathway, two enzymes encoded by two different genes modify chemical compounds in two sequential reactions to produce a purple pigment. Loss of either of the enzymes disrupts the pathway and no pigment is produced. (Original-Deyholos-CC:AN)

Strains with mutations in gene A can be represented as the genotype *aa*, while strains with mutations in gene B can be represented as *bb*. Given that there are two genes here, A and B, then each of these mutant strains can be more completely represented as *aaBB* and *AAbb* . (LEARNING NOTE: Student often forget that genotypes usually only show mutant loci, however, one must remember all the other genes are assumed to be wild type.)

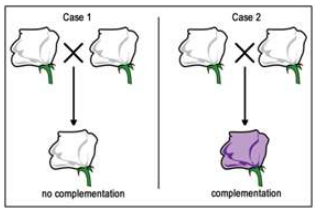
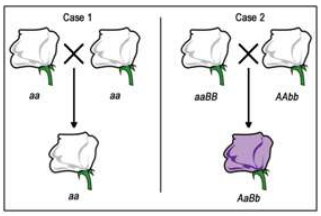
If these two strains are crossed together the resulting progeny will all be *AaBb*. They will have both a wild type, functional A gene and B gene and will thus have a pigmented, purple flower, a wild type phenotype. This is an example of **complementation**. Together, each strain provides what the other is lacking (*AaBb*). The mutations are in different genes and are thus called non-allelic mutations.

Now, if we are presented with a third pure-breeding, independently derived white-flower mutant strain, we won't initially know if it is mutant in gene A or gene B (or possibly some other gene altogether). We can use complementation testing to determine which gene is mutated. To perform a complementation test, two homozygous individuals with similar mutant phenotypes are crossed (Figure 9.6.10).

If the F1 progeny all have the same mutant phenotype (Case 1 - Figure 9.6.10A), then we infer that the same gene is mutated in each parent. These mutations would then be called allelic mutations - in the same gene locus. These mutations FAIL to COMPLEMENT one another (still mutant). These could be either the exact same mutant alleles, or different mutations in the same gene (allelic).

Conversely, if the F1 progeny all appear to be wild-type (Case 2 - Figure 9.6.10B), then each of the parents most likely carries a mutation in a different gene. These mutations would then be called non-allelic mutations - in a different gene locus. These mutations do COMPLEMENT one another.

Note: For mutations to be used in complementation tests they are (1) usually true-breeding (homozygous at the mutant locus), and (2) must be recessive mutations. Dominant mutation CANNOT be used in complementation tests. Also, remember, some mutant strains may have more than one gene locus mutated and thus would fail to complement mutants from more than one other locus (or group).

A.		B.	
	Figure 9.6.10***A*** **–** ***Observation:*** *In a typical complementation test, the genotypes of two parents are unknown (although they must be pure breeding, homozygous mutants). If the F1 progeny all have a mutant phenotype (Case 1), there is no complementation. If the F1 progeny are all wild-type, the mutations have successfully complemented each other.*		***Figure*** 9.6.10***B*** **–** ***Interpretation:*** *The pure breeding, homozygous mutant parents had unknown genotypes before the complementation test, but it could be assumed that they had either mutations in the same genes (Case 1) or in different genes (Case 2). In Case 1, all of the progeny would have the mutant phenotype, because they would all have the same, homozygous genotype as the parents. In Case 2, each parent has a mutation in a different gene, therefore none of the F_1 progeny would be homozygous mutant at either locus. Note that the genotype in Case 1 could be written as either aa or aaBB. (Original-Deyholos-CC:AN)*

This page titled 9.6: Complementation tests and Allelism is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.6: Complementation tests and Allelism** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.7: Example of Human Mutations

– Cystic Fibrosis (CF) – autosmal recessive

Cystic fibrosis (CF) is one of many diseases that geneticists have shown to be caused by mutation of a single, well-characterized gene. Cystic fibrosis is the most common (1/2,500) life-limiting autosomal recessive disease among people of European heritage, with ~ 1 in 25 people being carriers. The frequency varies in different populations. Most of the deaths caused by CF are the result of lung disease, but many CF patients also suffer from other disorders including infertility and gastrointestinal disease. The disease is due to a mutation in the **CFTR** (Cystic Fibrosis Transmembrane Conductance Regulator) gene, which was identified by **Lap-chee Tsui**'s group at the University of Toronto.

Epithelial tissues in some organs rely on the CFTR protein to transport ions (especially Cl^-) across their cell membranes. The passage of ions through a six-sided channel is gated by another part of the CFTR protein, which binds to ATP. If there is insufficient activity of CFTR, an imbalance in ion concentration results, which disrupts the properties of the liquid layer that normally forms at the epithelial surface. In the lungs, this causes mucus to accumulate and can lead to infection. Defects in CFTR also affect pancreas, liver, intestines, and sweat glands, all of which need this ion transport. CFTR is also expressed at high levels in the salivary gland and bladder, but defects in CFTR function do not cause problems in these organs, probably because other ion transporters are able to compensate.

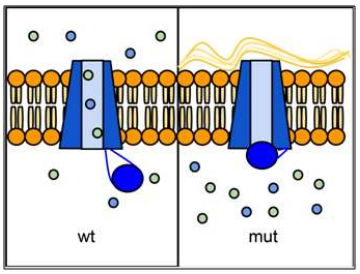


Figure 9.7.11: **Wild-type and mutant forms of CFTR in the cell membrane:** In wild-type, the CFTR ion channel is gated; when activated by ATP, the channel opens and allows ions to move across the membrane. In some CFTR mutants, the channel does not open. This prevents the movement of ions and water and allows mucus to build up on the lung epithelium. (Wikipedia- Lbudd14-CC: AS)

Over one thousand different mutant alleles of CFTR have been described. Any mutation that prevents CFTR from sufficiently transporting ions can lead to cystic fibrosis (CF). Worldwide, the most common CFTR allele among CF patients is called **ΔF508** (delta-F508; or PHE508DEL), which is a deletion of three nucleotides that eliminates a phenylalanine from position 508 of the 1480 aa wild-type protein. Mutation ΔF508 causes CFTR to be folded improperly in the endoplasmic reticulum (ER), which prevents CFTR from reaching the cell membrane. ΔF508 accounts for approximately 70% of CF cases in North America, with ~1/25 people of European descent being carriers. The high frequency of the ΔF508 allele has led to speculation that it may confer some selective advantage to heterozygotes, perhaps by reducing dehydration during cholera epidemics, or by reducing susceptibility to certain pathogens that bind to epithelial membranes.

CFTR is also notable because it is one of the well-characterized genetic diseases for which a drug has been developed that compensates for the effects of a specific mutation. The drug, **Kalydeco**, was approved by the FDA and Health Canada in 2012, decades after the CFTR gene was first mapped to DNA markers (in 1985) and cloned (in 1989). Kalydeco is effective on only some CFTR mutations, most notably G551D (i.e. where glycine is substituted by aspartic acid at position 551 of the protein; GLY551ASP). This mutation is found in less than 5% of CF patients. The G551D mutation affects the ability of ATP to bind to CFTR and open the channel it for transport. Kalydeco compensates for mutation by binding to CFTR and holding it in an open conformation. Kalydeco is expected to cost approximately $250,000 per patient per year.

This page titled 9.7: Example of Human Mutations is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.7: Example of Human Mutations** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

SECTION OVERVIEW

9.8: Linkage and Mapping

As we learned in Chapter 6, Mendel reported that the pairs of loci he observed behaved independently of each other; for example, the segregation of seed color alleles was independent from the segregation of alleles for seed shape. This observation was the basis for his Second Law (Independent Assortment), and contributed greatly to our understanding of heredity. However, further research showed that Mendel's Second Law did not apply to every pair of genes that could be studied. In fact, we now know that alleles of loci that are located close together on the same chromosome tend to be inherited together. This phenomenon is called **linkage**, and is a major exception to Mendel's **Second Law of Independent Assortment.** Researchers use linkage to determine the location of genes along chromosomes in a process called genetic mapping. The concept of gene linkage is important to the natural processes of heredity and evolution.



Figure 7.1: *Linkage affects the frequency at which some combinations of traits are observed. (Wikipedia-Abiyoyo-CC:AN)*

9.8.1: Linkage

9.8.2: Recombination

9.8.3: Linkage Reduces Recombination Frequency

9.8.4: Crossovers Allow Recombination of Linked Loci

9.8.5: Inferring Recombination From Genetic Data

9.8.6: Genetic Mapping

9.8.7: Mapping With Three-Point Crosses

9.8.E: Linkage and Mapping (Exercises)

9.8.S: Linkage and Mapping (Summary)

Template:HideTOC

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University)

The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 9.8: Linkage and Mapping is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

9.8.1: Linkage

As we learned in Chapter 6, Mendel reported that the pairs of loci he observed behaved independently of each other; for example, the segregation of seed color alleles was independent from the segregation of alleles for seed shape. This observation was the basis for his Second Law (Independent Assortment), and contributed greatly to our understanding of heredity. However, further research showed that Mendel's Second Law did not apply to every pair of genes that could be studied. In fact, we now know that alleles of loci that are located close together on the same chromosome tend to be inherited together. This phenomenon is called **linkage**, and is a major exception to Mendel's **Second Law of Independent Assortment.** Researchers use linkage to determine the location of genes along chromosomes in a process called genetic mapping. The concept of gene linkage is important to the natural processes of heredity and evolution.

This page titled 9.8.1: Linkage is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.1: Linkage** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.8.2: Recombination

The term “recombination” is used in several different contexts in genetics. In reference to heredity, **recombination** is defined as any process that results in gametes with combinations of alleles that were not present in the gametes of a previous generation (see Figure 9.8.2.2). **Interchromosomal recombination** occurs either through **independent assortment** of alleles whose loci are on different chromosomes (Chapter 6). **Intrachromosomal recombination** occurs through **crossovers** between loci on the same chromosomes (as described below). It is important to remember that in both of these cases, recombination is a process that occurs during meiosis (mitotic recombination may also occur in some species, but it is relatively rare). If meiosis results in recombination, the products are said to have a **recombinant genotype**. On the other hand, if no recombination occurs during meiosis, the products have their original combinations and are said to have a non-recombinant, or **parental genotype**. Recombination is important because it contributes to the genetic variation that may be observed between individuals within a population and acted upon by selection to produce evolution.

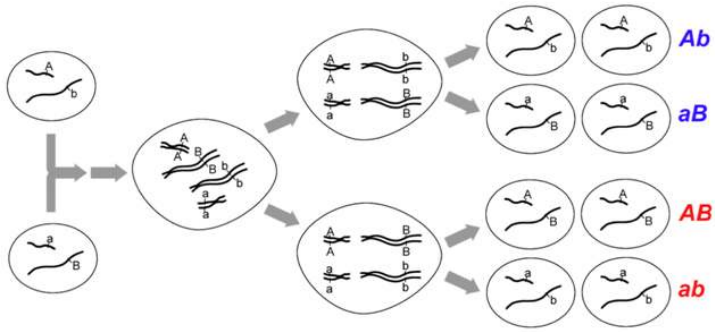


Figure 9.8.2.2: When two loci are on non-homologous chromosomes, their alleles will segregate in combinations identical to those present in the parental gametes (Ab, aB), and in recombinant genotypes (AB, ab) that are different from the parental gametes. (Original-Deyholos-CC:AN)

As an example of interchromosomal recombination, consider loci on two different chromosomes as shown in Figure 9.8.2.2. We know that if these loci are on different chromosomes, there are no physical connections between them, so they are **unlinked** and will segregate independently as did Mendel’s traits. The segregation depends on the relative orientation of each pair of chromosomes at metaphase. Since the orientation is random and independent of other chromosomes, each of the arrangements (and their meiotic products) is equally possible for two unlinked loci as shown in Figure 9.8.2.2. More precisely, there is a 50% probability for recombinant genotypes, and a 50% probability for parental genotypes within the gametes produced by a meiocyte with unlinked loci. Indeed, if we examined all of the gametes that could be produced by this individual (which are the products of multiple independent meioses), we would note that approximately 50% of the gametes would be recombinant, and 50% would be parental. **Recombination frequency** (RF) is simply the number of recombinant gametes, divided by the total number of gametes. A frequency of approximately 50% recombination is therefore a defining characteristic of unlinked loci. Thus the greatest recombinant frequency expected is ~50%.

This page titled 9.8.2: Recombination is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.2: Recombination** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.8.3: Linkage Reduces Recombination Frequency

Having considered unlinked loci above, let us turn to the opposite situation, in which two loci are so close together on a chromosome that the parental combinations of alleles always segregate together (Figure 9.8.3.3). This is because during meiosis they are so close that there are no crossover events between the two loci and the alleles at the two loci are physically attached on the same chromatid and so they always segregate together into the same gamete. In this case, no recombinants will be present following meiosis, and the recombination frequency will be 0%. This is **complete** (or **absolute**) **linkage** and is rare, as the loci must be so close together that crossovers are never detected between them.

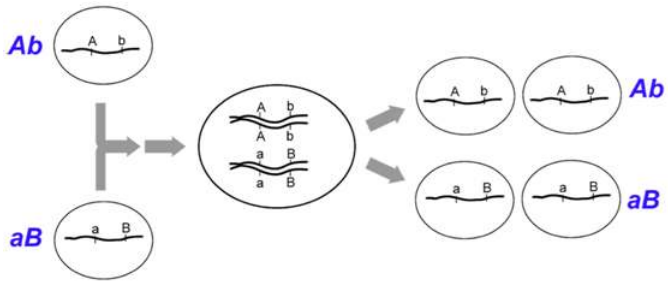


Figure 9.8.3.3: If two loci are completely linked, their alleles will segregate in combinations identical to those present in the parental gametes (Ab, aB). No recombinants will be observed. (Original-Deyholos-CC:AN)

This page titled 9.8.3: Linkage Reduces Recombination Frequency is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.3: Linkage Reduces Recombination Frequency** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.8.4: Crossovers Allow Recombination of Linked Loci

Thus far, we have only considered situations with either no linkage (50% recombination) or complete linkage (0% recombination). It is also possible to obtain recombination frequencies between 0% and 50%, which is a situation we call **incomplete** (or **partial**) **linkage**. Incomplete linkage occurs when two loci are located on the same chromosome but the loci are far enough apart so that crossovers occur between them during some, but not all, meioses. Genes that are on the same chromosome are said to be **syntenic** regardless of whether they are completely or incompletely linked. All linked genes are syntenic, but not all syntenic genes are linked, as we will learn later.

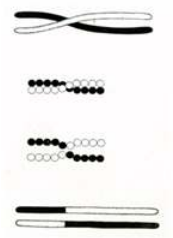


Figure 9.8.4.4: A depiction of a crossover from Morgan's 1916 manuscript. Only one pair of non-sister chromatids is shown, but remember there are 4 chromatids for each homologous pair in meiosis. (Wikipedia-Morgan-PD)

Crossovers occur during prophase I of meiosis, when pairs of homologous chromosomes have aligned with each other in a process called **synapsis**. Crossing over begins with the breakage of DNA of a pair of non-sister chromatids. The breaks occur at corresponding positions on two non-sister chromatids, and then the ends of non-sister chromatids are connected to each other resulting in a reciprocal exchange of double-stranded DNA (Figure 9.8.4.4). Generally every pair of chromosomes has at least one (and often more) crossovers during meioses (Figure 9.8.4.5).

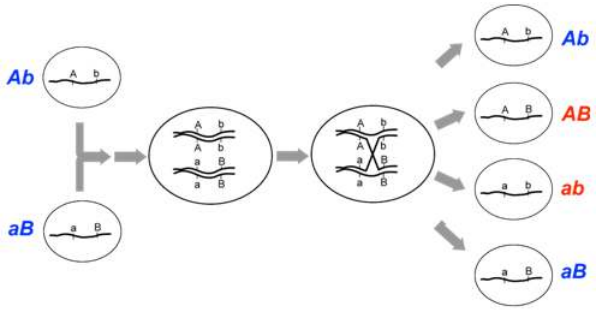


***Figure** 9.8.4.5**:** A crossover between two linked loci can generate recombinant genotypes (AB, ab), from the chromatids involved in the crossover. Remember that multiple, independent meioses occur in each organism, so this particular pattern of recombination will not be observed among all the meioses from this individual. (Original-Deyholos-CC:AN)*

Because the location of crossovers is essentially random along the chromosome, the greater the distance between two loci, the more likely a crossover will occur between them. Furthermore, loci that are on the same chromosome, but are sufficiently far apart from each other, will on average have multiple crossovers between them and they will behave as though they are completely unlinked. A recombination frequency of 50% is therefore the maximum recombination frequency that can be observed, and is indicative of loci that are either on separate chromosomes, or are located very far apart on the same chromosome.

This page titled 9.8.4: Crossovers Allow Recombination of Linked Loci is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.4: Crossovers Allow Recombination of Linked Loci** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.8.5: Inferring Recombination From Genetic Data

In the preceding examples, we had the advantage of knowing the approximate chromosomal positions of each allele involved, before we calculated the recombination frequencies. Knowing this information beforehand made it relatively easy to define the parental and recombinant genotypes, and to calculate recombination frequencies. However, in most experiments, we cannot directly examine the chromosomes, or even the gametes, so we must infer the arrangement of alleles from the phenotypes over two or more generations. Importantly, it is generally not sufficient to know the genotype of individuals in just one generation; for example, given an individual with the genotype *AaBb*, we do not know from the genotype alone whether the loci are located on the same chromosome, and if so, whether the arrangement of alleles on each chromosome is *AB* and *ab* or *Ab* and *aB* (Figure 9.8.5.6). The top cell has the two dominant alleles together and the two recessive alleles together and is said to have the genes in the **coupling** (or **cis**) **configuration**. The alternative shown in the cell below is that the genes are in the **repulsion** (or **trans**) **configuration**.

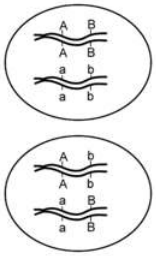


Figure 9.8.5.6: Alleles in coupling configuration (top) or repulsion configuration (bottom). (Original-Deyholos-CC:AN)

Fortunately for geneticists, the arrangement of alleles can sometimes be inferred if the genotypes of a previous generation are known. For example, if the parents of *AaBb* had genotypes *AA*BB and *aabb* respectively, then the parental gametes that fused to produce *AaBb* would have been genotype *AB* and genotype *ab*. Therefore, prior to meiosis in the dihybrid, the arrangement of alleles would likewise be *AB* and *ab* (Figure 9.8.5.7). Conversely, if the parents of *AaBb* had genotypes *aaBB* and *AAbb*, then the arrangement of alleles on the chromosomes of the dihybrid would be *aB* and *Ab*. Thus, the genotype of the previous generation can determine which of an individual's gametes are considered recombinant, and which are considered parental.

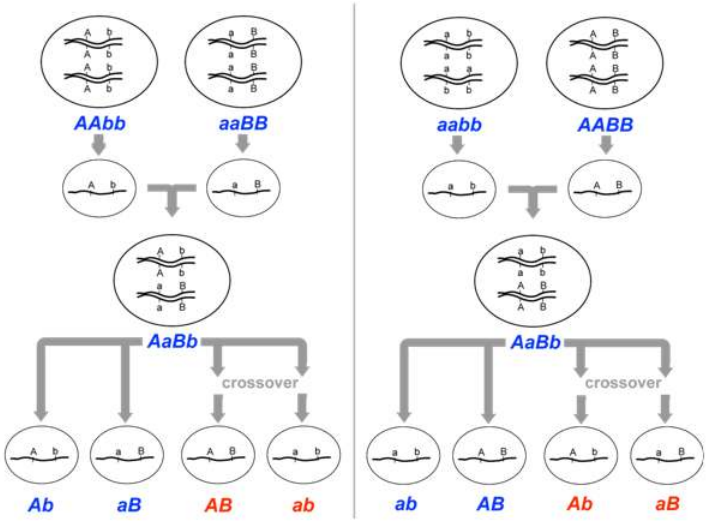


Figure 9.8.5.7: The genotype of gametes can be inferred unambiguously if the gametes are produced by homozygotes. However, recombination frequencies can only be measured among the progeny of heterozygotes (i.e. dihybrids). Note that the dihybrid on the left contains a different configuration of alleles than the dihybrid on the rightdue to differences in the genotypes of their respective parents. Therefore, different gametes are defined as recombinant and parental among the progeny of the two dihybrids. In the cross at left, the recombinant gametes will be genotype AB and ab, and in the cross on the right, the recombinant gametes will be Ab and aB. (Original-Deyholos-CC:AN)

Let us now consider a complete experiment in which our objective is to measure recombination frequency (Figure 9.8.5.8). We need at least two alleles for each of two genes, and we must know which combinations of alleles were present in the parental gametes. The simplest way to do this is to start with pure-breeding lines that have contrasting alleles at two loci. For example, we could

cross short-tailed mice, brown mice (*aaBB*) with long-tailed, white mice (*AAbb*). Based on the genotypes of the parents, we know that the parental gametes will be *aB* or *Ab* (but not *ab* or *AB*), and all of the progeny will be dihybrids, *AaBb*. We do not know at this point whether the two loci are on different pairs of homologous chromosomes, or whether they are on the same chromosome, and if so, how close together they are.

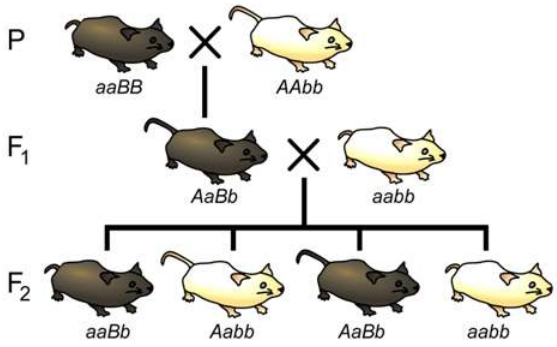


Figure 9.8.5.8: An experiment to measure recombination frequency between two loci. The loci affect coat color (B/b) and tail length (A/a). (Wikipedia-Modified Deyholos-CC:AN)

The recombination events that may be detected will occur during meiosis in the dihybrid individual. If the loci are completely or partially linked, then prior to meiosis, alleles *aB* will be located on one chromosome, and alleles *Ab* will be on the other chromosome (based on our knowledge of the genotypes of the gametes that produced the dihybrid). Thus, recombinant gametes produced by the dihybrid will have the genotypes *ab* or *AB*, and non-recombinant (i.e. parental) gametes will have the genotypes *aB* or *Ab*.

How do we determine the genotype of the gametes produced by the dihybrid individual? The most practical method is to use a testcross (Figure 9.8.5.8), in other words to mate *AaBb* to an individual that has only recessive alleles at both loci (*aabb*). This will give a different phenotype in the F_2 generation for each of the four possible combinations of alleles in the gametes of the dihybrid. We can then infer unambiguously the genotype of the gametes produced by the dihybrid individual, and therefore calculate the recombination frequency between these two loci. For example, if only two phenotypic classes were observed in the F_2 (i.e. short tails and brown fur (*aaBb*), and white fur with long tails (*Aabb*) we would know that the only gametes produced following meiosis of the dihybrid individual were of the parental type: *aB* and *Ab*, and the recombination frequency would therefore be 0%. Alternatively, we may observe multiple classes of phenotypes in the F_2 in ratios such as shown in Table 9.8.5.1:

Table 9.8.5.1: An example of quantitative data that may be observed in a genetic mapping experiment involving two loci. The data correspond to the F_2 generation in the cross shown in Figure 9.8.5.8.

tail phenotype	fur phenotype	number of progeny	gamete from dihybrid	genotype of F_2 from test cross	(P)arental or (R)ecombinant
short	brown	48	*aB*	*aaBb*	P
long	white	42	*Ab*	*Aabb*	P
short	white	13	*ab*	*aabb*	R
long	brown	17	*AB*	*AaBb*	R

Given the data in Table 9.8.5.1, the calculation of recombination frequency is straightforward:

$$\text{recombination frequency} = \frac{\text{number of recombinant gametes}}{\text{total number of gametes scored}} \tag{9.8.5.1}$$

$$\text{R.F.} = \frac{13+17}{48+42+13+17} \tag{9.8.5.2}$$

$$= 25\% \tag{9.8.5.3}$$

This page titled 9.8.5: Inferring Recombination From Genetic Data is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.5: Inferring Recombination From Genetic Data** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.8.6: Genetic Mapping

Because the frequency of recombination between two loci (up to 50%) is roughly proportional to the chromosomal distance between them, we can use recombination frequencies to produce genetic maps of all the loci along a chromosome and ultimately in the whole genome. The units of genetic distance are called **map units** (mu) or **centiMorgans** (cM), in honor of **Thomas Hunt Morgan** by his student, **Alfred Sturtevant**, who developed the concept. Geneticists routinely convert recombination frequencies into cM: the recombination frequency in percent is approximately the same as the map distance in cM. For example, if two loci have a recombination frequency of 25% they are said to be ~25cM apart on a chromosome (Figure 9.8.6.9). Note: this approximation works well for small distances (RF<30%) but progressively fails at longer distances because the RF reaches a maximum at 50%. Some chromosomes are >100 cM long but loci at the tips only have an RF of 50%. The method for mapping of these long chromosomes is shown below.

Note that the map distance of two loci alone does not tell us anything about the orientation of these loci relative to other features, such as centromeres or telomeres, on the chromosome.

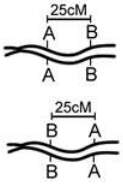


Figure 9.8.6.9: Two genetic maps consistent with a recombination frequency of 25% between A and B. Note the location of the centromere. (Original-Deyholos-CC:AN)

Map distances are always calculated for one pair of loci at a time. However, by combining the results of multiple pairwise calculations, a **genetic map** of many loci on a chromosome can be produced (Figure 9.8.6.10). A genetic map shows the map distance, in cM, that separates any two loci, and the position of these loci relative to all other mapped loci. The genetic map distance is roughly proportional to the physical distance, i.e. the amount of DNA between two loci. For example, in *Arabidopsis*, 1.0 cM corresponds to approximately 150,000bp and contains approximately 50 genes. The exact number of DNA bases in a cM depends on the organism, and on the particular position in the chromosome; some parts of chromosomes ("crossover hot spots") have higher rates of recombination than others, while other regions have reduced crossing over and often correspond to large regions of heterochromatin.

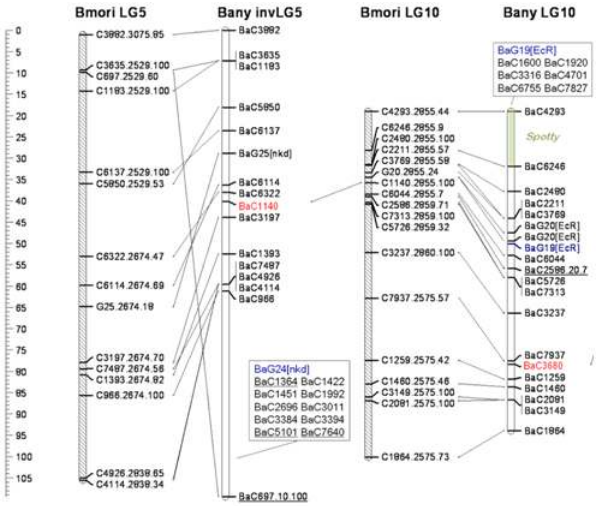


Figure 9.8.6.10: Genetic maps for regions of two chromosomes from two species of the moth, Bombyx. The scale at left shows distance in cM, and the position of various loci is indicated on each chromosome. Diagonal lines connecting loci on different chromosomes show the position of corresponding loci in different species. This is referred to as regions of **conserved synteny**. *(NCBI-NIH-PD)*

When a novel gene or locus is identified by mutation or polymorphism, its approximate position on a chromosome can be determined by crossing it with previously mapped genes, and then calculating the recombination frequency. If the novel gene and

the previously mapped genes show complete or partial linkage, the recombination frequency will indicate the approximate position of the novel gene within the genetic map. This information is useful in isolating (i.e. cloning) the specific fragment of DNA that encodes the novel gene, through a process called **map-based cloning**.

Genetic maps are also useful to track genes/alleles in breeding crops and animals, in studying evolutionary relationships between species, and in determining the causes and individual susceptibility of some human diseases.

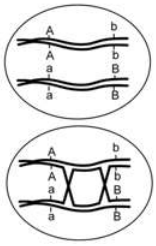


Figure 9.8.6.11: A double crossover between two loci will produce gametes with parental genotypes. (Original-Deyholos-CC:AN)

Genetic maps are useful for showing the order of loci along a chromosome, but the distances are only an approximation. The correlation between recombination frequency and actual chromosomal distance is more accurate for short distances (low RF values) than long distances. Observed recombination frequencies between two relatively distant markers tend to underestimate the actual number of crossovers that occurred. This is because as the distance between loci increases, so does the possibility of having a second (or more) crossovers occur between the loci. This is a problem for geneticists, because with respect to the loci being studied, these **double-crossovers** produce gametes with the same genotypes as if no recombination events had occurred (Figure 9.8.6.11) – they have parental genotypes. Thus a double crossover will appear to be a parental type and not be counted as a recombinant, despite having two (or more) crossovers. Geneticists will sometimes use specific mathematical formulae to adjust large recombination frequencies to account for the possibility of multiple crossovers and thus get a better estimate of the actual distance between two loci.

This page titled 9.8.6: Genetic Mapping is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.6: Genetic Mapping** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.8.7: Mapping With Three-Point Crosses

A particularly efficient method of mapping three genes at once is the **three-point cross**, which allows the order and distance between three potentially linked genes to be determined in a single cross experiment (Figure 9.8.7.12). This is particularly useful when mapping a new mutation with an unknown location to two previously mapped loci. The basic strategy is the same as for the dihybrid mapping experiment; pure breeding lines with contrasting genotypes are crossed to produce an individual heterozygous at three loci (a trihybrid), which is then testcrossed to determine the recombination frequency between each pair of genes.

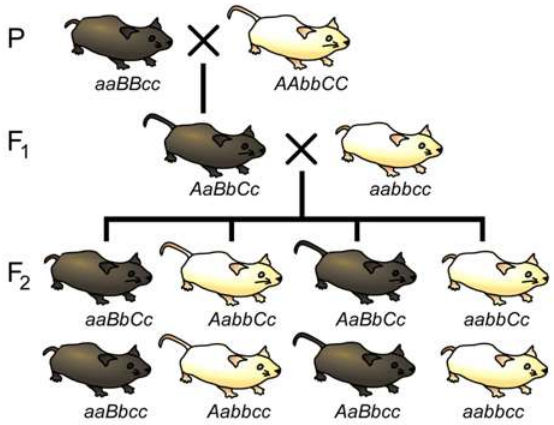


Figure 9.8.7.12: A three point cross for loci affecting tail length, fur color, and whisker length. (Original-Modified Deyholos-CC:AN)

One useful feature of the three-point cross is that the order of the loci relative to each other can usually be determined by a simple visual inspection of the F_2 segregation data. If the genes are linked, there will often be two phenotypic classes that are much more infrequent than any of the others. In these cases, the rare phenotypic classes are usually those that arose from two crossover events, in which the locus in the middle is flanked by a crossover on either side of it. Thus, among the two rarest recombinant phenotypic classes, the one allele that differs from the other two alleles relative to the parental genotypes likely represents the locus that is in the middle of the other two loci. For example, based on the phenotypes of the pure-breeding parents in Figure 9.8.7.12, the parental genotypes are *aBC* and *AbC* (remember the order of the loci is unknown, and it is not necessarily the alphabetical order in which we wrote the genotypes). Because we can deduce from the outcome of the testcross (Table 9.8.7.2) that the rarest genotypes were *abC* and *ABc*, we can conclude that locus *A* that is most likely located between the other two loci, since it would require a recombination event between both *A* and *B* and between *A* and *C* in order to generate these gametes. Thus, the order of loci is *BAC* (which is equivalent to *CAB*).

Table 9.8.7.2: An example of data that might be obtained from the F2 generation of the three-point cross shown in Figure 9.8.7.12. The rarest phenotypic classes correspond to double recombinant gametes ABc and abC. Each phenotypic class and the gamete from the trihybrid that produced it can also be classified as parental (P) or recombinant (R) with respect to each pair of loci (A,B), (A,C), (B,C) analyzed in the experiment.

tail phenotype	fur phenotype	whisker phenotype	number of progeny	gamete from trihybrid	genotype of F_2 from test cross	loci A, B	loci A, C	loci B, C
short	brown	long	5	*aBC*	*aaBbCc*	P	R	R
long	white	long	38	*AbC*	*AabbCc*	P	P	P
short	white	long	1	*abC*	*aabbCc*	R	R	P
long	brown	long	16	*ABC*	*AaBbCc*	R	P	R
short	brown	short	42	*aBc*	*aaBbcc*	P	P	P
long	white	short	5	*Abc*	*Aabbcc*	P	R	R
short	white	short	12	*abc*	*aabbcc*	R	P	R
long	brown	short	1	*ABc*	*AaBbcc*	R	R	P

Recombination frequencies may be calculated for each pair of loci in the three-point cross as we did before for one pair of loci in our dihybrid (Figure 7. 8).

$$\text{loci A,B R.F.} = \frac{1+16+12+1}{120} = 25\% \tag{9.8.7.1}$$

$$\text{loci A,C R.F.} = \frac{1+5+1+5}{120} = 10\% \tag{9.8.7.2}$$

$$\text{loci B,C R.F.} = \frac{5+16+12+5}{120} = 32\% \tag{9.8.7.3}$$

$$\text{(not corrected for double crossovers)} \tag{9.8.7.4}$$

$$\tag{9.8.7.5}$$

However, note that in the three point cross, the sum of the distances between A-B and A-C (35%) is less than the distance calculated for B-C (32%)(Figure 9.8.7.13). this is because of **double crossovers** between B and C, which were undetected when we considered only pairwise data for B and C. We can easily account for some of these double crossovers, and include them in calculating the map distance between B and C, as follows. We already deduced that the map order must be *BAC* (or *CAB*), based on the genotypes of the two rarest phenotypic classes in Table 9.8.7.2. However, these double recombinants, *ABc* and *abC*, were not included in our calculations of recombination frequency between loci *B* and *C*. If we included these double recombinant classes (multiplied by 2, since they each represent two recombination events), the calculation of recombination frequency between B and C is as follows, and the result is now more consistent with the sum of map distances between A-B and A-C.

$$\text{loci B,C R.F.} = \frac{5+16+12+5+2(1)+2(1)}{120} = 35\% \tag{9.8.7.6}$$

$$\text{(corrected for double recombinants)} \tag{9.8.7.7}$$

$$\tag{9.8.7.8}$$

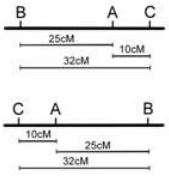


Figure 9.8.7.13: Equivalent maps based on the data in Table 9.8.7.2 (without correction for double crossovers). (Original-Deyholos-CC:AN)

Thus, the three point cross was useful for:

1. determining the order of three loci relative to each other,
2. calculating map distances between the loci, and
3. detecting some of the double crossover events that would otherwise lead to an underestimation of map distance.

However, it is possible that other, double crossovers events remain undetected, for example double crossovers between loci A,B or between loci A,C. Geneticists have developed a variety of mathematical procedures to try to correct for things like double crossovers during large-scale mapping experiments.

As more and more genes are mapped a better genetic map can be constructed. Then, when a new gene is discovered, it can be mapped relative to other genes of known location to determine its location. All that is needed to map a gene is two alleles, a wild type allele (e.g. A) and a mutant allele (e.g. 'a').

This page titled 9.8.7: Mapping With Three-Point Crosses is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.7: Mapping With Three-Point Crosses** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.8.E: Linkage and Mapping (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions

7.1 Compare recombination and crossover. How are these similar? How are they different?

7.2 Explain why it usually necessary to start with pure-breeding lines when measuring genetic linkage by the methods presented in this chapter.

7.3 If you knew that a locus that affected earlobe shape was tightly linked to a locus that affected susceptibility to cardiovascular disease human, under what circumstances would this information be clinically useful?

7.4 In a previous chapter, we said a 9:3:3:1 phenotypic ratio was expected among the progeny of a dihybrid cross, in absence of gene interaction.

a) What does this ratio assume about the linkage between the two loci in the dihybrid cross?

b) What ratio would be expected if the loci were completely linked? Be sure to consider every possible configuration of alleles in the dihybrids.

7.5 Given a dihybrid with the genotype *CcEe*:

a) If the alleles are in coupling (cis) configuration, what will be the genotypes of the parental and recombinant progeny from a test cross?

b) If the alleles are in repulsion (trans) configuration, what will be the genotypes of the parental and recombinant progeny from a test cross?

7.6 Imagine the white flowers are recessive to purple flowers, and yellow seeds are recessive to green seeds. If a green-seeded, purple-flowered dihybrid is testcrossed, and half of the progeny have yellow seeds, what can you conclude about linkage between these loci? What do you need to know about the parents of the dihybrid in this case?

7.7 In corn (i.e. maize, a diploid species), imagine that alleles for resistance to a particular pathogen are recessive and are linked to a locus that affects tassel length (short tassels are recessive to long tassels). Design a series of crosses to determine the map distance between these two loci. You can start with any genotypes you want, but be sure to specify the phenotypes of individuals at each stage of the process. Outline the crosses similar to what is shown in Figure 7.8, and specify which progeny will be considered recombinant. You do not need to calculate recombination frequency.

7.8 In a mutant screen in *Drosophila*, you identified a gene related to memory, as evidenced by the inability of recessive homozygotes to learn to associate a particular scent with the availability of food. Given another line of flies with an autosomal mutation that produces orange eyes, design a series of crosses to determine the map distance between these two loci. Outline the crosses similar to what is shown in Figure 7.8, and specify which progeny will be considered recombinant. You do not need to calculate recombination frequency.

7.9 Image that methionine heterotrophy, chlorosis (loss of chlorophyll), and absence of leaf hairs (trichomes) are each caused by recessive mutations at three different loci in Arabidopsis. Given a triple mutant, and assuming the loci are on the same chromosome, explain how you would determine the order of the loci relative to each other.

7.10 If the progeny of the cross *aaBB* x *AAbb* is testcrossed, and the following genotypes are observed among the progeny of the testcross, what is the frequency of recombination between these loci?

AaBb 135

Aabb 430

aaBb 390

aabb 120

7.11 Three loci are linked in the order B-C-A. If the A-B map distance is 1cM, and the B-C map distance is 0.6cM, given the lines *AaBbCc* and *aabbcc*, what will be the frequency of *Aabb* genotypes among their progeny if one of the parents of the dihybrid had the genotypes *AABBCC*?

7.12 Genes for body color (B black dominant to b yellow) and wing shape (C straight dominant to c curved) are located on the same chromosome in flies. If single mutants for each of these traits are crossed (i.e. a yellow fly crossed to a curved-wing fly), and their progeny is testcrossed, the following phenotypic ratios are observed among their progeny.

	black, straight	17
	yellow, curved	12
	black, curved	337
	yellow, straight	364

a) Calculate the map distance between B and C.

b) Why are the frequencies of the two smallest classes not exactly the same?

7.13 Given the map distance you calculated between B-C in question 12, if you crossed a double mutant (i.e. yellow body and curved wing) with a wild-type fly, and testcrossed the progeny, what phenotypes in what proportions would you expect to observe among the F_2 generation?

7.14 In a three-point cross, individuals *AAbbcc* and *aaBBCC* are crossed, and their F_1 progeny is testcrossed. Answer the following questions based on these F_2 frequency data.

aaBbCc	480
AaBbcc	15
AaBbCc	10
aaBbcc	1
aabbCc	13
Aabbcc	472
AabbCc	1
aabbcc	8

a) Without calculating recombination frequencies, determine the relative order of these genes.

b) Calculate pair-wise recombination frequencies (without considering double cross overs) and produce a genetic map.

c) Recalculate recombination frequencies accounting for double recombinants.

7.15 Wild-type mice have brown fur and short tails. Loss of function of a particular gene produces white fur, while loss of function of another gene produces long tails, and loss of function at a third locus produces agitated behaviour. Each of these loss of function alleles is recessive. If a wild-type mouse is crossed with a triple mutant, and their F_1 progeny is test-crossed, the following recombination frequencies are observed among their progeny. Produce a genetic map for these loci.

Chapter 7 - Answers

7.1

Crossovers are defined cytologically; they are observed directly under the microscope.

Recombination is defined genetically; it is calculated from observed phenotypic proportions.

Some crossovers lead to recombination, but not all crossovers result in recombination.

Some recombinations involve crossovers, but not all recombinations result from crossovers.

Crossovers happen between sister and non-sister chromatids. If the chromatids involved the crossover have identical alleles, there will not be any recombination.

Crossovers can also happen without causing recombination when there are two crossovers between the loci being scored for recombination.

Recombination can occur without crossover when loci are on different chromosomes.

7.2

The use of pure breeding lines allows the researcher to be sure that he/she is working with homozygous genotypes. If a parent is known to be homozygous, then all of its gametes will have the same genotype. This simplifies the definition of parental genotypes and therefore the calculation of recombination frequencies.

7.3

This would suggest that individuals with a particular earlobe phenotype may also carry one or more alleles that increased their risk of cardiovascular disease. These individuals could therefore be informed of their increased risk and have an opportunity to seek increased monitoring and reduce other risk factors.

7.4 a)

It assumes that the loci are completely unlinked.

b)

If the parental gametes were *AB* and *ab*, then the gametes produced by the dihybrids would also be *AB* and *ab*, and the offspring of a cross between the two dihybrids would all be genotype *AABB*:*AaBb*:*aabb*,in a 1:2:1 ratio.

If the parental gametes were *Ab* and *aB*, then the gametes produced by the dihybrids would also be *Ab* and *aB*, and the offspring of a cross between the two dihybrids would all be genotype *AAbb*:*AaBb*:*aaBB*, in a 1:2:1 ratio.

fur	tail	behaviour	
white	short	normal	16
brown	short	agitated	0
brown	short	normal	955
white	short	agitated	36
white	long	normal	0
brown	long	agitated	14
brown	long	normal	46
white	long	agitated	933

7.5 a) Parental: *CcEe* and *ccee;* Recombinant: *Ccee* and *ccEe*.

b) Parental: *Ccee* and *ccEe;* Recombinant: *CcEe* and *ccee*.

7.6 Let *WwYy* be the genotype of a purple-flowered (*W*), green seeded (*Y*) dihybrid . Half of the progeny of the cross *WwYy* × *wwyy* will have yellow seeds whether the loci are linked or not. The proportion of the seeds that are also either white or purple flowered would help you to know about the linkage between the two loci only if the genotypes of the parents of the dihybrid were also known.

7.7

Let *tt* be the genotype of a short tassels, and *rr* is the genotype of pathogen resistant plants. We need to start with homozygous lines with contrasting combinations of alleles, for example:

P: *RRtt* (pathogen sensitive, short tassels) × *rrTT* (pathogen resistant, long tassels)

F_1: *RrTt* (sensitive, long) × *rrtt* (resistant, short)

F_2: parental *Rrtt* (sensitive, short) , *rrTt* (resistant, long)

Recombinant *rrtt* (resistant, short) , *RrTt* (sensitive, long)

7.8 Let *mm* be the genotype of a mutants that fail to learn, and *ee* is the genotype of orange eyes. We need to start with homozygous lines with contrasting combinations of alleles, for example (wt means wild-type):

P: *MMEE* (wt eyes, wt learning) × *mmee* (orange eyes, failure to learn)

F_1: *MmEe* (wt eyes, wt learning) × *mmee* (orange eyes, failure to learn)

F_2: parental *MmEe* (wt eyes, wt learning) , *mmee* (orange eyes, failure to learn)

Recombinant *Mmee* (wt eyes, failure to learn) , *mmEe* (orange eyes, wt learning)

7.9 Given a triple mutant *aabbcc* , cross this to a homozygote with contrasting genotypes, i.e. *AABBCC*, then testcross the trihybrid progeny, i.e.

P: *AABBCC* × *aabbcc*

F_1: *AaBbCc* × *aabbcc*

Then, in the F_2 progeny, find the two rarest phenotypic classes; these should have reciprocal genotypes, e.g. *aaBbCc* and *AAbbcc*. Find out which of the three possible orders of loci (i.e. *A-B-C*, *B-A-C*, or *B-C-A*) would, following a double crossover that flanked the middle marker, produce gametes that correspond to the two rarest phenotypic classes. For example, if the rarest phenotypic classes were produced by genotypes *aaBbCc* and *AAbbcc*, then the dihybrid's contribution to these genotypes was *aBC* and *Abc*. Since the parental gametes were *ABC* and *abc* the only gene order that is consistent with *aBC* and *Abc* being produced by a double crossover flanking a middle marker is *B-A-C* (which is equivalent to *C-A-B*).

7.10 If the progeny of the cross *aaBB* x *AAbb* is testcrossed, and the following genotypes are observed among the progeny of the testcross, what is the frequency of recombination between these loci?

AaBb 135

Aabb 430

aaBb 390

aabb 120

(135 + 120)/(135+120+390+430)= 24%

7.11

Based on the information given, the recombinant genotypes with respect to these loci will be *Aabb* and *aaBb*. The frequency of recombination between A-B is 1cM=1%, based on the information given in the question, so each of the two recombinant genotypes should be present at a frequency of about 0.5%. Thus, the answer is 0.5%.

7.12

a) 4cM

b) Random sampling effects; the same reason that many human families do not have an equal number of boys and girls.

7.13

There would be approximately 2% of each of the recombinants: (yellow, straight) and (black, curved), and approximately 48% of each of the parentals: (yellow, curved) and (black, straight).

7.14

a) Without calculating recombination frequencies, determine the relative order of these genes.

A-C-B

b)

A-B 4.6%

A-C 2%

B-C 3%

B C A

|----------|---------|

3cM 2cM

	A-B	A-C	B-C
aBC	0	0	0
ABc	15	0	15
ABC	10	10	0
aBc	0	1	1
abC	13	0	13
Abc	0	0	0
AbC	0	1	1
abc	8	8	0
TOTAL	46	20	30
%	4.6	2	3

c) Recalculate recombination frequencies accounting for double recombinants

	A-B	A-C	B-C
aBC	0	0	0
ABc	15	0	15
ABC	10	10	0
aBc	1 x 2	1	1
abC	13	0	13
Abc	0	0	0
AbC	1 x 2	1	1
abc	8	8	0
TOTAL	50	20	30
%	5	2	3

7.15

A is fur color locus

B is tail length locus

C is behaviour locus

fur (A)	tail (B)	behaviour (C)			AB	AC	BC
white	short	normal	16	*aBC*	R	R	P
brown	short	agitated	0	*ABc*	P	R	R
brown	short	normal	955	*ABC*	P	P	P
white	short	agitated	36	*aBc*	R	P	R
white	long	normal	0	*abC*	P	R	R
brown	long	agitated	14	*Abc*	R	R	P
brown	long	normal	46	*AbC*	R	P	R
white	long	agitated	933	*abc*	P	P	P

B C A

|----------|---------|

4.1cM 1.5cM

Pairwise recombination frequencies are as follows (calculations are shown below):

A-B 5.6%

A-C 1.5%

B-C 4.1%

AB	AC	BC
16	16	0
0	0	0
0	0	0
36	0	36
0	0	0
14	14	0
46	0	46
0	0	0
112	30	82
5.6%	1.5%	4.1%

This page titled 9.8.E: Linkage and Mapping (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

9.8.S: Linkage and Mapping (Summary)

- Recombination is defined as any process that results in gametes with combinations of alleles that were not present in the gametes of a previous generation.
- The recombination frequency between any two loci depends on their relative chromosomal locations.
- Unlinked loci show a maximum 50% recombination frequency.
- Loci that are close together on a chromosome are linked and tend to segregate with the same combinations of alleles that were present in their parents.
- Crossovers are a normal part of most meioses, and allow for recombination between linked loci.
- Measuring recombination frequency is easiest when starting with pure-breeding lines with two alleles for each locus, and with suitable lines for test crossing.
- Because recombination frequency is usually proportional to the distance between loci, recombination frequencies can be used to create genetic maps.
- Recombination frequencies tend to underestimate map distances, especially over long distances, since double crossovers may be genetically indistinguishable from non-recombinants.
- Three-point crosses can be used to determine the order and map distance between of three loci, and can correct for some of the double crossovers between the two outer loci.

Key Terms

linkage recombination independent assortment crossover recombinant genotype parental genotype unlinked recombination frequency (RF)	complete (absolute) linkage incomplete (partial) linkage syntenic synapsis coupling (cis) configuration repulsion(trans) configuration	repulsion map units (mu) centiMorgans (cM) genetic map conserved synteny double-crossover three-point cross

This page titled 9.8.S: Linkage and Mapping (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **7.S: Linkage and Mapping (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

9.E: Mutation and Variation (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Q4.1

How are polymorphisms and mutations alike? How are they different?

Q4.2

What are some of the ways a substitution can occur in a DNA sequence?

Q4.3

What are some of the ways a deletion can occur in a DNA sequence?

Q4.4

What are all of the ways an insertion can occur in a DNA sequence?

Q4.5

In the context of this chapter, explain the health hazards of smoking tobacco.

Q4.6

You have a female fruit fly, whose father was exposed to a mutagen (she, herself, wasn't). Mating this female fly with another non-mutagenized, wild type male produces offspring that all appear to be completely normal, except there are twice as many daughters as sons in the F_1 progeny of this cross.

a. Propose a hypothesis to explain these observations.
b. How could you test your hypothesis?

Q4.7

You decide to use genetics to investigate how your favourite plant makes its flowers smell good.

a. What steps will you take to identify some genes that are required for production of the sweet floral scent? Assume that this plant is a self-pollinating diploid.
b. One of the recessive mutants you identified has fishy-smelling flowers, so you name the mutant (and the mutated gene) *fishy*. What do you hypothesize about the normal function of the wild-type *fishy* gene?
c. Another recessive mutant lacks floral scent altogether, so you call it *nosmell*. What could you hypothesize about the normal function of this gene?

Q4.8

Suppose you are only interested in finding dominant mutations that affect floral scent.

a. What do you expect to be the relative frequency of dominant mutations, as compared to recessive mutations, and why?
b. How will you design your screen differently than in the previous question, in order to detect dominant mutations specifically?
c. Which kind of mutagen is most likely to produce dominant mutations, a mutagen that produces point mutations, or a mutagen that produces large deletions?

Q4.9

hich types of transposable elements are transcribed?

Q4.10

You are interested in finding genes involved in synthesis of proline (Pro), an amino acid that is normally synthesizes by a particular model organism.

a. How would you design a mutant screen to identify genes required for Pro synthesis?
b. Imagine that your screen identified ten mutants (#1 through #10) that grew poorly unless supplemented with Pro. How could you determine the number of different genes represented by these mutants?
c. If each of the four mutants represents a different gene, what will be the phenotype of the F1 progeny if any pair of the four mutants are crossed?
d. If each of the four mutants represents the same gene, what will be the phenotype of the F1 progeny if any pair of the four mutants are crossed?

Chapter 4 - Answers

4.1 Polymorphisms and mutations are both variations in DNA sequence and can arise through the same mechanisms. We use the term polymorphism to refer to DNA variants that are relatively common in populations. Mutations affect the phenotype.

4.2 Misreading of bases during replication can lead to substitution and can be caused by things like tautomerism, DNA alkylating agents, and irradiation.

4.3 Looping out of DNA on the template strand during replication; strand breakage, due to radiation and other mutagens; and (discussed in earlier chapters) chromosomal aberrations such as deletions and translocations.

4.4 Looping out of DNA on the growing strand during replication; transposition; and (discussed in earlier chapters) chromosomal aberrations such as duplications, insertions, and translocation.

4.5 Benzopyrene is one of many hazardous compounds present in smoke. Benzopyrene is an intercalating agent, which slides between the bases of the DNA molecule, distorting the shape of the double helix, which disrupts transcription and replication and can lead to mutation.

4.6 a) One possible explanation is that original mutagenesis resulted in a loss-of-function mutation in a gene that is essential for early embryonic development, and that this mutation is X-linked recessive in the female. Because half of the sons will inherit the X chromosome that bears this mutation, half of the sons will fail to develop beyond very early development and will not be detected among the F_1 progeny. The proportion of male flies that were affected depends on what fraction of the female parent's gametes carried the mutation. In this case, it appears that half of the female's gametes carried the mutation.

b) To test whether a gene is X-linked, you can usually do a reciprocal cross. However, in this case it would be impossible to obtain adult male flies that carry the mutation; they are dead. If the hypothesis proposed in a) above is correct, then half of the females, and none of the living males in the F_1 should carry the mutant allele. You could therefore cross F_1 females to wild type males, and see whether the expected ratios were observed among the offspring (e.g. half of the F_1 females should have a fewer male offspring than expected, while the other half of the F_1 females and all of the males should have a roughly equal numbers of male and female offspring).

4.7 a) Treat a population of seeds with a mutagen such as EMS. Allow these seeds to self-pollinate, and then allow the F_1 generation to also self-pollinate. In the F_2 generation, smell each flower to find individuals with abnormal scent.

b) The fishy gene appears to be required to make the normal floral scent. Because the flowers smell fishy in the absence of this gene, one possibility explanation of this is that fishy makes an enzyme that converts a fishy-smelling intermediate into a chemical that gives flowers their normal, sweet smell.

Note that although we show this biochemical pathway as leading from the fishy-smelling chemical to the sweet-smelling chemical in one step, it is likely that there are many other enzymes that act after the fishy enzyme to make the final, sweet-smelling product. In either case, blocking the pathway at the step catalyzed by the *fishy* enzyme would explain the fishy smell.

c) In *nosmell* plants, the normal sweet smell disappears. Unlike *fishy*, the sweet smell is not replaced by any intermediate chemical that we can easily detect. Thus, we cannot conclude where in the biochemical pathway the *nosmell* mutant is blocked; *nosmell* may normally therefore act either before or after *fishy* normally acts in the pathway:

Alternatively, *nosmell* may not be part of the biosynthetic pathway for the sweet smelling chemical at all. It is possible that the normal function of this gene is to transport the sweet-smelling chemical into the cells from which it is released into the air, or maybe it is required for the development of those cells in the first place. It could even be something as general as keeping the plants healthy enough that they have enough energy to do things like produce floral scent.

4.8 a) Dominant mutations are generally much rarer than recessive mutations. This is because mutation of a gene tends to cause a loss of the normal function of this gene. In most cases, having just one normal (wt) allele is sufficient for normal biological function, so the mutant allele is recessive to the wt allele. Very rarely, rather than destroying normal gene function, the random act of mutation will cause a gene to gain a new function (e.g. to catalyze a new enzymatic reaction), which can be dominant (since it performs this new function whether the wt allele is present or not). This type of gain-of-function dominant mutation is very rare because there are many more ways to randomly destroy something than by random action to give it a new function (think of the example given in class of stomping on an iPod).

b) Dominant mutations should be detectable in the F_1 generation, so the F_1 generation, rather than the F_2 generation can be screened for the phenotype of interest.

c) Large deletions, such as those caused by some types of radiation, are generally less likely than point mutations to introduce a new function into a protein: it is hard for a protein to gain a new function if the entire gene has been removed from the genome by deletion.

4.9 Class I. see Figure 4.5 on Transposable Elements.

4.10 a) Mutagenize a wild type (auxotrophic) strain and screen for mutations that fail to grow on minimal media, but grow well on minimal media supplemented with proline.
b) Take mutants #1-#10) and characterize them, based on (1) genetic mapping of the mutants (different locations indicate different genes); (2) different response to proline precursors (a different response suggests different genes); (3) complementation tests among the mutations (if they complement then they are mutations in different genes).
c) If the mutations are in different genes then the F1 progeny would be wild type (able to grow on minimal medium without proline).
d) If the mutations are in the same gene then the F1 progeny would NOT be wild type (unable to grow on minimal medium without proline).

This page titled 9.E: Mutation and Variation (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

9.S: Mutation and Variation (Summary)

- When a variation in DNA sequence originated recently, and is rare in a population, we call that change a mutation.
- When variations in DNA sequence co-exist in a population, and neither one can be meaningfully defined as wild-type, we call the variations polymorphisms.
- Mutations may either occur spontaneously, or may be induced by exposure to mutagens.
- Mutations may result in either substitutions, deletions, or insertions.
- Mutation usually causes either a partial or complete loss of function, but sometimes results in a gain of function,including new functions.
- Spontaneous mutations arise from many sources including natural errors in DNA replication, usually associated with base mispairing, or else insertion deletion especially within repetitive sequences.
- Induced mutations result from mispairing, DNA damage, or sequence interruptions caused by chemical, biological, or physical mutagens.
- By randomly inducing mutations, then screening for a specific phenotype, it is possible to identify genes associated with specific biological pathways.
- Transposable elements are dynamic, abundant components of eukaryotic genomes and important forces in evolution.
- Transposable elements are dynamic, abundant components of eukaryotic genomes and important forces in evolution.
- The efficiency of mutant screening is limited by silent mutations, redundancy, and embyronic lethality.
- Mutation of different genes can produce a similar phenotype.
- Complementation testing determines whether two mutants are the result of mutation of the same gene (allelic mutations), or if each mutant is caused by mutation of a different gene (non-allelic mutations).

Key Terms

mutation		
mutant	P-element	
polymorphism	T-DNA	
insertion	copy-and-paste	
deletion	cut-and-paste	
substitution	alkylation agent	
mutagen	EMS	
DNA replication error	intercalating agent	
Strand slippage	benzopyrene	
biological mutagen	carcinogenic	
chemical mutagen	ethidium bromide	
physical mutagen	thymine dimer	
mispairing	mutant screen	
loop	loss-of-function	
SSR	gain-of-function	
insertional mutagen	amorph	
Class I, Class II	null	
transposon	hypomorph	
retrotransposon	hypermorph	
reverse transcriptase	neopmorph	
transposase	dominant negative	
non-autonomous	somatic cells	
autonomous	germline cells	
SINE, LINE, Alu		
		silent mutation
		inter-genic region
		redundancy
		essential gene
		recessive lethal allele
		complementation testing
		allelic / non-allelic
		cM_0, M_1, M_2
		redundancy
		lethality
		allelic
		non-allelic
		complementation group
		CFTR
		Cystic Fibrosis (CF)
		DF508(PHE508DEL)
		Kalydeco

This page titled 9.S: Mutation and Variation (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **4.S: Mutation and Variation (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

10: Changes in Chromosome Number and Structure

Previous chapters described chromosomes as simple linear DNA molecules on which genes are located. For example, your largest chromosome, chromosome 1, has about 3536 genes. To ensure that each of your cells possesses these genes the chromosome has features that allow it to be passed on during cell division. **Origins of replication** found along its length provide places for DNA replication to start, **telomeres** protect each end of the chromosome, and a single **centromere** near the middle provides a place for microtubules to attach and move the chromosome during mitosis and meiosis. This chapter examines: (1) changes in the number of whole chromosomes and how they affect the phenotype of an organism and (2) changes in the structure of individual chromosomes and how they affect meiotic pairing. Human examples will be used to show the phenotypic consequences and methods for detection.

10.1: Changes in Chromosome Number

10.2: Changes in Chromosome Structure

10.3: Chromosome Abnormalities in Humans

10.4: Diagnosing Human Chromosome Abnormalities

10.E: Changes in Chromosome Number and Structure (Exercises)

10.S: Changes in Chromosome Number and Structure (Summary)

Thumbnail: (Wikipedia-Pmx-CC:AS)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 10: Changes in Chromosome Number and Structure is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Changes in Chromosome Number

If something goes wrong during cell division, an entire chromosome may be lost and the cell will lack all of these genes. The causes behind these chromosome abnormalites and the consequences they have for the cell and the organism is the subject of this section.

Cause: Nondisjunction During Mitosis or Meiosis

Segregation occurs in anaphase. In mitosis and meiosis II, sister chromatids (of replicated chromosomes) are normally pulled to opposite ends of the cell (see Figure 10.1.1a). In Meiosis I, it is homologous chromosomes, which are synapsed at that time, that segregate and move apart.

In rare cases, the sister chromatids (or paired chromosomes in Meiosis I) fail to separate, or dis-join. This failure to segregate properly is called **nondisjunction** and it can happen during mitosis, meiosis I, or meiosis II. This nondisjunction results in both chromatids (or chromosomes) moving to one pole and none at the other. One cell will have an extra copy and the other will lack a copy. Thus failure to segregate properly leads to unbalanced products.

Consequence: Decreased Viability

The result of a non-disjunction event is daughter cells that have an abnormal number of chromosomes. Cells, such as the parent cell in Figure 10.1.1a, which have the proper number of chromosomes, are said to be **euploid**. The daughter cells have one too many or one too few chromosomes and are thus **aneuploid**. Even though both product cells have at least one copy of all genes, both cells will probably die. The reason is due to the loss or gain of a large number of genes. Genes produce an standard amount of product - either functional RNAs or proteins. The parent cell shown has a **balanced** genotype because it has two copies of all of its genes. Each of its genes produces half of the products needed by the cell. But if one of these cells suddenly had only one copy (or three copies) of an important gene, the amount of product would be either 50% (or 150%) of what was required. Such a change for a single gene could probably be tolerated by the cell and it would probably survive. But the sudden change to one copy (or three copies) of the hundreds or thousands of genes on an entire chromosome the results would be more than tolerable and be catastrophic for the daughter cells. They have what's called an **unbalanced** genotype, which usually decreases their viability.

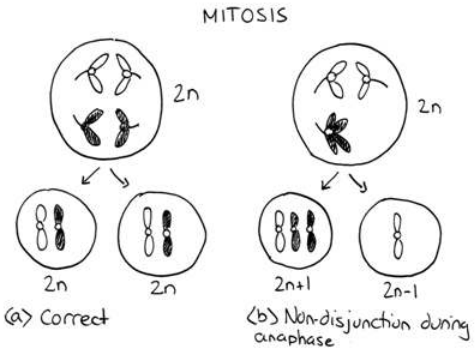


Figure 10.1.1: Mitosis done successfully (a) and unsuccessfully (b). The cell is diploid and the homologs of one chromosome are shown in white and black. (Original-Harrington-CC:AN)

If a **first division** or **second division nondisjunction** event occurs during meiosis the result is an unbalanced gamete (Figure 10.1.2b and 10.1.2c). The gamete will often be functional, but after fertilization the embryo will be genetically unbalanced. This usually leads to the death of the embryo. There are some exceptions to this in humans and these will be presented later in this chapter.

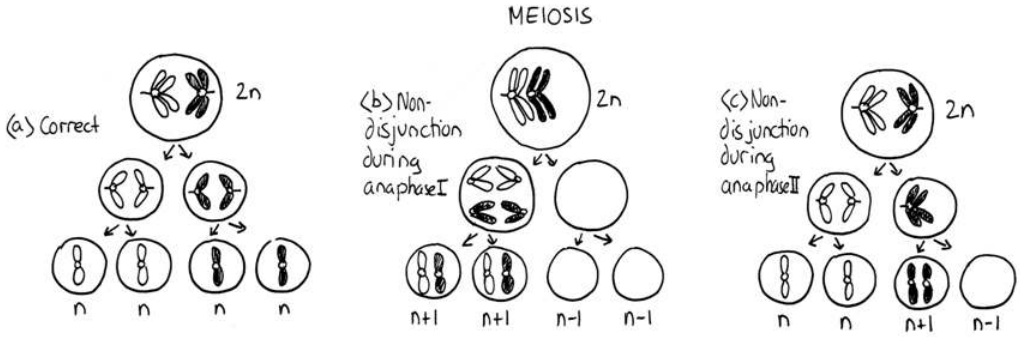


Figure 10.1.2: Meiosis done successfully (a) and unsuccessfully (b and c). (Original-Harrington-CC:AN)

This page titled 10.1: Changes in Chromosome Number is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **9.1: Changes in Chromosome Number** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

10.2: Changes in Chromosome Structure

If the chromosome is altered, but still retains the three critical features of a chromosome (centromeres, telomeres, and origin of replication), it will continue to be inherited during subsequent cell divisions, however the daughter cell may not retain all the genes. For example, if a segment of the chromosome has been lost, the cell may be missing some genes. The causes of chromosome structural abnormalites and the consequences they have for the cell and the organism is shown below. They all involve breaks in the DNA that makes up the chromosome.

Cause #1: Incorrect Repair of Double Strand DNA Breaks During Interphase

A chromosome is a very long but very thin molecule. In the phopho-diester backbone there are only two covalent bonds holding each base pair to the next. If one of these covalent bonds is broken the chromosome will still remain intact, although a DNA Ligase will be needed to repair the nick (Figure 10.2.1*a*). Problems arise when both strands are broken at or near the same location. This **double strand break** will cleave the chromosome into two independent pieces (Figure 10.2.1*b*). Because these events do occur in cells there is a repair system called the **non-homologous end joining** (NHEJ) **system** to fix them. Proteins bind to each broken end of the DNA and reattach them with new covalent bonds. This system is not perfect and sometimes leads to **chromosome rearrangements** (see next section).

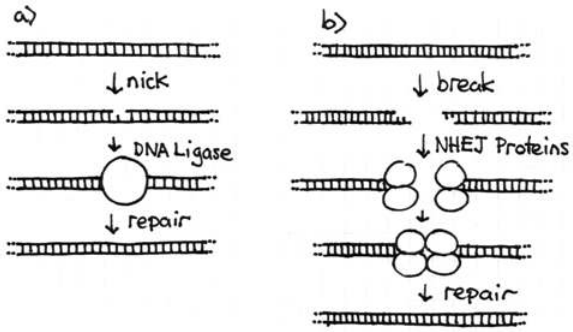


Figure 10.2.1: Repair of single strand nicks and double strand breaks in DNA. (Original-Harrington-CC:AN)

The NHEJ system proteins only function if required. If the chromosomes within an interphase nucleus are all intact the system is not active. The telomeres at the natural ends of chromosomes prevent the NHEJ system from attempting to join the normal ends of chromosomes together. If there is one double strand break the two broken ends can be recognized and joined. But if there are two double strand breaks at the same time there will be four broken ends in total. The NHEJ system proteins may join the ends together correctly but if they do not the result is a **chromosome rearrangement** (Figure 10.2.2).

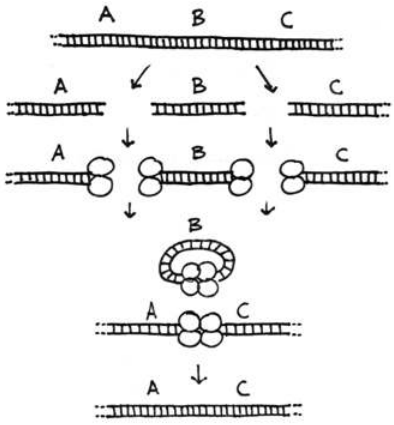


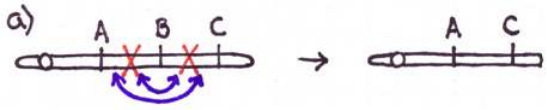
Figure 10.2.2 there has been breaks in the DNA, recruitment of NHEJ proteins, and repair. After the repairs are completed the small piece of DNA with gene B is lost and the chromosome now only has genes A and C. (Original-Harrington-CC:AN)

The Four Types of Chromosome rearrangements

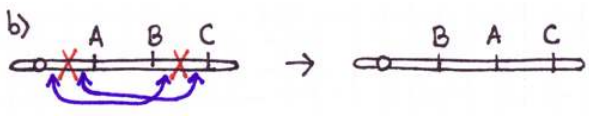
Errors during the repair of multiple double strand breaks can cause four types of chromosome rearrangements. The type of chromosome rearrangement is dependent upon where the two breaks were originally and how they are rejoined. Figure 10.2.5 shows some possibilities but more are shown below. In these there is a double strand DNA break between the B and C genes (shown here as a red X). A second DNA break occurs and the NHEJ proteins mend the damage incorrectly by joining the ends shown with the blue arrows. The chromosomes are drawn unreplicated as they are in G_1 phase but these events can happen anytime during interphase.

There are four major types of rearrangements:

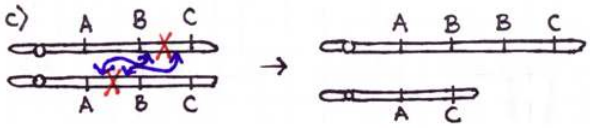
a) Deletions arise when both breaks are on one chromosome. If the ends are joined in this way the piece of DNA with the B gene on it does not have a centromere and will be lost during the next cell division.



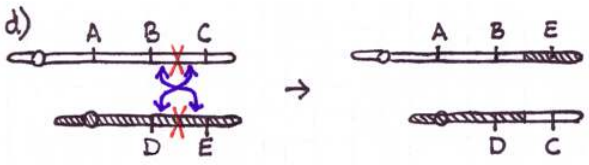
b) Inversions also occur when both breaks are on one chromosome. If the ends are joined in this way, part of the chromosome is inverted. This example shows a **paracentric inversion**, named because the inverted section does not include the centromere (para = beside). If the breaks occur on different chromosome arms the inverted section includes the centromere and the result is a **pericentric inversion** (peri = around).



c) Duplications can occur from two DNA breaks at different places in sister chromatids (in a replicated chromosome). The ends are joined together incorrectly to give a chromosome with a duplication (two “B” regions as shown above). Note: the reciprocal product has a deletion.



d) Translocations result from two breaks on different chromosomes (not homologs) and incorrect rejoining. This example shows a **reciprocal translocation** - two chromosomes have 'swapped' arms, the E gene is now part of the white chromosome and the C gene is now part of the shaded chromosome. **Robertsonian translocations** are those rare situations in which all of the genes end up together on one chromosome and the other chromosome is so small that it is typically lost.



Cause #2: Incorrect Crossovers During Meiosis

Meiotic crossovers occur at the beginning of meiosis for two reasons. They help hold the homologous chromosomes together until separation occurs during anaphase I (see Chapter 2). They also allow recombination to occur between linked genes (see Chapter 7). The event itself takes place during prophase I when a double strand break on one piece of DNA is joined with a double strand break on another piece of DNA and the ends are put together (Figure 10.2.3*a*). Most of the time the breaks are on non-sister chromatids and most of the time the breaks are at the same relative locations.

Problems occur when the wrong pieces of DNA are matched up along the chromosomes during crossover events. This can happen if the same or similar DNA sequence is found at multiple sites on the chromosomes (Figure 10.2.3*b*). For example, if there are two

***Alu* transposable elements** on a chromosome. When the homologous chromosomes pair during prophase I the wrong *Alu* sequences might line up. A crossover may occur in this region. If so, when the chromosomes separate during anaphase I one of the chromatids will have a duplication and one will have a deletion. Ultimately, of the four cells produced by this meiosis, two will be normal, one will have a chromosome with extra genes, and one will have a chromosome missing some genes. Errors of this type can also cause inversions and translocations.

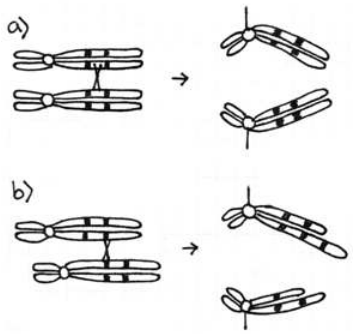


Figure 10.2.3: Errors during meiotic crossovers can cause duplications and deletions. This diagram shows homologous chromosomes pairing in prophase I and then separating in anaphase I. The shaded boxes are alu transposable elements. a) The homologous chromosomes pair properly, a crossover occurs, and all four chromatids in anaphase I are normal. b) The pairing is incorrect, a crossover occurs in the misspaired region, and in anaphase I one chromatid has a duplication and another has a deletion. (Original-Harrington-CC:AN)

Consequence #1 - Rearrangements Show Abnormal pairing at Meiosis

Homologous regions of chromosomes pair at meiosis I (prophase I). With rearranged chromosomes this can lead to visible abnormalities and segregation abnormalities.

Deletion chromosomes will pair up with a normal homolog along the shared regions and at the missing segment, the normal homolog will loop out (nothing to pair with) to form a **deletion loop**. This can be used to locate the deletion cytologically. The deleted region is also **pseudo-dominant**, in that it permits the mutant expression of recessive alleles on the normal homolog. Deletion mutations don't revert - nothing to replace the missing DNA.

When an **inversion** chromosome is paired up in meiosis there is an **inversion loop** formed. If there is a crossover within the loop then abnormal products will result and abnormal, unbalanced gametes will be produced. For example, a crossover event within the loop of a **paracentric inversion** will lead to a di-centric product that will break into deletion products and produce unbalanced gametes (Figure 10.2.4). Similarly, with a **pericentric inversion**, a crossover event leads to duplicate/deletion products that are unbalanced (Figure 10.2.5).

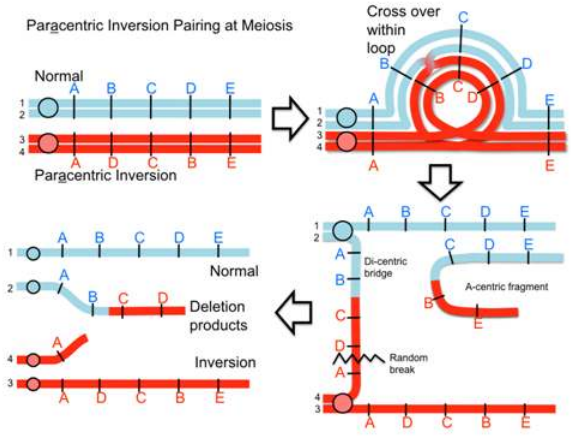


Figure 10.2.4: A paracentric inversion pairing at meiosis. A crossover within the loop causes the production of an acentric and a dicentric chromatids which leads to deletion product.(Original-Locke-CC:AN)

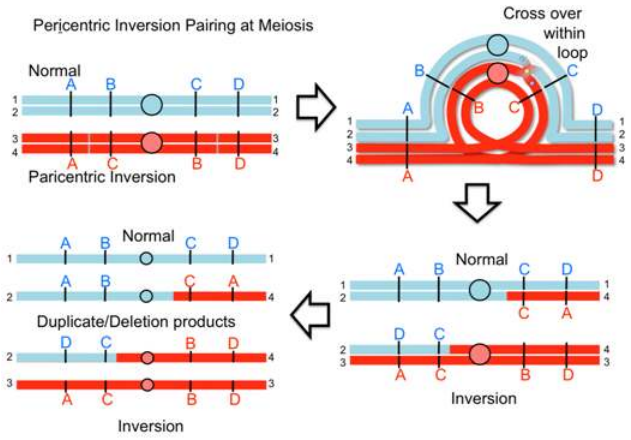


Figure 10.2.5: A pericentric inversion pairing at meiosis. A crossover within the loop causes the production of duplicate and deletion products. (Original-Locke-CC:AN)

If joined with a normal gamete, they will result in an unbalanced zygote, which are usually lethal. The consequence for this is that crossover products (recombinants) are lost and thus inversions appear to suppress crossovers within the inverted region.

> With both types of inversions, crossovers outside the loop are possible and fully viable as they don't alter the gene balance.

Duplications also produce a cytologically visible loop at meiotic pairing. Duplications can revert at a relatively high frequency by unequal crossing over. Duplicated genes offer new possibilities for mutational divergence followed by natural selection in the course of evolution.

For **translocations**, a consequence for the two chromosomes involved is that when they pair at meiosis both replicated chromosome pairs will be together, which can be seen cytologically as a **tetrad**. This tetrad can segregate in three ways. Two of which are shown below. This set of paired, replicated chromosomes can segregate as **Alternate** (balanced) where both normal and both translocated chromosomes go to the same polls. Or the chromosomes can segregate as **Adjacent-1** (unbalanced) where the normal and translocation chromosomes segregate as shown below. Each of these possibilities is approximately equally frequent and thus only about half the time do the gametes end up unbalanced (Figure 10.2.6).

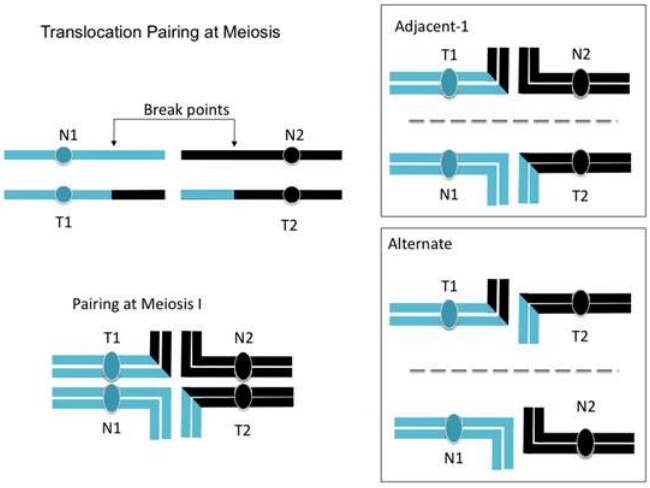


Figure 10.2.6: A reciprocal translocation pairing at meiosis. There are two main avenues for segregation: Adjacent-1 and Alternate. Adjacent-1 results in duplication and deletion for part of the chromosome segments. Alternate doesn't. (Original-Locke-CC:AN)

Consequence #2: Decreased Viability

All of the chromosome rearrangements shown above produce functional chromosomes. Each has one centromere, two telomeres, and thousands of origins of replication. Because inversions and translocations do not change the number of genes in a cell or

organism they are said to be balanced rearrangements. Unless one of the breakpoints occurred in the middle of a gene the cells will not be affected. On the other hand, deletions and duplications are unbalanced rearrangements. The larger they are (more genes involved) the more disruption they cause to the proper functioning of the cell or organism. As explained in Section 9.1.2 above having too much or too little gene action for a large number of genes can disrupt the cellular metabolism to generate a phenotype or reduce viability.

Consequence #3: Decreased Fertility

Recall that during meiosis I homologous chromosomes pair up. If a cell has a chromosome with a rearrangement this chromosome will have to pair with its normal homolog.

Cells heterozygous for balanced rearrangements actually have more difficulties in prophase I. Consider the chromosomes shown in Figure 10.2.7. There are different ways they might pair during prophase I - one is shown in Figure 10.2.8. But if a crossover occurs in the inverted region the result will be unbalanced gametes. Embryos made with unbalanced gametes rarely survive. The consequence is that the heterozygous organism will have **reduced fertility**.

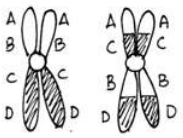


Figure 10.2.7: An unrearranged chromosome (left) and a homolog with a pericentric inversion (right). (Original-Harrington-CC:AN)

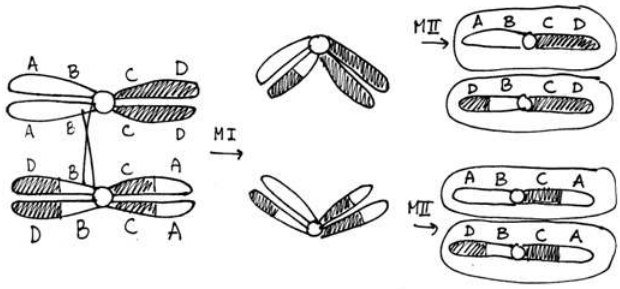


Figure 10.2.8: Meiosis in a cell heterozygous for the chromosomes shown in Figure 10.2.11. Note that of the four gametes one has a deletion of the A gene and a duplication of the D gene while another gamete has a duplication of A and a deletion of D. (Original-Harrington-CC:AN)

Note that an organism homozygous for this inversion chromosome will not be affected in this way because no loops are formed. The chromosomes can pair along their entire length and crossovers will not produce any unbalanced gametes. This is a general property of inversions and translocations. In heterozygotes there are problems during meiosis resulting in a lot of the gametes being unbalanced and an overall reduction in fertility. In homozygotes the rearranged chromosomes pair with one another just fine and there is no effect on fertility.

Consequence #4: Cancer

Some chromosome rearrangements have breakpoints within genes leading to the creation of hybrid genes – the first part of one gene with the last part of another. If the hybrid gene inappropriately promotes cell replication, the cell can become cancerous. An example of this is shown in Figure 10.2.1 where the chromosomes are from a patient with leukemia caused by a translocation between chromosomes 9 and 22 (the red and green spots side-by-side).

Consequence #5: Evolution

Those chromosome changes that duplicate genes are important for evolution. If an organism has an extra copy of important genes, one gene can be retained for their original function while others can mutate and potentially acquire new functions (Figure 10.2.9). An example of this is the multiple copies of the globin genes found in mammals (see Figure 10.2.13).

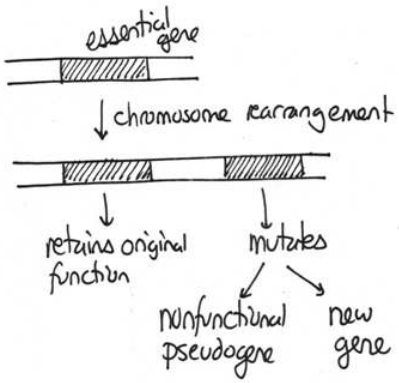


Figure 10.2.9: Duplicated genes can mutate without compromising the viability of the organism. Occasionally the result is a new gene. (Original-Harrington-CC:AN)

Chromosome rearrangements that decrease fertility are also important for the origin of new species. If a rearrangement, such as the inversion shown in Figure 10.2.9, becomes common in a small isolated population, that population has 100% fertility if they mate within their group, but a reduced fertility if they mate with members of the larger population. As rearrangements accumulate the small population will become more and more reproductively isolated. When members are incapable of forming viable, fertile offspring with the original population the group will have become a new species.

This page titled 10.2: Changes in Chromosome Structure is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **9.2: Changes in Chromosome Structure** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

10.3: Chromosome Abnormalities in Humans

The problems described above can affect all eukaryotes, unicellular and multicellular. To better understand the consequences let us consider those that affect people. As you will recall from Figure 10.3.12, humans are 2n=46. The convention when describing a person's **karyotype** (chromosome composition) is to list the total number of chromosomes, then the sex chromosomes, and then anything out of the ordinary. Most of us are 46,XX or 46,XY. What follows are some examples of chromosome number and chromosome structure abnormalities.

Down Syndrome

The most common chromosome number abnormality is trisomy-21 or, as it is more commonly known, **Down syndrome**. Having an extra copy of the smallest human chromosome, chromosome 21, causes substantial health problems.It is present in about 1 in 800 births. Infants with this condition have three copies of chromosome 21 rather than the normal two. Don't confuse **trisomy** - having three copies of one chromosome (i.e. 2n+1) with triploidy - having three entire chromosome sets (3x; see Section 2.6.) Females with trisomy-21 are 47,XX,+21 while males are 47,XY,+21. In general, people with Down syndrome are 47,sex,+21 where the word 'sex' signifies that the sex chromosomes may be XX or XY.

Trisomy-21 may arise from a nondisjunction event during meioisis in either parent or during mitosis very early during embryogenesis. However, most cases are due to a first division non-disjunction event occurring in the female parent (Figure 10.3.1).

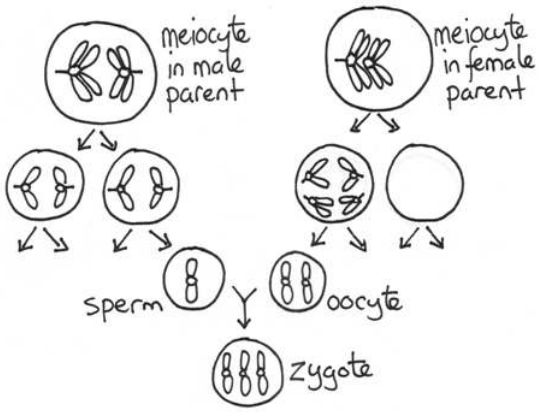


Figure 10.3.1: This diagram shows the errors during chromosome segregation that cause Down syndrome during meiosis in both parents and fusion of the gametes. Note that the cells that begin meiosis are called meiocytes and that this diagram only shows one of the four cells produced by meiosis. Meiosis occurred properly in the male parent but there was a nondisjunction event in the female parent in anaphase I. (Original-Harrington-CC:AN)

Having an extra copy of the smallest human chromosome, chromosome 21, causes substantial health problems. People with Down syndrome have various degrees of intellectual disability and often have other health problems such as heart defects. The disease was first described by John Down in 1866 but it was not until 1959 when its chromosomal basis was discovered. Current research suggests that at least some of the mental problems are due to having three copies of the *DYRK* gene on chromosome 21. This gene is active in the brain and there is evidence from humans and from mice that neurons are damaged if there is too much DYRK protein synthesized.

XYY and XXX

While fetuses trisomic for one of the other larger autosomes seldom survive to term, the situation is quite different for the sex chromosomes. Approximately 1 in 1000 males have an extra Y chromosome and yet most are unaware of it! There is little harm in having two Y chromosomes because they have relatively few genes. Similarly, 1 in 1000 females have an extra X chromosome. This situation is also relatively harmless although for a different reason. Normally in female mammals one of the two X chromosomes is inactivated in each cell so that there can be genetic balance (Figure 10.3.2 and see Section 3.5.2). In 46,XX females one of the X chromosomes is inactivated while in 47,XXX females two are inactivated.

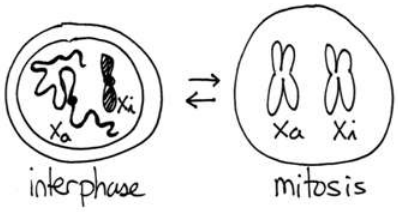


Figure 10.3.2: A simplified view of dosage compensation in mammals. Because females have more X chromosomes than males in females one X chromosome is marked as inactive (X_i). It is replicated and transmitted during cell division but most of its genes are silent. It appears as a condensed mass called the Barr body within the interphase nucleus. (Original-Harrington-CC:AN)

Turner Syndrome

Monosomy (2n-1) for autosomal chromosomes does occur at conception, but these embryos almost never survive to term. Similarly, embryos that are 45,Y are also non-viable because they lack many essential genes found on the X chromosome. The only viable monosomy in humans is 45,X, also known as **Turner syndrome**. These people are phenotypically female because they do not have a Y chromosome (Section 2.5.2). They are viable because in females only one X is active in most cells anyway. People with this condition do have health problems though: they are shorter than average, they have an elevated risk of heart defects, and they are infertile.

The reason for the health problems is that there are a few genes that are found on both the X and the Y chromosome. Because these genes are found in two copies in both XY males and XX females they are in what is called the **pseudoautosomal region**. This region escapes X chromosome inactivation. One of the genes in this region is called *SHOX*. It makes a protein that promotes bone growth. 46,XX and 46,XY people have two functioning copies and have average height. People with 47,XYY and 47,XXX genomes have three copies and are taller than average. And people with 45,X have one copy and are short. It is the single copy of *SHOX* and a few of the other genes in the pseudoautosomal region that causes health problems for women with Turner syndrome.

The reason for the infertility is that the X chromosome inactivation system only acts in somatic cells - it is not needed in the germline cells. Ovaries naturally have two functional X chromosomes. Women with Turner syndrome can not perform oogenesis because this process only works if there are two active X chromosomes. Recently, it has become possible for these women to become pregnant with donated eggs and in vitro fertilization.

Klinefelter Syndrome

There are four common sex-chromosome aneuploidies: 47,XYY, 47,XXX, 45,X, and 47,XXY. This last situation is known as **Klinefelter syndrome**. These people are male (because they have a Y chromosome) and tall (because they have three *SHOX* genes). They do not have health problems because the X chromosome inactivation system is independent of sex. In the embryonic nuclei the X chromosome are counted and all but one are shut down. It does not matter whether the embryo is male or female. Men with Klinefelter syndrome have a Barr bodies in their nuclei, the same as 46,XX females. They do have fertility problems because there are two active X chromosome in their testes and this interferes with spermatogenesis. They make enough sperm to conceive children using intracytoplasmic sperm injection though.

Cri-du-Chat Syndrome

Cri-du-chat syndrome occurs when a child inherits a defective chromosome 5 from one parent (Figure 10.3.16n). This condition is rare - it is present in only 1 in 20,000 to 1 in 50,000 births but it does account for 1% of cases of profound intellectual disability. The specific defect is a deletion that removes 2 Mb or more from the tip of the short arm of the chromosome. In most cases the deletion is the result of a chromosomal rearrangement in one of the parent's germ line cells. People with cri-du-chat have a karyotype of 46,sex,deletion(5).

As with Down syndrome this condition is associated with intellectual disability and other health problems. These problems include an improperly formed larynx which leads to infants making high pitched cat-like crying sounds (hence the name "cry of the cat"). It is suspected that at least some of the intellectual disability phenotype is due to having only a single copy of the *CTNND2* gene. This gene is active during embryogenesis and makes a protein essential for neuron migration. Down syndrome and cri-du-chat syndrome are two examples of the need for genomes to contain the proper number of genes. Having too many copies of key genes (Down syndrome) or too few (cri-du-chat syndrome) can lead to substantial developmental problems.



Figure 10.3.3: A boy with cri-du-chat syndrome. The pictures were taken at 8 months (A), 2 years (B), 4 years (C), and 9 years (D). (Wikipedia-Paola Cerruti Mainardi-CC:AN)

Inversion(9)

The most common chromosome rearrangements in humans are inversions of chromosome 9. About 2% of the world's population is heterozygous or homozygous for inversion(9). This rearrangement does not affect a person's health because the genes on the chromosome are all present - all that has changed is their relative locations. Inversion(9) is different from deletion(5) in two main respects. As mentioned above because it is a balanced rearrangement it does not cause harm. And because of this nearly everyone with an inversion(9) chromosome has inherited it from a parent who had inherited it from one of their parents and so on. In contrast, most cases of deletion(5) are due to new mutations occurring in a parent.

This page titled 10.3: Chromosome Abnormalities in Humans is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **9.3: Chromosome Abnormalities in Humans** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

10.4: Diagnosing Human Chromosome Abnormalities

Bright Field Microscopy

How can we confirm that a person has a specific chromosomal abnormality? The first method was simply to obtain a sample of their cells, stain the chromosomes with **Giemsa** dye, and examine the results with a light microscope (Figure 10.4.1). Each chromosome can be recognized by its length, the location of its centromere, and the characteristic pattern of purple bands produced by the Giemsa. For example, if mitotic cells from a person consistently contained forty seven chromosomes in total with three chromosome 21s this would be indicative of Down syndrome. Bright field microscopy does has its limitations though - it only works with mitotic chromosomes and many chromosome rearrangements are either too subtle or too complex for even a skilled cytogeneticist to discern.

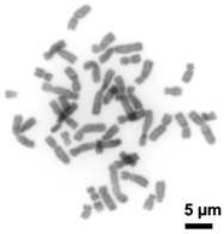


Figure 10.4.1: Human chromosomes. One way to obtain chromosomes is to take a blood sample, culture the cells for three days in the presence of a T-cell growth factor, arrest the cells in metaphase with a microtubule inhibitor, and then drop the cells onto a slide. The cells burst and the chromosomes stick to the slide. The chromosomes can then be stained or probed. Because the cells are in metaphase it is possible to see 46 replicated chromosomes here. There will be dozens of collections of chromosomes like this over the entire slide. (Wikipedia-Steffen Dietzel-CC:AS)

Fluorescence In Situ Hybridization

The solution to these problems was **fluorescence in situ hybridization** (FISH). The technique is similar to a Southern blot in that a single stranded **DNA probe** is allowed to hybridize to denatured target DNA (see Section 8.6). However, instead of the probe being radioactive it is fluorescent and instead of the target DNA being restriction fragments on a nylon membrane it is denatured chromosomes on a glass slide. Because there are several fluorescent colours available it is common to use more than one probe at the same time. Typically the chromosomes are also labeled with a fluorescent stain called **DAPI** which gives them a uniform blue colour. If the chromosomes have come from a mitotic cell it is possible to see all forty six of them spread out in a small area. Alternatively, if the chromosomes are within the nucleus of an interphase cell they appear together within a large blue circle.

Using FISH to Diagnose Down Syndrome

Most pregnancies result in healthy children. However in some cases there is an elevated chance that the fetus has trisomy-21. Older women are at a higher risk because the non-disjunction events that lead to trisomy become more frequent with age. The second consideration is what the fetus looks like during an ultrasound examination. Fetuses with trisomy-21 and some other chromosome abnormalities have a swelling in the back of the neck called a nuchal translucency. If either or both factors is present the woman may choose amniocentesis. In this test some amniotic fluid is withdrawn so that the fetal cells within it can be examined. Figure 10.4.2 shows a positive result for trisomy-21. Based upon this image the fetus has two X chromosomes and three chromosome 21s and therefore has a karyotype of 47,XX,+21.

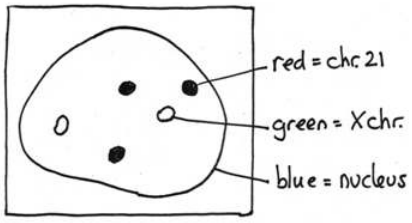


Figure 10.4.2: Confirmation of Down syndrome in a fetal cell. This diagram is based upon actual results. The DNA has been stained blue with DAPI. A red fluorescent probe is binding to the centromeres of chromosome 21 (shown here as filled circles). A green fluorescent probe is binding to the centromeres of the X chromosome (open circles). Source: Figure 4 in Antonarakis, S. E. et al. 2004. Chromosome 21 and Down syndrome: From genomics to pathophysiology. Nature Reviews Genetics 5:725-738 PubMed ID: 15510164.

Using FISH to Diagnose Cri-du-Chat Syndrome

A physician may suspect that a patient has a specific genetic condition based upon the patient's physical appearance, mental abilities, health problems, and other factors. FISH can be used to confirm the diagnosis. For example, Figure 10.4.3 shows a positive result for cri-du-chat syndrome. The probes are binding to two long arms of chromosome 5 but only one short arm. One of the chromosome 5s must therefore be missing part of its short arm.

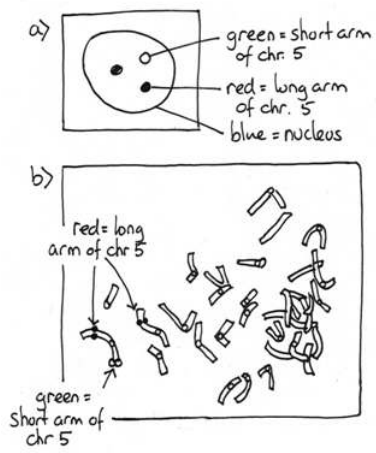


Figure 10.4.3: A positive result for cri-du-chat syndrome. This diagram is based upon actual results. Cells from a patient's blood were prepared to show an interphase nucleus (a) and mitotic chromosomes (b). The DNA has been coloured blue with DAPI. The green fluorescent probe is binding to the tip of the short arm of chromosome 5 (shown here as open circles). This is the region absent in cri-du-chat. The red fluorescent probe is binding to the middle of the long arm of the same chromosome (filled circles). This probe is used as a control. Source: Figure 1 in Fang J.-S. et al. 2008 Cytogenetic and molecular characterization of three-generation family with chromosome 5p terminal deletion. Clinical Genetics 73:585-590 PubMed ID: 18400035.

Newer Techniques

FISH is an elegant technique that produces dramatic images of our chromosomes. Unfortunately, FISH is also expensive, time consuming, and requires a high degree of skill. For these reasons, FISH is slowly being replaced with PCR and DNA chip based methods. Versions of these techniques have been developed that can accurately quantify a person's DNA. For example a sample of DNA from a person with Down syndrome will contain 150% more DNA from chromosome 21 than the other chromosomes. Likewise DNA from a person with cri-du-chat syndrome will contain 50% less DNA from the end of chromosome 5. These techniques are very useful if the suspected abnormality is a deletion, a duplication, or a change in chromosome number. They are less useful for diagnosing chromosome inversions and translocations because these rearrangements often involve no net loss or gain of genes.

In the future all of these techniques will likely be replaced with DNA sequencing. Each new generation of genome sequencing machines can sequence more DNA in less time. Eventually it will be cheaper just to sequence a patient's entire genome than to use FISH or PCR to test for specific chromosome defects.

This page titled 10.4: Diagnosing Human Chromosome Abnormalities is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **9.4: Diagnosing Human Chromosome Abnormalities** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

10.E: Changes in Chromosome Number and Structure (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

9.1

Make diagrams showing how an improper crossover event during meiosis can lead to: (a) an inversion or (b) a translocation.

9.2

Make a diagram showing how a nondisjunction event can lead to a child with a 47,XYY karyotype.

9.3

How many Barr bodies would you expect to see in cells from people who are: (a) 46, XY, (b) 46,XX, (c) 47, XYY, (d) 47,XXX, (e) 45,X, and (f) 47,XXY ?

9.4

Why can people survive with trisomy-21 (47,sex,+21) but not monosomy-21 (45,sex,-21)?

9.5

If Drosophila geneticists want to generate mutant strains with deletions they expose flies to gamma rays. What does this imply about gamma rays?

9.6

What would happen if there was a nondisjunction event involving chromosome 21 in a 46,XY zygote?

9.7

Design a FISH based experiment to find out if your lab partner is a 47,XXX female or a 47,XYY male.

9.8

What would Figure 9.18 look like if it also showed metaphase chromosomes from another cell?

This page titled 10.E: Changes in Chromosome Number and Structure (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

10.S: Changes in Chromosome Number and Structure (Summary)

- Errors during anaphase in mitosis or meiosis can lead to trisomy and other forms of aneuploidy.
- Errors during the repair of DNA breaks or during meiotic crossing over can lead to chromosome rearrangements.
- Five common forms of aneuploidy in humans are 47,sex,+21 (Down syndrome), 47,XYY, 47,XXX, 45,X (Turner syndrome) and 47,XXY (Klinefelter syndrome).
- Deletion(5) causes a serious condition (cri-du-chat syndrome) because deletions are unbalanced chromosome rearrangements.
- Inversion(9) causes few health consequences because inversions are balanced chromosome rearrangements.
- Bright field microscopy can be used to detect chromosome number abnormalities and some chromosome rearrangements.
- Fluorescence in situ hybridization can be used to detect all types of chromosome abnormalities.
- PCR and DNA chip based techniques can be used to detect chromosome number abnormalities, deletions, and duplications.

Key Terms:

origin of replication
telomere
centromere
non-disjunction
euploid
aneuploid
balanced
unbalanced
first division nondisjunction
second division nondisjunction
double strand break
nonhomologous end joining
DNA repair system
chromosome rearrangement
deletion
inversion
paracentric inversion
pericentric inversion
tandem duplication
translocation
reciprocal translocation
Robertsonian translocation
meiotic crossover
deletion loop
karyotype
46,XX
46,XY
47,sex,+21 (Down syndrome)
trisomy
47,XYY
47,XXX
monosomy
45,X (Turner syndrome)
pseudoautosomal region
47,XXY (Klinefelter syndrome)
46,sex,deletion(5) (cri-du-chat syndrome)
46,sex,inversion(9)
bright field microscopy
Giemsa stain
fluorescence in situ hybridization
fluorescent probe
DAPI stain
amniocentesis

This page titled 10.S: Changes in Chromosome Number and Structure (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **9.S: Changes in Chromosome Number and Structure (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

11: Pedigrees and Populations

The basic concepts of genetics described in the preceding chapters can be applied to almost any eukaryotic organism. However, some techniques, such as test crosses, can only be performed with model organisms or other species that can be experimentally manipulated. To study the inheritance patterns of genes in humans and other species for which controlled matings are not possible, geneticists use the analysis of pedigrees and populations.

11.1: Prelude to Pedigrees and Populations
11.2: Pedigree Analysis
11.3: Inferring the Mode of Inheritance
11.4: Sporadic and Non-Heritable Diseases
11.5: Calculating Probabilities
11.6: Population Genetics
11.E: Pedigrees and Populations (Exercises)
11.S: Pedigrees and Populations (Summary)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 11: Pedigrees and Populations is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Prelude to Pedigrees and Populations

To study the inheritance patterns of genes in humans and other species for which controlled matings are not possible, geneticists use the analysis of pedigrees and populations.



Figure 11.1.1: Polydactyly (six fingers in this case) is an example of a human trait that can be studied by pedigree analysis. (Wikipedia-en:User:DRGNU23-GFDL)

This page titled 11.1: Prelude to Pedigrees and Populations is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu.

- **5.1: Prelude to Pedigrees and Populations** is licensed CC BY-SA 4.0.

11.2: Pedigree Analysis

Pedigree charts are diagrams that show the phenotypes and/or genotypes for a particular organism and its ancestors. While commonly used in human families to track genetic diseases, they can be used for any species and any inherited trait. Geneticists use a standardized set of symbols to represent an individual's sex, family relationships and phenotype. These diagrams are used to determine the **mode of inheritance** of a particular disease or trait, and to predict the probability of its appearance among offspring. Pedigree analysis is therefore an important tool in both basic research and **genetic counseling**.

Each pedigree chart represents all of the available information about the inheritance of a single trait (most often a disease) within a family. The pedigree chart is therefore drawn using factual information, but there is always some possibility of errors in this information, especially when relying on family members' recollections or even clinical diagnoses. In real pedigrees, further complications can arise due to **incomplete penetrance** (including age of onset) and **variable expressivity** of disease alleles, but for the examples presented in this book, we will presume complete accuracy of the pedigrees. A pedigree may be drawn when trying to determine the nature of a newly discovered disease, or when an individual with a family history of a disease wants to know the probability of passing the disease on to their children. In either case, a tree is drawn, as shown in Figure 11.2.2, with circles to represent females, and squares to represent males. Matings are drawn as a line joining a male and female, while a consanguineous mating (closely related is two lines.

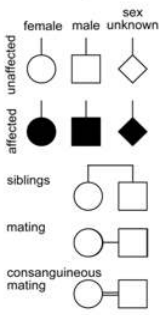


Figure 11.2.2: Symbols used in drawing a pedigree. (Original-Deyholos-CC:AN)

The affected individual that brings the family to the attention of a geneticist is called the **proband** (or propositus). If an individual is known to have symptoms of the disease (**affected**), the symbol is filled in. Sometimes a half-filled in symbol is used to indicate a known **carrier** of a disease; this is someone who does not have any symptoms of the disease, but who passed the disease on to subsequent generations because they are a heterozygote. Note that when a pedigree is constructed, it is often unknown whether a particular individual is a carrier or not, so not all carriers are always explicitly indicated in a pedigree. For simplicity, in this chapter we will assume that the pedigrees presented are accurate, and represent fully penetrant traits.

This page titled 11.2: Pedigree Analysis is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **5.2: Pedigree Analysis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.3: Inferring the Mode of Inheritance

Given a pedigree of an uncharacterized disease or trait, one of the first tasks is to determine which modes of inheritance are possible and then which mode of inheritance is most likely. This information is essential in calculating the probability that the trait will be inherited in any future offspring. We will mostly consider five major types of inheritance: autosomal dominant (AD), autosomal recessive (AR), X-linked dominant (XD), X-linked recessive (XR), and Y-linked (Y).

Autosomal Dominant (AD)

When a disease is caused by a dominant allele of a gene, every person with that allele will show symptoms of the disease (assuming complete penetrance), and only one disease allele needs to be inherited for an individual to be affected. Thus, every affected individual must have an affected parent. A pedigree with affected individuals in every generation is typical of AD diseases. However, beware that other modes of inheritance can also show the disease in every generation, as described below. It is also possible for an affected individual with an AD disease to have a family without any affected children, if the affected parent is a heterozygote. This is particularly true in small families, where the probability of every child inheriting the normal, rather than disease allele is not extremely small. Note that AD diseases are usually rare in populations, therefore affected individuals with AD diseases tend to be heterozygotes (otherwise, both parents would have had to been affected with the same rare disease). Achondroplastic dwarfism, and polydactyly are both examples of human conditions that may follow an AD mode of inheritance.

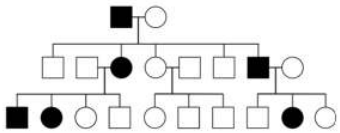


Figure 11.3.3: A pedigree consistent with AD inheritance. (Original-Deyholos-CC:AN)

X-linked dominant (XD)

In X-linked dominant inheritance, the gene responsible for the disease is located on the X-chromosome, and the allele that causes the disease is dominant to the normal allele in females. Because females have twice as many X-chromosomes as males, females tend to be more frequently affected than males in the population. However, not all pedigrees provide sufficient information to distinguish XD and AD. One definitive indication that a trait is inherited as AD, and not XD, is that an affected father passes the disease to a son; this type of transmission is not possible with XD, since males inherit their X chromosome from their mothers.

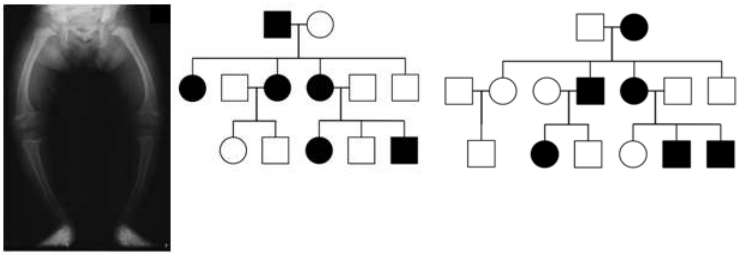


Figure 11.3.5: Some types of rickets may follow an XD mode of inheritance. (Wikipedia-Mrish-CC:AS) (middle and left) Two pedigrees consistent with XD inheritance. (Original-Deyholos_CC:AN)

Autosomal recessive (AR)

Diseases that are inherited in an autosomal recessive pattern require that both parents of an affected individual carry at least one copy of the disease allele. With AR traits, many individuals in a pedigree can be carriers, probably without knowing it. Compared to pedigrees of dominant traits, AR pedigrees tend to show fewer affected individuals and are more likely than AD or XD to “skip a generation”. Thus, the major feature that distinguishes AR from AD or XD is that unaffected individuals can have affected offspring.



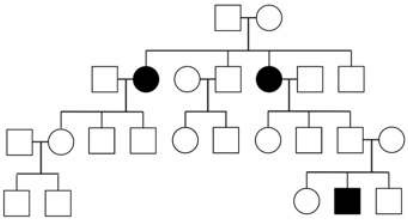


Figure 11.3.6: (left) Many inborn errors of metabolism, such as phenylketonuria (PKU) are inherited as AR. Newborns are often tested for a few of the most common metabolic diseases. (Wikipedia-U.S. Air Force photo/Staff Sgt. Eric T. Sheler-PD) (right) A pedigree consistent with AR inheritance. (Original-Deyholos_CC:AN)

X-linked recessive (XR)

Because males have only one X-chromosome, any male that inherits an X-linked recessive disease allele will be affected by it (assuming complete penetrance). Therefore, in XR modes of inheritance, males tend to be affected more frequently than females in a population. This is in contrast to AR and AD, where both sexes tend to be affected equally, and XD, in which females are affected more frequently. Note, however, in the small sample sizes typical of human families, it is usually not possible to accurately determine whether one sex is affected more frequently than others. On the other hand, one feature of a pedigree that can be used to definitively establish that an inheritance pattern is not XR is the presence of an affected daughter from unaffected parents; because she would have had to inherit one X-chromosome from her father, he would also have been affected in XR.

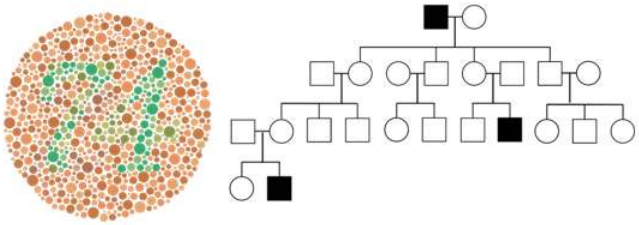


Figure 11.3.8: Some forms of colour blindness are inherited as XR-traits. Colour blindness is diagnosed using tests such as this. Ishihara Test. (Wikipedia-unknown-PD) (right) A pedigree consistent with XR inheritance. (Original-Deyholos-CC:AN)

Y-linked and Mitochondrial Inheritance.

Two additional modes are **Y-linked** and **Mitochondrial inheritance**.

Only males are affected in human Y-linked inheritance (and other species with the X/Y sex determining system). There is only father to son transmission. This is the easiest mode of inheritance to identify, but it is one of the rarest because there are so few genes located on the Y-chromosome. An example of Y-linked inheritance is the **hairy-ear-rim** phenotype seen in some Indian families. As expected this trait is passed on from father to all sons and no daughters. Y-chromosome DNA polymorphisms can be used to follow the male lineage in large families or through ancient ancestral lineages. For example, the Y-chromosome of Mongolian ruler Genghis Khan (1162-1227 CE), and his male relatives, accounts for ~8% of the Y-chromosome lineage of men in Asia (0.5% world wide).

Mutations in Mitochondrial DNA are inherited through the maternal line (from your mother). There are some human diseases associated with mutations in mitochondria genes. These mutations can affect both males and females, but males cannot pass them on as the mitochondria are inherited via the egg, not the sperm. Mitochondrial DNA polymorphisms are also used to investigate evolutionary lineages, both ancient and recent. Because of the relative similarity of sequence mtDNA is also used in species identification in ecology studies.

This page titled 11.3: Inferring the Mode of Inheritance is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **5.3: Inferring the Mode of Inheritance** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.4: Sporadic and Non-Heritable Diseases

Not all of the characterized human traits and diseases are attributed to mutant alleles at a single gene locus. Many diseases that have a heritable component, have more complex inheritance patterns due to (1) the involvement of multiple genes, and/or (2) environmental factors. On the other hand, some non-genetic diseases may appear to be heritable because they affect multiple members of the same family, but this is due to the family members being exposed to the same toxins or other environmental factors (e.g. in their homes).

Finally, diseases with similar symptoms may have different causes, some of which may be genetic while others are not. One example of this is ALS (amyotrophic lateral sclerosis); approximately 5-10% of cases are inherited in an AD pattern, while the majority of the remaining cases appear to be **sporadic**, in other words, not caused by a mutation inherited from a parent. We now know that different genes or proteins are affected in the inherited and sporadic forms of ALS. The physicist Stephen Hawking (Figure 11.4.10) and baseball player Lou Gehrig both suffered from sporadic ALS.



Figure 11.4.10: Stephen Hawking(Wikipedia-NASA-PD)

This page titled 11.4: Sporadic and Non-Heritable Diseases is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **5.4: Sporadic and Non-Heritable Diseases** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.5: Calculating Probabilities

Once the mode of inheritance of a disease or trait is identified, some inferences about the genotype of individuals in a pedigree can be made, based on their phenotypes and where they appear in the family tree. Given these genotypes, it is possible to calculate the probability of a particular genotype being inherited in subsequent generations. This can be useful in genetic counseling, for example when prospective parents wish to know the likelihood of their offspring inheriting a disease for which they have a family history.

Probabilities in pedigrees are calculated using knowledge of Mendelian inheritance and the same basic methods as are used in other fields. The first formula is the **product rule**: the joint probability of two independent events is the product of their individual probabilities; this is the probability of one event AND another event occurring. For example, the probability of a rolling a "five" with a single throw of a single six-sided die is 1/6, and the probability of rolling "five" in each of three successive rolls is 1/6 x 1/6 x 1/6 = 1/216. The second useful formula is the **sum rule**, which states that the combined probability of two independent events is the sum of their individual probabilities. This is the probability of one event OR another event occurring. For example, the probability of rolling a five or six in a single throw of a dice is 1/6 + 1/6 = 1/3.

With these rules in mind, we can calculate the probability that two carriers (i.e. heterozygotes) of an AR disease will have a child affected with the disease as ½ x ½ = ¼, since for each parent, the probability of any gametes carrying the disease allele is ½. This is consistent with what we already know from calculating probabilities using a Punnett Square (e.g. in a monohybrid cross *Aa* x *Aa*, ¼ of the offspring are *aa*).

We can likewise calculate probabilities in the more complex pedigree shown in Figure 11.5.11.

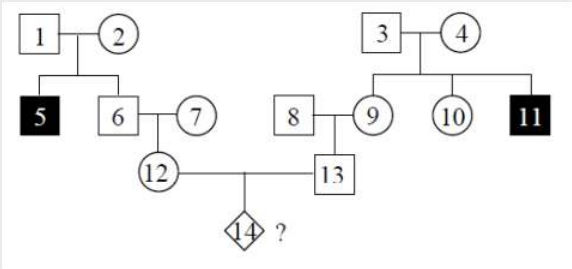


Figure 11.5.11: Individuals in this pedigree are labeled with numbers to make discussion easier. (Original-Deyholos-CC:AN)

Assuming the disease has an AR pattern of inheritance, what is the probability that individual 14 will be affected? We can assume that individuals #1, #2, #3 and #4 are heterozygotes (*Aa*), because they each had at least one affected (*aa*) child, but they are not affected themselves. This means that there is a 2/3 chance that individual #6 is also *Aa*. This is because according to Mendelian inheritance, when two heterozygotes mate, there is a 1:2:1 distribution of genotypes *AA*:*Aa*:*aa*. However, because #6 is unaffected, he can't be *aa*, so he is either *Aa* or *AA*, but the probability of him being *Aa* is twice as likely as *AA*. By the same reasoning, there is likewise a 2/3 chance that #9 is a heterozygous carrier of the disease allele.

If individual 6 is a heterozygous for the disease allele, then there is a ½ chance that #12 will also be a heterozygote (i.e. if the mating of #6 and #7 is *Aa* × *AA*, half of the progeny will be *Aa*; we are also assuming that #7, who is unrelated, does not carry any disease alleles). Therefore, the combined probability that #12 is also a heterozygote is 2/3 x 1/2 = 1/3. This reasoning also applies to individual #13, i.e. there is a 1/3 probability that he is a heterozygote for the disease. Thus, the overall probability that both individual #12 and #13 are heterozygous, and that a particular offspring of theirs will be homozygous for the disease alleles is 1/3 x 1/3 x 1/4 = 1/36.

This page titled 11.5: Calculating Probabilities is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **5.5: Calculating Probabilities** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.6: Population Genetics

A **population** is a large group of individuals of the same species, who are capable of mating with each other. It is useful to know the frequency of particular alleles within a population, since this information can be used to calculate disease risks. Population genetics is also important in ecology and evolution, since changes in allele frequencies may be associated with migration or natural selection.



Figure 11.6.12: (Flickr-Zach Stern-CC:AND)

Allele frequencies may also be studied at the population level

The frequency of different alleles in a population can be determined from the frequency of the various phenotypes in the population. In the simplest system, with two alleles of the same locus (e.g. *A*,*a*), we use the symbol **p** to represent the frequency of the dominant allele within the population, and **q** for the frequency of the recessive allele. Because there are only two possible alleles, we can say that the frequency of p and q together represent 100% of the alleles in the population (**p+q=1**).

We can calculate the values of p and q, in a representative sample of individuals from a population, by simply counting the alleles and dividing by the total number of alleles examined. For a given allele, homozygotes will count for twice as much as heterozygotes.

For example:

genotype	number of individuals
AA	320
Aa	160
aa	20

	A (p)	*a* (q)
A (p)	p^2	pq
a (q)	pq	q^2

$$\mathbf{p} = \frac{\mathbf{2(AA) + Aa}}{\mathbf{total\ alleles\ counted}} \tag{11.6.1}$$

$$= \frac{2(320) + 160}{2(320) + 2(160) + 2(20)} = 0.8 \tag{11.6.2}$$

$$\mathbf{q} = \frac{\mathbf{2(aa) + Aa}}{\mathbf{total\ alleles\ counted}} \tag{11.6.3}$$

$$= \frac{2(20) + 160}{2(320) + 2(160) + 2(20)} = 0.2 \tag{11.6.4}$$

Hardy-Weinberg formula

With the allele frequencies of a population we can use an extension of the Punnett Square, and the product rule, to calculate the expected frequency of each genotype following random matings within the entire population. This is the basis of the *Hardy-Weinberg* formula:

$$p^2 + 2pq + q^2 = 1 \tag{11.6.5}$$

Here p^2 is the frequency of homozygotes *AA*, $2pq$ is the frequency of the heterozygotes, and q^2 is the frequency of homozygotes *aa*. Notice that if we substitute the allele frequencies we calculated above (p=0.8, q=0.2) into the Equation 11.6.5**,** we obtain expected probabilities for each of the genotypes that exactly match our original observations:

$$\mathrm{p}^2 = 0.8^2 = 0.64 \qquad 0.64 \times 500 = 320 \tag{11.6.6}$$

$$2\mathrm{pq} = 2(0.8)(0.2) = 0.32 \qquad 0.32 \times 500 = 160 \tag{11.6.7}$$

$$\mathrm{q}^2 = 0.2^2 = 0.04 \qquad 0.04 \times 500 = 20 \tag{11.6.8}$$

This is a demonstration of the **Hardy-Weinberg Equilibrium**, where both the genotype frequencies and allele frequencies in a population remain unchanged following successive matings within a population, *if* certain conditions are met. These conditions are listed in Table 11.6.1. Few natural populations actually satisfy all of these conditions. Nevertheless, large populations of many species, including humans, appear to approach Hardy-Weinberg equilibrium for many loci. In these situations, deviations of a particular gene from Hardy-Weinberg equilibirum can be an indication that one of the alleles affects the reproductive success of organism, for example through natural selection or **assortative mating**.

Conditions for the Hardy-Weinberg equilibrium

- **Random mating**: Individuals of all genotypes mate together with equal frequency. Assortative mating, in which certain genotypes preferentially mate together, is a type of non-random mating.
- **No natural selection**: All genotypes have equal fitness.
- **No migration**: Individuals do not leave or enter the population.
- **No mutation**: The allele frequencies do not change due to mutation.
- **Large population**: Random sampling effects in mating (i.e. genetic drift) are insignificant in large populations.

The Hardy-Weinberg formula can also be used to estimate allele frequencies, when only the frequency of one of the genotypic classes is known. For example, if 0.04% of the population is affected by a particular genetic condition, and all of the affected individuals have the genotype aa, then we assume that $\mathrm{q}^2 = 0.0004$ and we can calculate p, q, and 2pq as follows:

$$\mathrm{q}^2 = 0.04\% = 0.0004 \tag{11.6.9}$$

$$\mathrm{q} = \sqrt{0.0004} = 0.02 \tag{11.6.10}$$

$$\mathrm{p} = 1 - \mathrm{q} = 0.98 \tag{11.6.11}$$

$$2\mathrm{pq} = 2(0.98)(0.02) = 0.04 \tag{11.6.12}$$

Thus, approximately 4% of the population is expected to be heterozygous (i.e. a carrier) of this genetic condition. Note that while we recognize that the population is probably not exactly in Hardy-Weinberg equilibrium for this locus, application of the Hardy-Weinberg formula nevertheless can give a reasonable estimate of allele frequencies, in the absence of any other information.

This page titled 11.6: Population Genetics is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

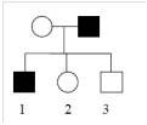
- **5.6: Population Genetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.E: Pedigrees and Populations (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions

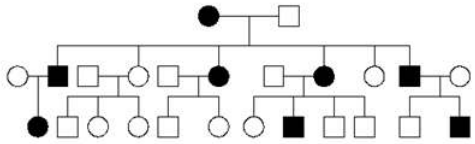
1. What are some of the modes of inheritance that are consistent with this pedigree?



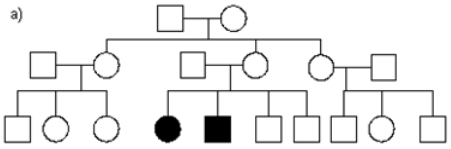
2. In this pedigree in question 1, the mode of inheritance cannot be determined unamibguously.What are some examples of data (e.g. from other generations) that, if added to the pedigree would help determine the mode of inheritance?

3. For each of the following pedigrees, name the most likely mode of inheritance (AR=autosomal recessive, AD=autosomal dominant, XR=X-linked recessive, XD=X-linked dominant).(These pedigrees were obtained from various external sources).

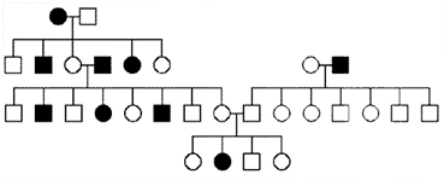
a)



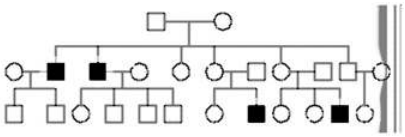
b)



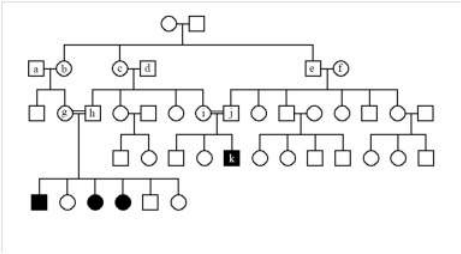
c)



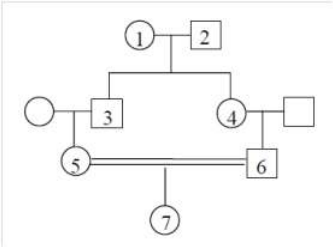
d)



4. The following pedigree represents a rare, autosomal recessive disease. What are the genotypes of the individuals who are indicated by letters?



5. If individual #1 in the following pedigree is a heterozygote for a rare, AR disease, what is the probability that individual #7 will be affected by the disease? Assume that #2 and the spouses of #3 and #4 are not carriers.



6. You are studying a population in which the frequency of individuals with a recessive homozygous genotype is 1%. Assuming the population is in Hardy-Weinberg equilibrium, calculate:

a) The frequency of the recessive allele.

b) The frequency of dominant allele.

c) The frequency of the heterozygous phenotype.

d) The frequency of the homozygous dominant phenotype.

7. Determine whether the following population is in Hardy-Weinberg equilibrium.

genotype	number of individuals
AA	432
Aa	676
aa	92

8. Out of 1200 individuals examined, 432 are homozygous dominant (*AA*)for a particular gene. What numbers of individuals of the other two genotypic classes (*Aa*, *aa*) would be expected if the population is in Hardy-Weinberg equilibrium?

9. Propose an explanation for the deviation between the genotypic frequencies calculated in question 8 and those observed in the table in question 7.

Chapter 5 - Answers

5.1 The pedigree could show an AD, AR or XR mode of inheritance. It is most likely AD. It could be AR if the mother was a carrier, and the father was a homozygote. It could be XR if the mother was a carrier, and the father was a hemizygote. It cannot be XD, since the daughter (#2) would have necessarily inherited the disease allele on the X chromosome she received from her father.

5.2 There are many possible answers. Here are some possibilities: if neither of the parents of the father were affected (i.e. the paternal grandparents of children 1, 2, 3), then the disease could not be dominant. If only the paternal grandfather was affected, then the disease could only be X-linked recessive if the paternal grandmother was a heterozygote (which would be unlikely given that this is a rare disease allele).

5.3 a) The mode of inheritance is most likely AD, since every affected individual has an affected parent, and the disease is inherited even in four different matings to unrelated, unaffected individuals. It is very unlikely that it is XD or XR, in part because affected father had an affected son.

b) The mode of inheritance cannot be AD or XD, because affected individuals must have an affected parent when a disease allele is dominant. Neither can it be XR, because there is an affected daughter of a normal father. Therefore, it must be AR, and this is consistent with the pedigree.

c) The mode of inheritance cannot be AD or XD, because, again, there are affected individuals with unaffected parents. It is not XR, because there are unaffected sons of an affected mother. It is therefore likely AR, but note that the recessive alleles for this condition appear to be relatively common in the population (note that two of the marriages were to unrelated, affected individuals).

d) The mode of inheritance cannot be AD or XD, because, again, there are affected individuals with unaffected parents. It could be either XR or AR, but because all of the affected individuals are male, and no affected males pass the disease to their sons, it is likely XR.

5.4 If *a* represents the disease allele, individuals a, d, f (who all married into this unusual family) are *AA*, while b, c, e, g, h, i, j are all *Aa*, and k is *aa*.

5.5 There is a ½ chance that an offspring of any mating Aa x AA will be a carrier (Aa). So, there is a ½ chance that #3 will be Aa, and likewise for #4. If #3 is a carrier, there is again a ½ chance that #5 will be a carrier, and likewise for #6. If #5 and #6 are both Aa, then there is a ¼ chance that this monohybrid cross will result in #7 having the genotype aa, and therefore being affected by the disease. Thus, the joint probability is 1/2 x 1/2 x 1/2 x 1/2 x 1/4 =**1/64**.

5.6 a) q = -0.01. = 0.1

b) 1-q = p; 1-0.1 = 0.9

c) 2pq = 2(0.1)(0.9) = 0.18

d) p^2 = 0.81

5.7 First, calculate allele frequencies:

$$\mathrm{p} = \frac{2(\mathrm{AA}) + (\mathrm{Aa})}{\text{total number of alleles scored}} = \frac{2(432) + 676}{2(432 + 676 + 92)} = 0.6417 \tag{11.E.1}$$

$$\mathrm{q} = \frac{2(\mathrm{aa}) + (\mathrm{Aa})}{\text{total number of alleles scored}} = \frac{2(92) + 676}{2(432 + 676 + 92)} = 0.3583 \tag{11.E.2}$$

Next, given these observed allele frequencies, calculate the genotypic frequencies that would be expected if the population was in Hardy-Weinberg equilibrium.

$$\mathrm{p}^2 = 0.6417^2 = 0.4118 \tag{11.E.3}$$

$$2\mathrm{pq} = 2(0.6417)(0.3583) = 0.4598 \tag{11.E.4}$$

$$\mathrm{q}^2 = 0.3583^2 = 0.1284 \tag{11.E.5}$$

Finally, given these expected frequencies of each class, calculate the expected numbers of each in your sample of 1200 individuals, and compare these to your actual observations.

expected	observed	(reported in the original question)
AA 0.4118 × 1200 = 494	432	
Aa 0.4598 × 1200 = 552	676	
aa 0.1284 × 1200 = 154	92	

The population does not appear to be at Hardy-Weinberg equilibrium, since the observed genotypic frequencies do not match the expectations. Of course, you could do a chi-square test to determine how significant the discrepancy is between observed and expected.

5.8 If in this theoretical question, the frequency of genotype of *AA* is set at 432/1200 and we are asked what frequencies of the other classes would fit a Hardy-Weinberg equilibrium. So, given that p^2 = 432/1200, then p=0.6, and q=0.4. Given these allele frequencies and a sample size of 1200 individuals, then there should be 576 *Aa* individuals (2pq × 1200 = 2(0.6)(0.4) × 1200=576)

and 192 *aa* individuals ($q^2 \times 1200 = 0.4^2 \times 1200 = 192$), if the population was at Hardy-Weinberg equilibrium with 432 *AA* individuals.

5.9 The actual population appear to have more heterozygotes and few recessive homozygotes than would be expected for Hardy-Weinberg equilibrium. There are many possible reasons that a population may not be in equilibrium (see Table 5.1). In this case, there is possibly some selection against homozygous recessive genotypes, in favour of heterozygotes in particular. Perhaps the heterozygotes have some selective advantage that increases their fitness.

It is also worth noting the discrepancies between the allele frequencies calculated in 5.8 and 5.7. In question 5.7, we calculated the frequencies directly from the genotypes (this is the most accurate method, and does not require the population to be in equilibrium). In 5.8, we essentially estimated the frequency base on one of the phenotypic classes. The discrepancy between these calculations shows the limitations of using phenotypes to estimate allele frequencies, when a population is not in equilibrium.

This page titled 11.E: Pedigrees and Populations (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

11.S: Pedigrees and Populations (Summary)

- Pedigree analysis can be used to determine the mode of inheritance of specific traits such as diseases. Loci can be X- or Y-linked or autosomal in location and alleles either dominant or recessive with respect to wild type.
- If the mode of inheritance is known, a pedigree can be used to calculate the probability of inheritance of a particular genotype by an individual.
- The frequencies of all alleles and genotypes remain unchanged through successive generations of a population in Hardy-Weinberg equilibrium.
- Populations in true Hardy-Weinberg equilibrium have random mating, and no genetic drift, no migration, no mutation, and no selection with respect to the gene of interest.
- The Hardy-Weinberg formula can be used to estimate allele and genotype frequencies given only limited information about a population.

Key Terms

Pedigree charts		
mode of inheritance	X-linked recessive	
genetic counseling	Y-linked inheritance	Hardy-Weinberg formula
incomplete penetrance	Y-linked	$p^2 + 2pq + q^2=1$
variable expressivity	mitochondrial inheritance (mtDNA)	Hardy-Weinberg equilibrium
proband	sporadic	assortative mating
affected	product rule	random mating
carrier	sum rule	migration
autosomal dominant	population	genetic drift
autosomal recessive	p+q=1	
X-linked dominant		

This page titled 11.S: Pedigrees and Populations (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **5.S: Pedigrees and Populations (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

12: Techniques of Molecular Genetics

Genetics is the study of the inheritance and variation of biological traits. We have previously noted that it is possible to conduct genetic research without directly studying DNA. Indeed some of the greatest geneticists had no special knowledge of DNA at all, but relied instead on analysis of phenotypes, inheritance patterns, and their ratios in carefully designed crosses.

12.1: Prelude to Molecular Genetics
12.2: Isolating Genomic DNA
12.3: Isolating or Detecting a Specific Sequence by PCR
12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
12.5: Cloning DNA - Plasmid Vectors
12.6: DNA Analysis - Gel Electrophoresis
12.7: DNA Analysis- Blotting and Hybridization
12.8: Transgenic organisms
12.9: Techniques of Molecular Genetics (Exercises)
12.S: Techniques of Molecular Genetics (Summary)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 12: Techniques of Molecular Genetics is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

12.1: Prelude to Molecular Genetics

Genetics is the study of the inheritance and variation of biological traits. We have previously noted that it is possible to conduct genetic research without directly studying DNA. Indeed some of the greatest geneticists had no special knowledge of DNA at all, but relied instead on analysis of phenotypes, inheritance patterns, and their ratios in carefully designed crosses. Today, **classical genetics** is often complemented by **molecular biology**, to give **molecular genetics**, which involves the study of DNA and other **macromolecules** that have been isolated from an organism. Usually, molecular genetics experiments involve some combination of techniques to isolate and analyze the DNA or RNA transcribed from a particular gene. In some cases, the DNA may be subsequently manipulated by mutation or by recombination with other DNA fragments. Techniques of molecular genetics have wide application in many fields of biology, as well as forensics, biotechnology, and medicine.

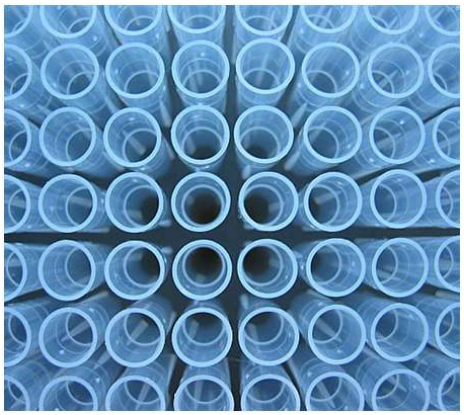


Figure 12.1.1***:*** *Disposable tips for a pipette are used to distribute microliter volumes of liquid in molecular biology. (Flickr-esstherase-CC:ANS)*

This page titled 12.1: Prelude to Molecular Genetics is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.1: Prelude to Molecular Genetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.2: Isolating Genomic DNA

DNA purification strategies rely on the chemical properties of DNA that distinguish it from other molecules in the cell, namely that it is a very long, negatively charged molecule. To extract purified DNA from a tissue sample, cells are broken open by grinding or **lysing** in a solution that contains chemicals that protect the DNA while disrupting other components of the cell (Figure 12.2.2). These chemicals may include **detergents**, which dissolve lipid membranes and denature proteins. A cation such as Na^+ helps to stabilize the negatively charged DNA and separate it from proteins such as histones. A chelating agent, such as EDTA, is added to protect DNA by sequestering Mg^{2+} ions, which can otherwise serve as a necessary co-factor for **nucleases** (enzymes that digest DNA). As a result, free, double-stranded DNA molecules are released from the chromatin into the extraction buffer, which also contains proteins and all other cellular components. (The basics of this procedure can be done with household chemicals and are presented on YouTube.)

The free DNA molecules are subsequently isolated by one of several methods. Commonly, proteins are removed by adjusting the salt concentration so they precipitate. The **supernatant**, which contains DNA and other, smaller metabolites, is then mixed with ethanol, which causes the DNA to precipitate. A small **pellet** of DNA can be collected by centrifugation, and after removal of the ethanol, the DNA pellet can be dissolved in water (usually with a small amount of EDTA and a pH buffer) for the use in other reactions. Note that this process has purified all of the DNA from a tissue sample; if we want to further isolate a specific gene or DNA fragment, we must use additional techniques, such as PCR.

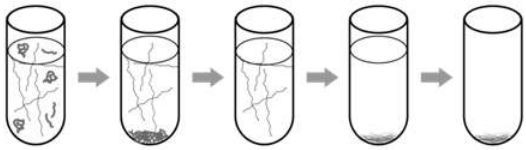


Figure 12.2.2: Extraction of DNA from a mixture of solubilized cellular components by successive precipitations. Proteins are precipitated, then DNA (in the supernatant) is precipitated in ethanol, leaving a pellet of DNA. (Original-Deyholos-CC:AN)

This page titled 12.2: Isolating Genomic DNA is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.2: Isolating Genomic DNA** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.3: Isolating or Detecting a Specific Sequence by PCR

Components of the PCR Reaction

The **Polymerase Chain Reaction (PCR)** is a method of DNA replication that is performed in a test tube (i.e. *in vitro*). Here "polymerase" refers to a DNA polymerase enzyme extracted and purified from bacteria, and "chain reaction" refers to the ability of this technique produce millions of copies of a DNA molecule, by using each newly replicated double helix as a template to synthesize two new DNA double helices. PCR is therefore a very efficient method of amplifying DNA.

Besides its ability to make large amounts of DNA, there is a second characteristic of PCR that makes it extremely useful. Recall that most DNA polymerases can only add nucleotides to the end of an existing strand of DNA, and therefore require a **primer** to initiate the process of replication. For PCR, chemically synthesized primers of about 20 nucleotides are used. In an ideal PCR, primers only hybridize to their exact complementary sequence on the template strand (Figure 12.3.3).



Figure 12.3.3: The primer-template duplex at the top part of the figure is perfectly matched, and will be stable at a higher temperature than the duplex in the bottom part of the figure, which contains many mismatches and therefore fewer hydrogen bonds. If the annealing temperature is sufficiently high, only the perfectly matched primer will be able to initiate extension (grey arrow) from this site on the template. (Original-Deyholos-CC:AN)

The experimenter can therefore control exactly what region of a DNA template is amplified by controlling the sequence of the primers used in the reaction.

To conduct a PCR amplification, an experimenter combines in a small, thin-walled tube (Figure 12.3.4), all of the necessary components for DNA replication, including DNA polymerase and solutions containing nucleotides (dATP, dCTP, dGTP, dTTP), a DNA template, DNA primers, a pH buffer, and ions (e.g. Mg^{2+}) required by the polymerase. Successful PCR reactions have been conducted using only a single DNA molecule as a template, but in practice, most PCR reactions contain many thousands of template molecules. The template DNA (e.g. total genomic DNA) has usually already been purified from cells or tissues using the techniques described above. However, in some situations it is possible to put whole cells directly in a PCR reaction for use as a template.



Figure 12.3.4: A strip of *PCR tubes (Wikipedia-madprime-GFDL)*

An essential aspect of PCR is **thermal-cycling**, meaning the exposure of the reaction to a series of precisely defined temperatures (Figure 12.3.5). The reaction mixture is first heated to 95°C. This causes the hydrogen bonds between the strands of the template DNA molecules to melt, or **denature**. This produces two single-stranded DNA molecules from each double helix (Figure 12.3.6). In the next step (**annealing**), the mixture is cooled to 45-65°C. The exact temperature depends on the primer sequence used and the objectives of the experiment. This allows the formation of double stranded helices between complementary DNA molecules, including the annealing of primers to the template. In the final step (**extension**) the mixture is heated to 72°C. This is the temperature at which the particular DNA polymerase used in PCR is most active. During extension, the new DNA strand is synthesized, starting from the 3' end of the primer, along the length of the template strand. The entire PCR process is very quick, with each temperature phase usually lasting 30 seconds or less. Each cycle of three temperatures (denaturation, annealing, extension) is usually repeated about 30 times, amplifying the target region approximately 2^{30}-fold. Notice from the figure that most of the newly synthesized strands in PCR begin and end with sequences either identical to or complementary to the primer sequences; although a few strands are longer than this, they are in such a small minority that they can almost always be ignored.

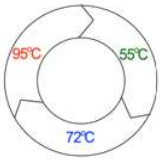


Figure 12.3.5: Example of a thermalcycle, in which the annealing temperature is 55°C. (Original-Deyholos-CC:AN)

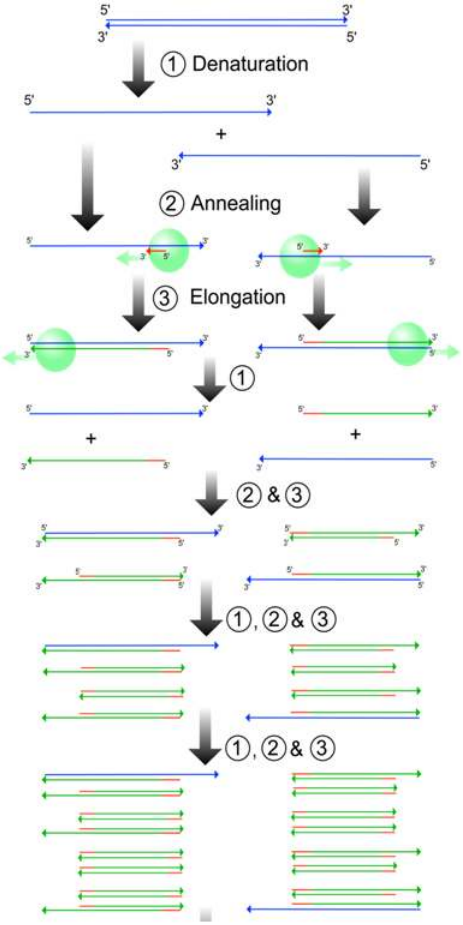


Figure 12.3.6***:*** *PCR with the three phases of the thermalcycle numbered. The template strand (blue) is replicated from primers (red), with newly synthesized strands in green. The green strands flanked by two primer binding sites will increase in abundance exponentially through successive PCR cycles. (Wikipedia-madprime-GFDL)*

The earliest PCR reactions used a polymerase from *E. coli*. Because the high temperature of the denaturation step destroyed the enzyme, new polymerase had to be added after each cycle. To overcome this, researchers identified **thermostable** DNA polymerases such as **Taq DNA pol**, from *Thermus acquaticus*, a thermophilic bacterium that lives in hot springs. Taq, and similar thermostable polymerases from other hot environments, are able to remain functional in the repeated cycles of amplification. Taq polymerase cannot usually amplify fragments longer than about 3kbp, but under some specialized conditions, PCR can amplify fragments up to approximately 10kbp. Other polymerases, either by themselves or in combination with Taq, are used to increase the length of amplified fragments or to increase the fidelity of the replication.

After completion of the thermalcycling (amplification), an aliquot from the PCR reaction is usually loaded onto an **electrophoretic gel** (described below) to determine whether a DNA fragment of the expected length was successfully amplified or not. Usually, the original template DNA will be so dilute that it will not be visible on the gel, only the amplified PCR product. The presence of a sharp band of the expected length indicates that PCR was able to amplify its target. If the purpose of the PCR was to test for the presence of a particular template sequence, this is the end of the experiment. Otherwise, the remaining PCR product can be used as starting material for a variety of other techniques such as sequencing or cloning.

An Application of PCR: the StarLink Affair

PCR is very sensitive (meaning it can amplify very small starting amounts of DNA), and specific (meaning it can amplify only the target sequence from a mixture of many DNA sequences). This made PCR the perfect tool to test whether genetically modified corn was present in consumer products on supermarket shelves. Although currently (2013) 85% of corn in the United States is genetically modified, and contains genes that government regulators have approved for human consumption, back in 2000, environmental groups showed that a strain of genetically modified corn, which had only been approved for use as animal feed, had been mixed in with corn used in producing human food, like taco shells.

To do this, the groups purchased taco shells from stores in the Washington DC area, extracted DNA from the taco shells and used it as a template in a PCR reaction with primers specific to the unauthorized gene (*Cry9C*). Their suspicions were confirmed when they ran this PCR product on an agarose gel and saw a band of expected size. The PCR test was able to detect one transgenic kernel in a whole bushel of corn (1 per 100,000). The company (Aventis) that sold the transgenic seed to farmers had to pay for the destruction of large amounts of corn, and was the target of a class action law-suit by angry consumers who claimed they had been made sick by the taco shells. While no legitimate cases of harm were ever proven, and the plaintiffs were awarded $9 million, of which $3 million went to the legal fees, and the remainder of the judgment went to the consumers in the form of coupons for taco shells. The affair damaged the company, and exposed a weakness in the way the genetically modified crops were handled in the United States at the time.

This page titled 12.3: Isolating or Detecting a Specific Sequence by PCR is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.3: Isolating or Detecting a Specific Sequence by PCR** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation

Restriction Enzymes

Many bacteria have enzymes that recognize specific DNA sequences (usually 4 or 6 nucleotides) and then cut the double stranded DNA helix at this sequence (Figure 12.4.7). These enzymes are called site-specific **restriction endonucleases**, or more simply "**restriction enzymes**", and they naturally function as part of bacterial defenses against viruses and other sources of foreign DNA. To cut DNA at known locations, researchers use restriction enzymes that have been purified from various bacterial species, and which can be purchased from various commercial sources. These enzymes are usually named after the bacterium from which they were first isolated. For example, ***EcoRI*** and *EcoRV* are both enzymes from *E. coli*. *EcoRI* cuts double stranded DNA at the sequence GAATTC, but note that this enzyme, like many others, does not cut in exactly the middle of the restriction sequence (Figure 12.4.8). The ends of a molecule cut by *EcoRI* have an overhanging region of single stranded DNA, and so are sometimes called **sticky-ends**. On the other hand, *EcoRV* is an example of an enzyme that cuts both strands in exactly the middle of its recognition sequence, producing what are called **blunt-ends**, which lack overhangs.



Figure 12.4.7: An EcoRI dimer (blue, purple) sits like a saddle on a double helix of DNA (one strand is green, one is brown). This image is looking down the center of the helix.(NCBI-?-PD)



Figure 12.4.8: The recognition sequence for EcoRI (blue) is cleaved by the enzyme (grey). This particular enzyme cuts DNA at a position offset from the center of the restriction site. This creates an overhanging, sticky-end. (Original-Deyholos-CC:AN)

DNA Ligation

The process of **DNA ligation** occurs when DNA strands are covalently joined, end-to-end through the action of an enzyme called **DNA ligase**. Sticky-ended molecules with complementary overhanging sequences are said to have **compatible ends**, which facilitate their joining to form recombinant DNA. Likewise, two blunt-ended sequences are also considered compatible to join together, although they do not ligate together as efficiently as sticky-ends. Note: sticky-ended molecules with non-complementary sequences will not ligate together with DNA ligase. Ligation is therefore central to the production of recombinant DNA, including the insertion of a double stranded DNA fragment into a plasmid vector.

This page titled 12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.5: Cloning DNA - Plasmid Vectors

Plasmids are Naturally Present in Some Bacteria

Many bacteria contain extra-chromosomal DNA elements called **plasmids**. These are usually small (a few 1000 bp), circular, double stranded molecules that replicate independently of the chromosome and can be present in high copy numbers within a cell. In the wild, plasmids can be transferred between individuals during bacterial mating and are sometimes even transferred between different species. Plasmids are particularly important in medicine because they often carry genes for pathogenicity and drug-resistance. In the lab, plasmids can be inserted into bacteria in a process called transformation.

Using Plasmids as Cloning Vectors

To insert a DNA fragment into a plasmid, both the fragment and the circular plasmid are cut using a restriction enzyme that produces compatible ends (Figure 12.5.1). Given the large number of restriction enzymes that are currently available, it is usually not too difficult to find an enzyme for which corresponding recognition sequences are present in both the plasmid and the DNA fragment, particularly because most plasmid vectors used in molecular biology have been engineered to contain recognition sites for a large number of restriction endonucleases.

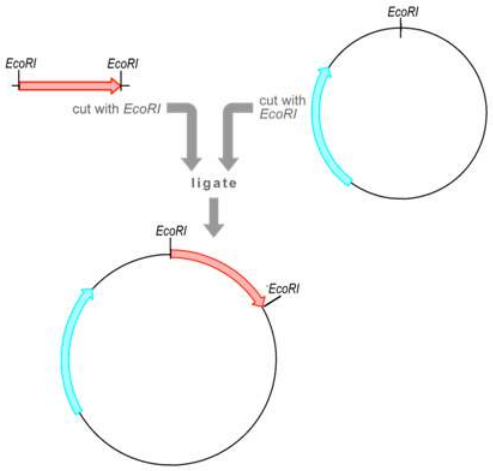


Figure 12.5.1: Cloning of a DNA fragment (red) into a plasmid vector. The vector already contains a selectable marker gene (blue) such as an antibiotic resistance gene. (Original-Deyholos-CC:AN)

After restriction digestion, the desired fragments may be further purified or selected before they are mixed together with ligase to join them together. Following a short incubation, the newly ligated plasmids, containing the gene of interest are **transformed** into *E. coli*. Transformation is accomplished by mixing the ligated DNA with *E. coli* cells that have been specially prepared (i.e. made **competent**) to uptake DNA. Competent cells can be made by exposure to compounds such as $CaCl_2$ or to electrical fields (**electroporation**). Because only a small fraction of cells that are mixed with DNA will actually be transformed, a **selectable marker**, such as a gene for antibiotic resistance, is usually also present on the plasmid. After transformation (combining DNA with competent cells), bacteria are spread on a bacterial agar plate containing an appropriate antibiotic so that only those cells that have actually incorporated the plasmid will be able to grow and form colonies. This can then be picked and used for further study.

Molecular biologists use plasmids as **vectors** to contain, amplify, transfer, and sometimes express genes of interest that are present in the cloned DNA. Often, the first step in a molecular biology experiment is to **clone** (i.e. copy) a gene into a plasmid, then transform this recombinant plasmid back into bacteria so that essentially unlimited copies of the gene (and the plasmid that carries it) can be made as the bacteria reproduce. This is a practical necessity for further manipulations of the DNA, since most techniques of molecular biology are not sensitive enough to work with just a single molecule at a time. Many molecular cloning and recombination experiments are therefore iterative processes in which:

1. a DNA fragment (usually isolated by PCR and/or restriction digestion) is cloned into a plasmid cut with a compatible restriction enzyme
2. the recombinant plasmid is transformed into bacteria

3. the bacteria are allowed to multiply, usually in liquid culture
4. a large quantity of the recombinant plasmid DNA is isolated from the bacterial culture
5. further manipulations (such as site directed mutagenesis or the introduction of another piece of DNA) are conducted on the recombinant plasmid
6. the modified plasmid is again transformed into bacteria, prior to further manipulations, or for expression

An Application of Molecular Cloning: Production of Recombinant Insulin

Purified insulin protein is critical to the treatment of diabetes. Prior to ~1980, insulin for clinical use was isolated from human cadavers or from slaughtered animals such as pigs. Human-derived insulin generally had better pharmacological properties, but was in limited supply and carried risks of disease transmission. By cloning the human insulin gene and expressing it in *E. coli*, large quantities of insulin identical to the human hormone could be produced in fermenters, safely and efficiently. Production of recombinant insulin also allows specialized variants of the protein to be produced: for example, by changing a few amino acids, longer-acting forms of the hormone can be made. The active insulin hormone contains two peptide fragments of 21 and 30 amino acids, respectively. Today, essentially all insulin is produced from recombinant sources (Figure 12.5.2), i.e. human genes and their derivatives expressed in bacteria or yeast.



Figure 12.5.2: A vial of insulin. Note that the label lists the origin as “rDNA”, which stands for recombinant DNA.(Flickr-DeathByBokeh-CC:AN)

This page titled 12.5: Cloning DNA - Plasmid Vectors is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.5: Cloning DNA - Plasmid Vectors** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.6: DNA Analysis - Gel Electrophoresis

A solution of DNA is colorless, and except for being viscous at high concentrations, is visually indistinguishable from water. Therefore, techniques such as **gel electrophoresis** have been developed to detect and analyze DNA (Figure 12.6.11).



Figure 12.6.11: Apparatus for agarose gel electrophoresis. A waterproof tank is used to pass current through a slab gel, which is submerged in a buffer in the tank. The current is supplied by an adjustable power supply. A gel (stained blue by a dye sometimes used when loading DNA on the gel) sits in a tray, awaiting further analysis, such as photography under a UV light source. (Flickr-457088634_585df11af5_o-pending)

This analysis starts when a solution of DNA is deposited at one end of a gel slab. This gel is made from polymers such as **agarose**, which is a polysaccharide isolated from seaweed. The DNA is then forced through the gel by an electrical current, with DNA molecules moving toward the positive electrode (Figure 12.6.12).

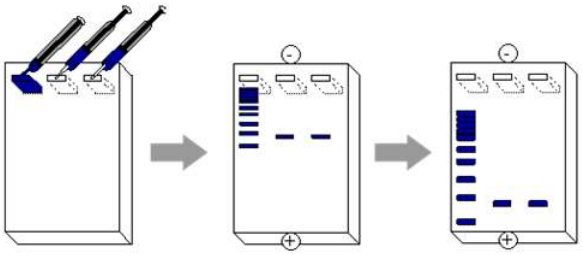


Figure 12.6.12: Agarose gel electrophoresis. DNA is loaded into wells at the top of a gel. A current is passed through the gel, pulling DNA towards the positively charged electrode. The DNA fragments are separated by size, with smaller fragments moving fastest towards the electrode. (Wikipedia-Magnus Manske_PD)

As it migrates, each piece of DNA threads its way through the pores, which form between the polymers in the gel. Because shorter pieces can move through these pores faster than longer pieces, gel electrophoresis separates molecules based on their size (length), with smaller DNA pieces moving faster than long ones. DNA molecules of a similar size migrate to a similar location in each gel, called a **band**. This feature makes it easy to see DNA after staining the DNA with a fluorescent dye such as **ethidium bromide** (Figure 12.6.13). By separating a mixture of DNA molecules of known size (**size markers**) in adjacent lanes on the same gel, the length of an uncharacterized DNA fragment can be estimated. Gel segments containing the DNA bands can also be cut out of the gel, and the size-selected DNA extracted and used in other types of reactions, such as sequencing and cloning.

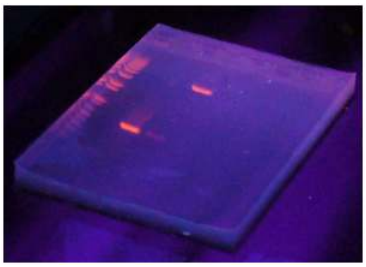


Figure 12.6.13: An agarose gel stained with ethidium bromide and illuminated by UV light. The stain associated with DNA is fluorescent. (Wikipedia-Transcontrol-GFDL)

This page titled 12.6: DNA Analysis - Gel Electrophoresis is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.6: DNA Analysis - Gel Electrophoresis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.7: DNA Analysis- Blotting and Hybridization

Bands of DNA in an electrophoretic gel form only if most of the DNA molecules are of the same size, such as following a PCR reaction, or restriction digestion of a plasmid. In other situations, such as after restriction digestion of chromosomal (genomic) DNA, there will be a large number of variable size fragments in the digest and it will appear as a continuous smear of DNA, rather than distinct bands. In this case, it is necessary to use additional techniques to detect the presence of a specific DNA sequence within the smear of DNA separated on an electrophoretic gel. This can be done using a “Southern Blot”.

Southern Blots

A **Southern blot** (also called a **Southern Transfer**) is named after Ed Southern, its inventor. In the first step, DNA is digested with restriction enzymes and separated by gel electrophoresis (as discussed above). Then a sheet or **membrane** of nylon or similar material is laid under the gel and the DNA, in its separated position (bands or smear), is transferred to the membrane by drawing the liquid out of the gel, in a process called **blotting** (Figure 12.7.1). The blotted DNA is usually covalently attached to the nylon membrane by briefly exposing the blot to UV light. Transferring the DNA to the sturdy membrane is necessary because the fragile gel would fall apart during the next two steps in the process. Next, the membrane is bathed in a solution to **denature** (double stranded made single stranded) the attached DNA. Then a **hybridization** solution containing a small amount of single-stranded **probe** DNA that is complementary in sequence to a target molecule on the membrane. This probe DNA is labeled using fluorescent or radioactive molecules, and if the hybridization is performed properly, the probe DNA will form a stable duplex only with those DNA molecules on the membrane that are exactly complementary to it. Then, the unhybridized probe is washed off and remaining radioactive or fluorescent signal will appear in a distinct band when appropriately detected. The band represents the presence of a particular DNA sequence within the mixture of DNA fragments.

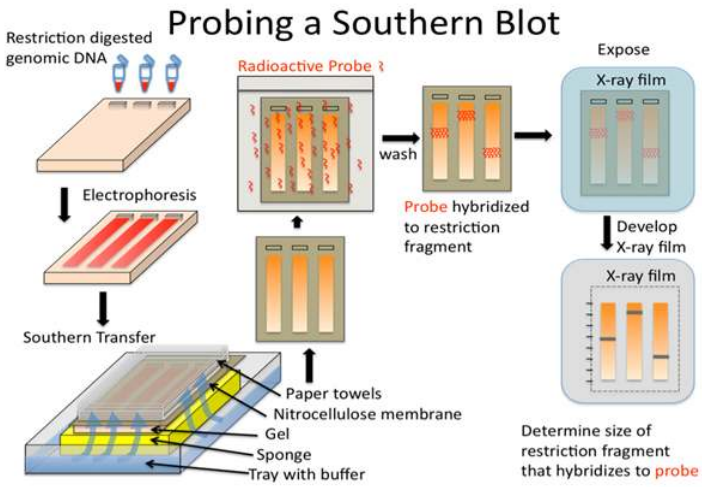


Figure 12.7.1: A diagram of Southern blotting. Genomic DNA that has been digested with a restriction enzyme is separated on an agarose gel, then the DNA is transferred from the gel to a nylon membrane (grey sheet) by blotting. The DNA is immobilized on the membrane, then probed with a radioactively labeled DNA fragment that is complementary to a target sequence. After stringent washing, the blot is exposed to X-ray film to detect what size frament the probe is bound. In this case, the probe bound to different-sized fragments in lanes 1, 2, and 3. In the last image the orange represent the position of the digested DNA, but it is not actually present on the X-ray film. (Original-J. Locke-CC:AN)

The probe is sequence specific (requires complementarity). However, variation in hybridization temperature and washing solutions can alter the **stringency** of the probe. At maximum stringency (higher temperature) hybridization conditions, probes will only hybridize with the exact target sequences that are perfectly complementary (maximum number of hydrogen bonds). At lower temperatures, probes will be able to hybridize to targets to which they do not match exactly, but only are roughly complementary for part of the sequence.

Southern blotting is useful not only for detecting the presence of a DNA sequence within a mixture of DNA molecules, but also for determining the size of a restriction fragment in a DNA sample. Southern blots are useful for detecting fragments larger than those normally amplified by PCR, and when trying to detect fragments that may be only distantly related to a known sequence. Applications of Southern blotting will be discussed further in the context of molecular markers in a subsequent chapter. Southern

blotting was invented before PCR, but PCR has replaced blotting in many applications because of its simplicity, speed, and convenience. Following the development of the Southern blot, other types of blotting techniques were invented.

Northern Blots

The **Northern blot** involves the size separation of RNA in gels like that of DNA. Because we wish to determine the native size of the RNA transcript (and because RNA is single stranded) no restriction enzymes are ever used. Because most RNA is single stranded and can fold into various conformations thorough intra-molecular base pairing, the electrophoresis separation is more haphazard and the bands are often less sharp, compared to that of double stranded DNA.

Western Blots

In a **Western blot,** protein is size separated on a gel (usually an acrylamide gel) before transferring to a membrane, which is then probed with an antibody that specifically binds to an antigenic site on the target protein. This antibody is then detected by other antibodies with some fluorescent or color production marker system. It will also give bands proportional to the amount and size of the target protein (Figure 12.7.2).

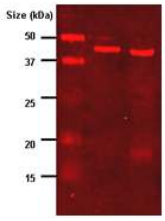


Figure 12.7.2: Western Blot using a anti-lipoic acid primary antibody and an IR-dye labelled secondary antibody in Leishmania major extracts. (Wikipedia-TimVickers–CC:AS)

A comparison of all three blotting methods is shown in Figure 12.7.3.

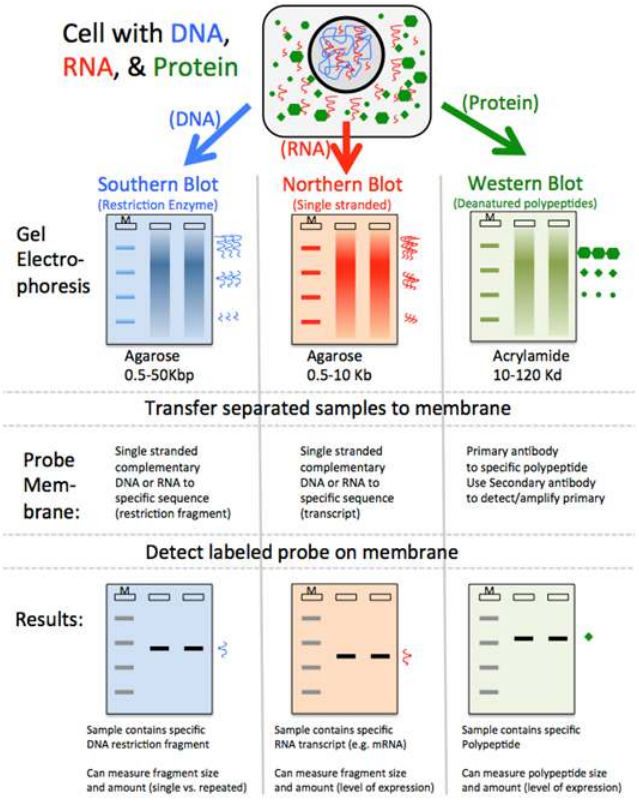


Figure 12.7.3: Comparison of Southern, Northern, and Western blots. Size and amount of DNA, RNA, and polypeptides can be determined using similar blotting methods. DNA is in blue, RNA in red, and polypeptides in green. A marker lane is shown in the left of each gel to determine size. A eukaryote cell is shown, but the same methods can be applied to prokaryotes, too. (Original-Locke-CC:AN)

This page titled 12.7: DNA Analysis- Blotting and Hybridization is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.7: DNA Analysis- Blotting and Hybridization** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.8: Transgenic organisms

General principles of transgenesis

Transgenic organisms contain foreign DNA that has been introduced using biotechnology. Foreign DNA (the **transgene**) is defined here as DNA from another species, or else recombinant DNA from the same species that has been manipulated in the laboratory then reintroduced. The terms transgenic organism and genetically modified organism (**GMO**) are generally synonymous. The process of creating transgenic organisms or cells to be come whole organisms with a permanent change to their germline has been called either **transformation** or **transfection**. (Unfortunately, both words have alternate meanings. Transformation also refers to the process of mammalian cell becoming cancerous, while transfection also refers to the process of introducing DNA into cells in culture, either bacterial or eukaryote, for a temporary use, not germ line changes.) Transgenic organisms are important research tools, and are often used when exploring a gene's function. Transgenesis is also related to the medical practice of gene therapy, in which DNA is transferred into a patient's cells to treat disease. Transgenic organisms are widespread in agriculture. Approximately 90% of canola, cotton, corn, soybean, and sugar beets grown in North America are transgenic. No other transgenic livestock or crops (except some squash, papaya, and alfalfa) are currently produced in North America.

To make a transgenic cell, DNA must first be transferred across the cell membrane, (and, if present, across the cell wall), without destroying the cell. In some cases, **naked DNA** (meaning plasmid or linear DNA that is not bound to any type of **carrier**) may be transferred into the cell by adding DNA to the medium and temporarily increasing the porosity of the membrane, for example by **electroporation**. When working with larger cells, naked DNA can also be **microinjected** into a cell using a specialized needle. Other methods use **vectors** to transport DNA across the membrane. Note that the word "vector" as used here refers to any type of carrier, and not just plasmid vectors. Vectors for transformation/transfection include **vesicles** made of lipids or other polymers that surround DNA; various types of particles that carry DNA on their surface; and infectious viruses and bacteria that naturally transfer their own DNA into a host cell, but which have been engineered to transfer any DNA molecule of interest. Usually the foreign DNA is a complete expression unit that includes its own cis-regulators (e.g. promoter) as well as the gene that is to be transcribed.

When the objective of an experiment is to produce a **stable** (i.e. heritable) transgenic eukaryote, the foreign DNA must be incorporated into the host's chromosomes. For this to occur, the foreign DNA must enter the host's nucleus, and recombine with one of the host's chromatids. In some species, the foreign DNA is inserted at a random location in a chromatid, probably wherever strand breakage and non-homologous end joining happen to occur. In other species, the foreign DNA can be targeted to a particular locus, by flanking the foreign DNA with DNA that is homologous to the host's DNA at that locus. The foreign DNA is then incorporated into the host's chromosomes through homologous recombination.

Furthermore, to produce multicellular organisms in which all cells are transgenic and the transgene is stably inherited, the cell that was originally transformed must be either a gamete or must develop into tissues that produce gametes. Transgenic gametes can eventually be mated to produce homozygous, transgenic offspring. The presence of the transgene in the offspring is typically confirmed using PCR or Southern blotting, and the expression of the transgene can be measured using reverse-transcription PCR (RT-PCR), RNA blotting, and Western (protein blotting).

The rate of transcription of a transgene is highly dependent on the state of the chromatin into which it is inserted (i.e. **position effects**), as well as other factors. Therefore, researchers often generate several independently transformed/transfected lines with the same transgene, and then screen for the lines with the highest expression. It is also good practice to clone and sequence the transgenic locus from a newly generated transgenic organism, since errors (truncations, rearrangements, and other mutations) can be introduced during transformation/transfection.

Producing a transgenic plant

The most common method for producing transgenic plants is **Agrobacterium-mediated transformation** (Figure 12.8.1). *Agrobacterium tumifaciens* is a soil bacterium that, as part of its natural pathogenesis, injects its own tumor-inducing (**T_i**) plasmid into cells of a host plant. The natural T_i plasmid encodes growth-promoting genes that cause a gall (i.e. tumor) to form on the plant, which also provides an environment for the pathogen to proliferate. Molecular biologists have engineered the T_i plasmid by removing the tumor-inducing genes and adding restriction sites that make it convenient to insert any DNA of interest. This engineered version is called a **T-DNA** (transfer-DNA) plasmid; the bacterium transfers a linear fragment of this plasmid that

includes the conserved “left-border (LB)”, and right-border (RB)” DNA sequences, and anything in between them (up to about 10 kb). The linear T-DNA fragment is transported into the nucleus, where it recombines with the host-DNA, probably wherever random breakages occur in the host’s chromosomes.

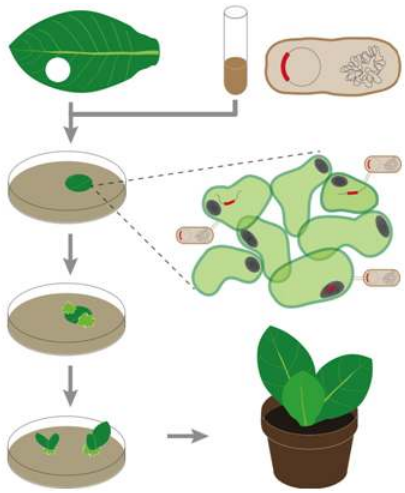


Figure 12.8.1: Production of a transgenic plant using Agrobacterium-mediated transformation. The bacterium has been transformed with a T-DNA plasmid that contains the transgene and a selectable marker that confers resistance to a herbicide or antibiotic. A bacterial culture and plant tissue (e.g. a leaf punch) are co-cultured on growth medium in a Petri dish. Some of the plant cells will become infected by the bacterium, which will transfer the T-DNA into the plant cytoplasm. In some cases the transgene will become integrated into the chromosomal DNA of a plant cell. In the presence of certain combinations of hormones, the plant cells will dedifferentiate into a mass of cells called callus. The presence of a selective agent (e.g. herbicide or antibiotic) in the growth medium prevents untransformed cells from dividing. Therefore, each callus ideally consists only of transgenic plant cells. The resistant calli are transferred to media with other combinations of hormones that promote organogenesis, i.e. differentiation of callus cells into shoots and then roots. The regenerated transgenic plants are transferred to soil. Their seeds can be harvested and tested to ensure that the transgene is stably inherited. (Original-Deyholos-CC:AN)

In Arabidopsis and a few other species, flowers can simply be dipped in a suspension of Agrobacterium, and ~1% of the resulting seeds will be transformed. In most other plant species, cells are induced by hormones to form a mass of undifferentiated tissues called a callus. The Agrobacterium is applied to a callus and a few cells are transformed, which can then be induced by other hormones to regenerate whole plants (Figure 12.8.2). Some plant species are resistant (i.e. “**recalcitrant**”) to transformation by Agrobacterium. In these situations, other techniques must be used such as **particle bombardment**, whereby DNA is non-covalently attached to small metallic particles, which are accelerated by compressed air into callus tissue, from which complete transgenic plants can sometimes be regenerated. In all transformation methods, the presence of a selectable marker (e.g. a gene that confers antibiotic resistance or herbicide resistance) is useful for distinguishing transgenic cells from non-transgenic cells at an early stage of the transformation process.

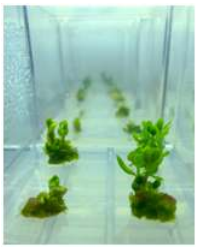


Figure 12.8.2: Organogenesis of flax shoots from calli. (Original-J. McDill-CC:AN)

Producing a transgenic mouse

In a commonly used method for producing a transgenic mouse, **stem cells** are removed from a mouse embryo, and a transgenic DNA construct is transferred into the stem cells using electroporation, and some of this transgenic DNA enters the nucleus, where it may undergo homologous recombination (Figure 12.8.3). The transgenic DNA construct contains DNA homologous to either side of a locus that is to be targeted for replacement. If the objective of the experiment is simply to delete (“**knock-out**”) the targeted locus, the host’s DNA can simply be replaced by selectable marker, as shown. It is also possible to replace the host’s DNA at this

locus with a different version of the same gene, or a completely different gene, depending on how the transgenic construct is made. Cells that have been transfected and express the selectable marker (i.e. resistance to the antibiotic neomycin resistance, neoR, in this example) are distinguished from unsuccessfully transfected cells by their ability to survive in the presence of the selective agent (e.g. an antibiotic). Transfected cells are then injected into early stage embryos, and then are transferred to a foster mother. The resulting pups are chimeras, meaning that only some of their cells are transgenic. Some of the chimeras will produce gametes that are transgenic, which when mated with a wild-type gamete, will produce mice that are hemizygous for the transgene. Unlike the chimeras, these hemizygotes carry the transgene in all of their cells. Through further breeding, mice that are homozygous for the transgene can be obtained.

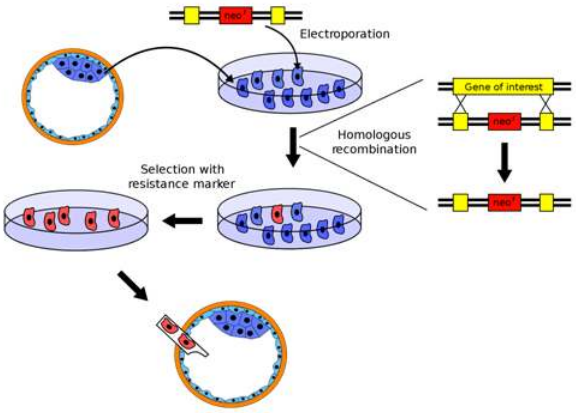


Figure 12.8.3: Production of a transgenic mouse. Stem cells are removed from an embryo, and are transfected (using electroporation) with a transgenic construct that bears a neomycin resistance gene (neo^r) flanked by two segments of DNA homologous to a gene of interest. In the nucleus of a transgenic cell, some of the foreign DNA will recombine with the targeted gene, disrupting the targeted gene and introducing the selectable marker. Only cells in which neo^r has been incorporated will survive selection. These neomycin resistant cells are then transplanted into another embryo, which will grow into a chimera within a foster mother. (Wikipedia-Kiaergaard-CC:AN)

Human gene therapy

Many different strategies for human gene therapy are under development. In theory, either the **germline** or **somatic** cells may be targeted for transfection, but most research has focused on somatic cell transfection, because of risks and ethical issues associated with germline transformation. Gene therapy approaches may be further classified as either ***ex vivo*** or ***in vivo***, with the former meaning that cells (e.g. stem cells) are transfected in isolation before being introduced to the body, where they replace defective cells. *Ex vivo* gene therapies for several blood disorders (e.g. immunodeficiencies, thalassemias) are undergoing clinical trials. For *in vivo* therapies, the transfection occurs within the patient. The objective may be either stable integration, or **non-integrative** transfection. As described above, stable transfection involves integration into the host genome. In the clinical context, stable integration may not be necessary, and carries with it higher risk of inducing mutations in either the transgene or host genome). In contrast, transient transfection does not involve integration into the host genome and the transgene may therefore be delivered to the cell as either RNA or DNA. Advantages of RNA delivery include that no promoter is needed to drive expression of the transgene. Besides mRNA transgenes, which could provide a functional version of a mutant protein, there is great interest in delivery of **siRNA** (small-inhibitory RNAs), which can be used to silence specific genes in the host cell's genome.

Vectors for *in vivo* gene therapy must be capable of delivering DNA or RNA to a large proportion of the targeted cells, without inducing a significant immune response, or having any toxic effects. Ideally, the vectors should also have high specificity for the targeted cell type. Vectors based on viruses (e.g. **lentiviruses**) are being developed for in both *in vivo* and *ex vivo* gene therapies. Other, non-viral vectors (e.g. vesicles and nanoparticles) are also being developed for gene therapy as well.

This page titled 12.8: Transgenic organisms is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.8: Transgenic organisms** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.9: Techniques of Molecular Genetics (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions:

8.1 What information, and what reagents would you need to use PCR to detect HIV in a blood sample?

8.2 A 6.0 kbp PCR fragment flanked by recognition sites for the *HindIII* restriction enzyme is cut with *HindIII* then ligated with a 3kb plasmid vector that has also been cut with *HindIII*. This recombinant plasmid is transformed into *E. coli*. From one colony a plasmid is prepared and digested with HindIII.

a) When the product of the *HindIII* digestion is analyzed by gel electrophoresis, what will be the size of the bands observed?

b) What bands would be observed if the recombinant plasmid was cut with *EcoRI*, which has only one site, directly in the middle of the PCR fragment?

c) What bands would be observed if the recombinant plasmid was cut with both *EcoRI* and *HindIII* at the same time?

8.3 If you started with 10 molecules of double stranded DNA template, what is the maximum number of molecules you would you have after 10 PCR cycles?

8.4 What is present in a PCR tube at the end of a successful amplification reaction? With this in mind, why do you usually only see a single, sharp band on a gel when it is analyzed by electrophoresis?

8.5 A coat protein from a particular virus can be used to immunize children against further infection. However, inoculation of children with proteins extracted from natural viruses sometimes causes fatal disease, due to contamination with live viruses. How could you use molecular biology to produce an optimal vaccine?

8.6 How would cloning be different if there were no selectable markers?

8.7 Research shows that a particular form of cancer is caused by a 200bp deletion in a particular human gene that is normally 2kb long. Only one mutant copy is needed to cause the disease.

a) Explain how you would use Southern blotting to diagnose the disease.

b) How would any of the blots appear if you hybridized and washed at very low temperature?

8.8 Refer to question 8.7.

a) Explain how you would detect the presence of the same deletion using PCR, rather than a Southern blot.

b) How would PCR products appear if you annealed at very low temperature?

8.9 You have a PCR fragment for a human olfactory receptor gene (perception of smells). You want to know what genes a dog might have that are related to this human gene. How can you use your PCR fragment and genomic DNA from a dog to find this out? Do you think dogs have more or less of these genes?

8.10 You add ligase to a reaction containing a sticky-ended plasmid and sticky-ended insert fragment, which both have compatible ends. Unbeknownst to you, someone in the lab left the stock of ligase enzyme out of the freezer overnight and it degraded (no longer works). Explain in detail what will happen in your ligation experiment in this situation should you try and transform with it.

This page titled 12.9: Techniques of Molecular Genetics (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

12.S: Techniques of Molecular Genetics (Summary)

- Molecular biology involves the isolation and analysis of DNA and other macromolecules
- Isolation of total genomic DNA involves separating DNA from protein and other cellular components, for example by ethanol precipitation of DNA.
- PCR can be used as part of a sensitive method to detect the presence of a particular DNA sequence
- PCR can also be used as part of a method to isolate and prepare large quantities of a particular DNA sequence
- Restriction enzymes are natural endonucleases used in molecular biology to cut DNA sequences at specific sites.
- DNA fragments with compatible ends can be joined together through ligation.If the ligation produces a sequence not found in nature, the molecule is said to be recombinant.
- Transformation is the introduction of DNA (usually recombinant plasmids) into bacteria.
- Cloning of genes in *E. coli* is a common technique in molecular biology, since it allows large quantities of a DNA for gene to made, which allows further analysis or manipulation
- Cloning can also be used to produce useful proteins, such as insulin, in microbes.
- Southern blotting can be used to detect the presence of any sequence that matches a probe, within a mixture of DNA (such as total genomic DNA).
- The stringency of hybridization in blotting and in PCR is dependent on physical parameters such as temperature and washing solution content.

Key Terms:

macromolecules	ligation	
lysis	ligase	
detergent	plasmid	GMO
chelating agent	transformation	transformation
EDTA	competent	transfection
nuclease	electroporation	naked DNA
pellet	selectable marker	vector
PCR	agarose	electroporation
primer	vector	microinjection
thermalcycle	band	vesicles
denature	ethidium bromide	callus
anneal	Southern blot	knock-out
extend	membrane	recalcitrant
Taq DNApol	denaturation	Agrobacterium
electrophoretic gel	hybridization	particle bombardment
restriction endonuclease	washing	T-DNA
EcoRI	probe	Ti plasmid
sticky end	stringency	position effects
blunt end	northern blot	stem cells
compatible end	western blot	
	transgene	

This page titled 12.S: Techniques of Molecular Genetics (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **8.S: Techniques of Molecular Genetics (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

13: Cancer Genetics

Cancer is a group of diseases that exhibit uncontrolled growth, invasion of adjacent tissues, and sometimes **metastasis** (the movement of cancer cells through the blood or lymph). In cancer cells, the regulatory mechanisms that control cell division and limit abnormal growth have been disrupted, usually by the accumulation of several mutations. Cancer is therefore essentially a genetic disease. Although some cancer--related mutations may be heritable, most cancers are sporadic, meaning they arise from new mutations that occur in the individual who has the disease. In this chapter we will examine the connection between cancer and genes.

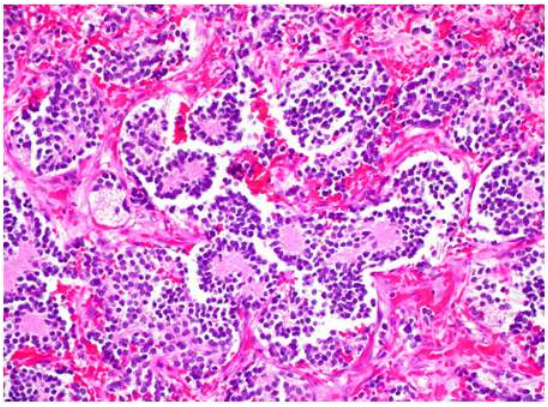


Figure 13.1: *Stained histological section of a neuroblastoma in an adrenal gland. (Flickr-Ed Uthman-CC:AS)*

13.1: Classification of Cancers
13.2: Cancer Cell Biology
13.3: Hallmarks of Cancer
13.4: Mutagens and Carcinogens
13.5: Oncogenes
13.6: Tumor Suppressor Genes
13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
13.8: Genes and COVID-19 Susceptibility in Humans
13.8.1: Introduction
13.8.2: How Does SARS-Cov-2 Infect Humans?
13.8.3: Genes Implicated in Severe COVID-19 Infection in Humans
13.8.4: Approaches for Vaccine Development Against SARS-Cov-2
13.8.5: Chapter 13 Summary
13.8.6: Key Terms in Chapter 13
13.E: Cancer Genetics (Exercises)
13.S: Cancer Genetics (Summary)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 13: Cancer Genetics is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

13.1: Classification of Cancers

Cancers can be classified based on the tissues in which they originate. **Sarcomas** are cancers that originate in mesoderm tissues, such as bone or muscle, and cancers arising in glandular tissues (e.g. breast, prostate) are classified as **adenocarcinomas**. **Carcinomas** originate in epithelial cells (both inside the body and on its surface) and are the most common types of cancer (~85%). Each of these classifications may be further sub--divided. For example, **squamous cell carcinoma (SCC)**, **basal cell carcinoma (BCC)**, and **melanoma** are all types of skin cancers originating respectively in the squamous cells, basal cells, or melanocytes of the skin.

This page titled 13.1: Classification of Cancers is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.1: Classification of Cancers** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

13.2: Cancer Cell Biology

Cancer is a progressive disease that usually begins with increased frequency of cell division (Figure 13.2.2). Under the microscope, this may be detectable as increased cellular and nuclear size, and an increased proportion of cells undergoing mitosis. As the disease progresses, cells typically lose their normal shape and tissue organization. Tissues with increased cell division and abnormal tissue organization exhibit **dysplasia**. Eventually a tumor develops, which can grow rapidly and expand into adjacent tissues.

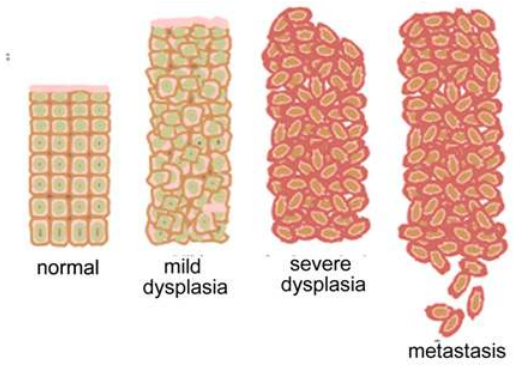


Figure 13.2.2: Progressive increases in cell division and abnormal cell morphology associated with cancer. (Wikipedia-NIH-PD)

As cellular damage accumulates and additional control mechanisms are lost, some cells may break free of the primary tumor, pass into the blood or lymph system, and be transported to another organ, where they develop into new tumors (Figure 13.2.3). The early detection of tumors is important so that they can be treated or removed before the onset of metastasis, but note that not all tumors will lead to cancer. Tumors that do not metastasize are classified as **benign,** and are not usually considered life threatening. In contrast, **malignant** tumors become invasive, and ultimately result in cancer.



Figure 13.2.3: Secondary tumors (white) develop in the liver from cells of a metastatic pancreatic cancer. (Wikipedia-J. Hayman-PD)

This page titled 13.2: Cancer Cell Biology is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.2: Cancer Cell Biology** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

13.3: Hallmarks of Cancer

Researchers have identified six molecular and cellular traits that characterize most cancers. These six hallmarks of cancer are summarized in Table 13.3.1. In this chapter, we will focus on the first two hallmarks, namely growth signal autonomy and insensitivity to anti--growth signals.

Table 13.3.1 *Ten Hallmarks of Cancer (Hanahan and Weinberg, 2000; Hanahan 2011)*

1. Growth signal autonomy Cancer cells can divide without the external signals normally required to stimulate division.
2. Insensitivity to growth inhibitory signals Cancer cells are unaffected by external signals that inhibit division of normal cells.
3. Evasion of apoptosis When excessive DNA damage and other abnormalities are detected, apoptosis (a type of programmed cell death) is induced in normal cells, but not in cancer cells.
4. Reproductive potential not limited by telomeres Each division of a normal cell reduces the length of its telomeres. Normal cells arrest further division once telomeres reach a certain length. Cancer cells avoid this arrest and/or maintain the length of their telomeres.
5. Sustained angiogenesis Most cancers require the growth of new blood vessels into the tumor. Normal angiogenesis is regulated by both inhibitory and stimulatory signals not required in cancer cells.
6. Tissue invasion and metastasis Normal cells generally do not migrate (except in embryo development). Cancer cells invade other tissues including vital organs.
7. Deregulated metabolic pathways Cancer cells use an abnormal metabolism to satisfy a high demand for energy and nutrients.
8. Evasion of the immune system Cancer cells are able to evade the immune system.
9. Chromosomal instability Severe chromosomal abnormalities are found in most cancers.
10. Inflammation Local chronic inflammation is associated with many types of cancer.

This page titled 13.3: Hallmarks of Cancer is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.3: Hallmarks of Cancer** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

13.4: Mutagens and Carcinogens

A **carcinogen** is any agent that directly increases the incidence of cancer. Most, but not all carcinogens are mutagens. Carcinogens that do not directly damage DNA include substances that accelerate cell division, thereby leaving less opportunity for cell to repair induced mutations, or errors in replication. Carcinogens that act as mutagens may be biological, physical, or chemical in nature, although the term is most often used in relation to chemical substances.

Human Papilloma Virus (**HPV**, Figure 13.4.4) is an example of a biological carcinogen. Almost all cervical cancers begin with infection by HPV, which contains genes that disrupt the normal pattern of cell division within the host cell. Any gene that leads to an uncontrolled increase in cell division is called an **oncogene**. The HPV E6 and E7 genes are considered oncogenes because they inhibit the host cell's natural tumor suppressing proteins (include p53, described below). The product of the E5 gene mimics the host's own signals for cell division, and these and other viral gene products may contribute to dysplasia, which is detected during a Pap smear (Figure 13.4.5). Detection of abnormal cell morphology in a Pap smear is not necessarily evidence of cancer. It must be emphasized again that cells have many regulatory mechanisms to limit division and growth, and for cancer to occur, each of these mechanisms must be disrupted. This is one reason why only a minority of individuals with HPV infections ultimately develop cancer. Although most HPV-related cancers are cervical, HPV infection can also lead to cancer in other tissues, in both women and men.

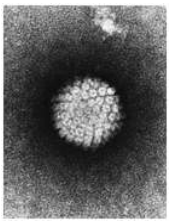


Figure 13.4.4***:*** *Electron micrograph of HPV. (Wikipedia-Unknown-PD)*

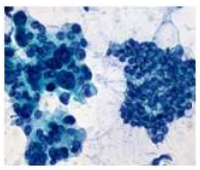


Figure 13.4.5***:*** *Dysplastic (left) and normal (right) cells from a Pap smear. (Flickr-Ed Uthman-CC:AS)*

Radiation is a well-known physical carcinogen, because of its potential to induce DNA damage within the body. The most damaging type of radiation is **ionizing**, meaning waves or particles with sufficient energy to strip electrons from the molecules they encounter, including DNA or molecules that can subsequently react with DNA. Ionizing radiation, which includes x-rays, gamma rays, and some wavelengths of ultraviolet rays, is distinct from the non-ionizing radiation of microwave ovens, cell phones, and radios. As with other carcinogens, mutation of multiple, independent genes that normally regulate cell division is required before cancer develops.

Chemical carcinogens (Table 13.4.2) can be either natural or synthetic compounds that, based on animal feeding trials or **epidemiological** (i.e. human population) studies, increase the incidence of cancer. The definition of a chemical as a carcinogen is problematic for several reasons. Some chemicals become carcinogenic only after they are metabolized into another compound in the body; not all species or individuals may metabolize chemicals in the same way. Also, the carcinogenic properties of a compound are usually dependent on its dose. It can be difficult to define a relevant dose for both lab animals and humans.

Nevertheless, when a correlation between cancer incidence and chemical exposure is observed, it is usually possible to find ways to reduce exposure to that chemical.

Table 13.4.2***:*** *Some classes of chemical carcinogens (Pecorino 2008)*

1. PAHs (polycyclic aromatic hydrocarbons)
e.g. benzo[a]pyrene and several other components of the smoke of cigarettes, wood, and fossil fuels
2. Aromatic amines
e.g. formed in food when meat (including fish, poultry) are cooked at high temperature
3. Nitrosamines and nitrosamides
e.g. found in tobacco and in some smoked meat and fish
4. Azo dyes
e.g. various dyes and pigments used in textiles, leather, paints.
5. Carbamates
e.g. ethyl carbamate (urethane) found in some distilled beverages and fermented foods
6. Halogenated compounds
e.g. pentachlorophenol used in some wood preservatives and pesticides.
7. Inorganic compounds
e.g. asbestos; may induce chronic inflammation and reactive oxygen species
8. Miscellaneous compounds
e.g. alkylating agents, phenolics

This page titled 13.4: Mutagens and Carcinogens is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.4: Mutagens and Carcinogens** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

13.5: Oncogenes

The control of cell division involves many different genes. Some of these genes act as signaling molecules to activate normal progression through the cell cycle. One of the pre-requisites for cancer occurs when one or more of these activators of cell division become mutated.

The mutation may involve a change in the coding sequence of the protein, so that it is more active than normal, or a change in the regulation of its expression, so that it is produced at higher levels than normal, or persists in the cell longer than normal. Genes that are a part of the normal regulation of cell division, but which after mutation contribute to cancer, are called **proto-oncogenes**. Once a proto-oncogene has been abnormally activated by mutation, it is called an oncogene. More than 100 genes have been defined as proto-oncogenes. These include genes at almost every step of the signaling pathways that normally induce cell to divide, including growth factors, **receptors**, **signal transducers**, and transcription factors.

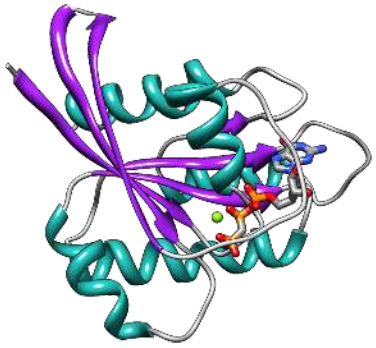


Figure 13.5.6: HRas structure PDB 121p, ribbon showing strands in purple, helices in aqua, loops in gray. Also shown are the bound GTP analog and magnesium ion. (CC-BY-SA ElaineMeng)

ras is an example of a proto-oncogene. *ras* acts as a switch within signal transduction pathways, including the regulation of cell division. When a receptor protein receives a signal for cell division, the receptor activates *ras*, which in turn activates other signaling components, ultimately leading to activation of genes involved in cell division. Certain mutations of the *ras* sequence causes it to be in a permanently active form, which can lead to constitutive activation of the cell cycle. This mutation is dominant as are most oncogenes. An example of the role of ras in relaying a signal for cell division in the EGF pathway is shown in Figure 13.5.7.

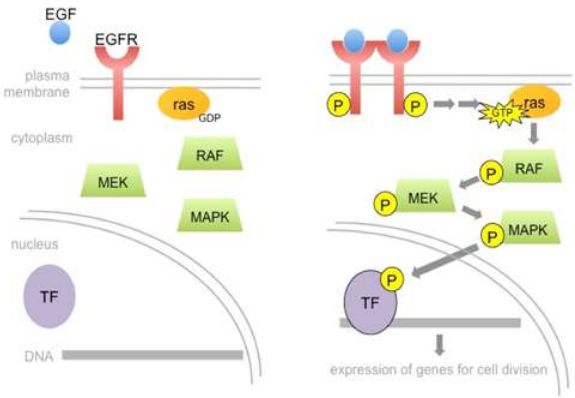


Figure 13.5.7: Simplified representation of the epidermal growth factor (EGF) signaling pathway. In the panel on the left, the components are shown in their inactive forms, prior to stimulation of the pathway. The components include the soluble ligand, EGF, its receptor (EGFR, a tyrosine kinase), ras (a G protein), several kinases (RAF, MEK, MAPK), and a transcription factor (TF). In the right panel, the activate pathway is shown. Binding of the ligand to its receptor leads to autophosphorylation of the receptor. Through a series of proteins not shown here, the phosphorpylated simulates conversion of ras to its active, GTP-bound form. The activated ras then stimulates phosphorylation of a series of kinases, which ultimately activate transcription factors and the expression of genes required for cell proliferation. (Original-Deyholos-CC:AN)

This page titled 13.5: Oncogenes is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.5: Oncogenes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

13.6: Tumor Suppressor Genes

More than 30 genes are classified as **tumor suppressors**. The normal functions of these genes include repair of DNA, induction of programmed cell death (**apoptosis**) and prevention of abnormal cell division. In contrast to proto-oncogenes, in tumor suppressors it is loss-of-function mutations that contribute to the progression of cancer. This means that tumor suppressor mutations tend to be recessive, and thus both alleles must be mutated in order to allow abnormal growth to proceed. It is perhaps not surprising that mutations in tumor suppressor genes, are more likely than oncogenes to be inherited. An example is the tumor suppressor gene, ***BRCA1***, which is involved in DNA-repair. Inherited mutations in *BRCA1* increase a woman's lifetime risk of breast cancer by up to seven times, although these heritable mutations account for only about 10% of breast cancer. Thus, sporadic rather than inherited mutations are the most common sources of both oncogenes and disabled tumor suppressor genes.

An important tumor suppressor gene is a transcription factor named **p53**. Other proteins in the cell sense DNA damage, or abnormalities in the cell cycle and activate p53 through several mechanisms including **phosphorylation** (attachment of phosphate to specific site on the protein) and transport into the nucleus. In its active form, p53 induces the transcription of genes with several different types of tumor suppressing functions, including DNA repair, cell cycle arrest, and apoptosis. Over 50% of human tumors contain mutations in p53. People who inherit only one function copy of p53 have a greatly increased incidence of early onset cancer. However, as with the other cancer related genes we have discussed, most mutations in p53 are sporadic, rather than inherited. Mutation of p53, through formation of pyrimidine dimers in the genes following exposure to UV light, has been causally linked to squamous cell and basal cell carcinomas (but not melanomas, highlighting the variety and complexities of mechanisms that can cause cancer).

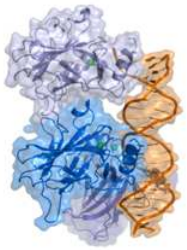


Figure 13.6.8: p53 bound to its target site on a DNA molecule. (Wikipedia-Thomas Spettstoesser from Cho etal, Science 265PP346, 1994-CC:AS)

This page titled 13.6: Tumor Suppressor Genes is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.6: Tumor Suppressor Genes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)

Chronic myelogenous leukemia (CML)

Chronic myelogenous leukemia (**CML**) is a type of cancer of white blood cells, myeloid cells, that are mutated and proliferate uncontrollably through three stages (chronic, accelerated, and blast crisis) and lead eventually to death. Cytogenetics showed the myeloid cells of CML patients usually also have a consistent chromosome translocation (the mutant event) between the long arms of chromosomes 9 and 22, t(9:22)(q34;q11). It is also known as the **Philadelphia chromosome** (Ph^+). This translocation involves breaks in two genes, *c-abl* and *bcr*, on chromosomes 9 and 22, respectively. The fusion of the translocation breaks result in a chimeric gene, called ***bcr-abl***, that contains exons 1 and/or 2 from *bcr* (this varies from patient to patient) and 2-11 from *abl* and it produces a chimeric protein (BCR-ABL or $p185^{bcr\text{-}abl}$) that is transcribed like *bcr* and contains *abl* enzyme sequences. This chimeric protein has a tyrosine-kinase from the *abl* gene sequences that is unique to the CML mutant cell. The consistent, unregulated expression of this gene and its kinase product causes activation of a variety of intracellular signaling pathways, promoting the uncontrolled proliferative and survival properties of CML cells (the cancer). Thus the BCR-ABL tyrosine kinase enzyme exists only in cancer cells (and not in healthy cells) and a drug that inhibits this activity could be used to target and prevent the uncontrolled growth of the cancerous CML cells.

Inhibiting the Bcr-Abl tyrosine kinase activity

Knowing that the kinase activity was the key to treatment, pharmaceutical companies screened chemical libraries of potential kinase inhibitory compounds. After initially finding low potency inhibitors, a relationship between structure and activity suggested other compounds that were optimized to inhibit the BCR-ABL tyrosine kinase activity. The lead compound was STI571, now called **Gleevec**™ or imatinib (Figure 13.7.9). This drug was shown to inhibit the BCR-ABL tyrosine kinase activity and to inhibit CML cell proliferation *in vitro* and *in vivo*. Gleevec™ works via targeted therapy—only the kinase activity in cancer cells was targeted and thereby killed through the drug's action. In this regard, Gleevec™ was one of the first cancer therapies to show the potential for this type of targeted action. It was dependent upon the genetic identification of the cause and protein target and is often cited as a paradigm for genetic research in cancer therapeutics.

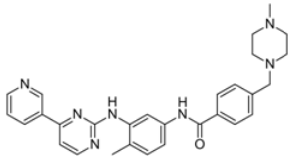


Figure 13.7.9***:*** *Biochemical structure of Gleevec™ or Imatinib. . (Wikipedia-Fuse809-CC:AN)*

Caution

This is a simplified presentation of the CML/cancer targeting by the drug Gleevec™. There are many more details than could be presented here. It is represents as a model of finding a drug for each type of cancer, rather than the one, single “magic bullet” that kills all cancers. Remember, there are always complexities in this type of research to treatment process, such as patient genetic and environmental variation that leads to differences in drug metabolism, uptake, and binding. Also, changes in drug dose, mutation of the *bcr-abl* gene, and other events can affect the effectiveness of the treatment and the relapse rate. Biological systems are extremely complex and difficult to modulate in the specific, targeted manner necessary to treat cancer ideally.

Remember, the drug, Gleevec™, is not a cure, but only a treatment. It prevents the uncontrolled proliferation of the CML cells, but doesn’t kill them directly. The arrested cells will die eventually, but there is always a small pool of CML cells that will proliferate if the drug is discontinued. While sustained use of this expensive drug is beneficial to the pharmaceutical companies, it is certainly not the ideal situation for the patient.

This page titled 13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.7: The "Poster Boy" of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

SECTION OVERVIEW

13.8: Genes and COVID-19 Susceptibility in Humans

13.8.1: Introduction

13.8.2: How Does SARS-Cov-2 Infect Humans?

13.8.3: Genes Implicated in Severe COVID-19 Infection in Humans

13.8.4: Approaches for Vaccine Development Against SARS-Cov-2

13.8.5: Chapter 13 Summary

13.8.6: Key Terms in Chapter 13

This page titled 13.8: Genes and COVID-19 Susceptibility in Humans is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

13.8.1: Introduction

Learning Objectives

- Describe the basic mechanism by which SARS-Cov-2 infects humans.
- Identify some of the genes implicated in severe COVID-19 infection in humans.
- Describe the various approaches for vaccine development against SARS-Cov-2.

Coronaviruses are a group of RNA viruses (i.e., their genetic material is RNA, rather than DNA) which cause diseases in mammals and birds, such as respiratory tract infections, which generally range from mild to lethal. Mild illnesses in humans include some cases of the common cold (which is also caused by other viruses, such as rhinoviruses), while more lethal varieties can cause SARS, MERS, and COVID-19. The most recent common ancestor of all coronaviruses is estimated to have existed as recently as 8000 BCE, although some models place the common ancestor as far back as 55 million years or more, implying long term co-evolution with bat and avian species.

The Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 – **Figure 13.1.1**). The structural proteins of SARS-CoV-2 include membrane glycoprotein (M), envelope protein (E), nucleocapsid protein (N), and the spike protein (S). The S-protein is the viral component that attaches to the host receptor via the ACE2 receptors, which is an enzyme on the surface of many cell types which generates small proteins by cleaving the large protein angiotensinogen, which that then go on to regulate functions in the cell. The first known infections from SARS-CoV-2 were discovered in Wuhan, China, somewhere between November – December 2019. The disease has since spread worldwide, leading to an ongoing pandemic. The original source of viral transmission to humans remains unclear, as does whether the virus became pathogenic before or after the spillover event.

SARS-CoV-2 infects people of all ages. However, evidence to date suggests that two groups of people are at a higher risk of getting severe COVID-19 disease. These are older people (people over 60 years old) and those with underlying medical conditions (such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer). The risk of severe disease gradually increases with age starting from around 40 years.

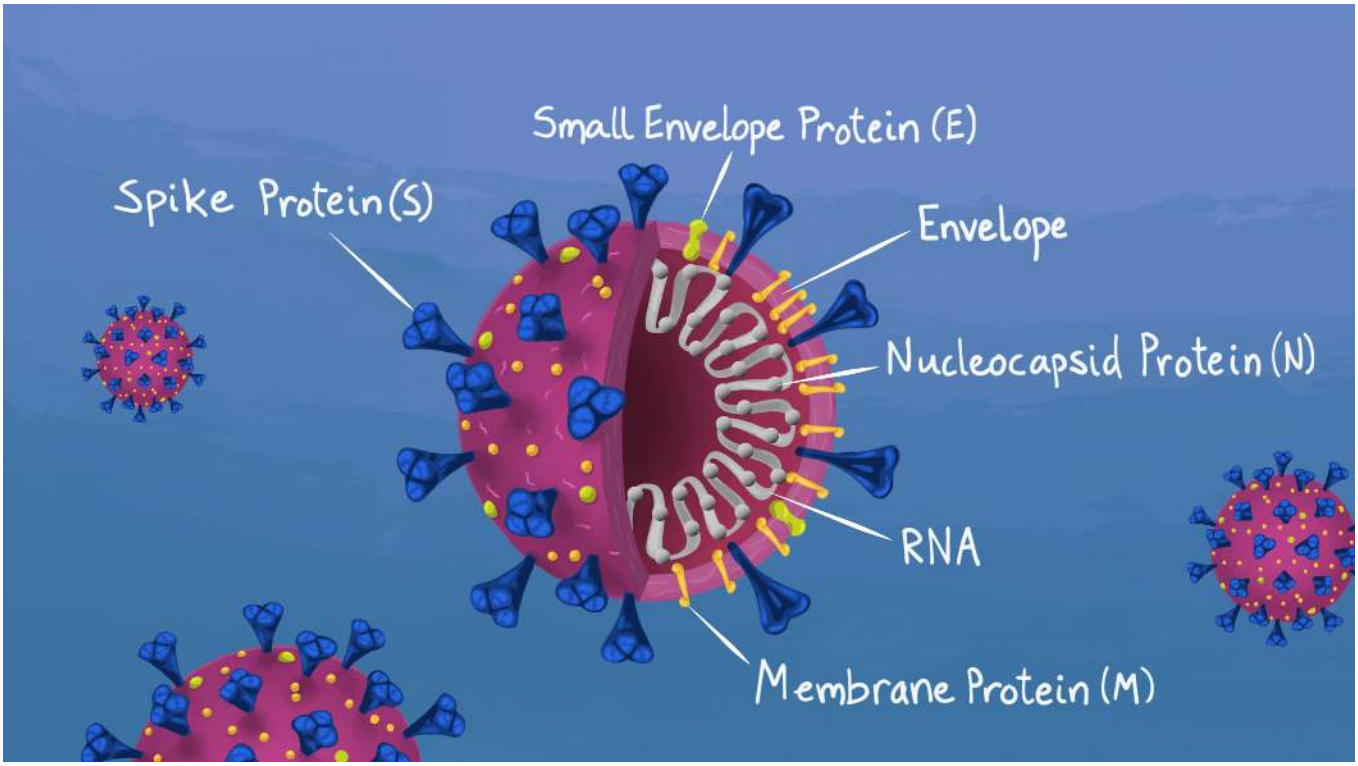
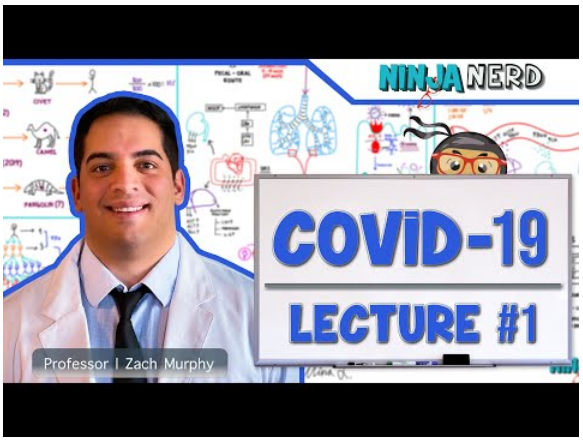


Figure 13.1.1 Anatomy of SARS-CoV-2 [Long description]

Take a look at the video below, *COVID-19 | Coronavirus: Epidemiology, Pathophysiology, Diagnostics*, by Ninja Nerd (2020) on YouTube, which summarizes some important and interesting facts about coronaviruses from the Government of Canada (2020).



A YouTube element has been excluded from this version of the text. You can view it online here: https://opengenetics.pressbooks.tru.ca/?p=2020

Media Attribution

- **Figure 13.1.1** SARS-CoV-2, anatomy/anatomía by Maya Peters Kostman for the CC BY-NC-SA 4.0, Innovative Genomics Institute

Reference

Health Canada. (2020, March). COVID-19 | Coronavirus: Epidemiology, pathophysiology, diagnostics [Video file]. YouTube. https://www.youtube.com/watch?v=PWzbArPgo-o

Long Description

- **Figure 13.1.1** The anatomy of the SARS-Cov-2 virus, with the spike proteins on the surface shown, along with various membrane proteins, nucleocapsid proteins, small envelope proteins, along with membrane envelopes and RNA molecules within. [Back to Figure 13.1.1]

This page titled 13.8.1: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.1: Introduction** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

13.8.2: How Does SARS-Cov-2 Infect Humans?

In order to develop effective strategies to diagnose, treat, and manage this disease, it is vital to understand exactly how SARS-CoV-2 enters human cells. The virus' surface spike protein mediates SARS-CoV-2 entry into cells by binding to the ACE2 (Angiotensin-converting enzyme 2) receptor in humans through its receptor-binding domain and is proteolytically activated by human proteases. Cell entry of SARS-CoV-2 is preactivated by proprotein convertase furin, reducing its dependence on target cell proteases for entry, thus making it more efficient in cell entry and infection.

In humans, the ACE2 receptor protein is present in many cell types (especially epithelial cells) and tissues including the nose, mouth, lungs, heart, blood vessels, kidneys, liver, and gastrointestinal tract. ACE2 assists in modulating the activities of a protein called angiotensin II (ANG II) which increases blood pressure and inflammation, thereby increasing damage to blood vessel linings and promotes various types of tissue injury. ACE2 converts ANG II to other molecules which effectively counteract the effects of ANG II, such as inflammation and cell death. When the SARS-CoV-2 virus binds to the ACE2 receptor, it prevents ACE2 from performing its normal function to regulate ANG II signaling. As such, ACE2 action is inhibited, removing the protective mechanism from ANG II signaling through increased availability of ANG II to injure tissues, especially in the lungs and heart. **Figure 13.2.1** summarizes the transmission and life-cycle of SARS-CoV-2 causing COVID-19.

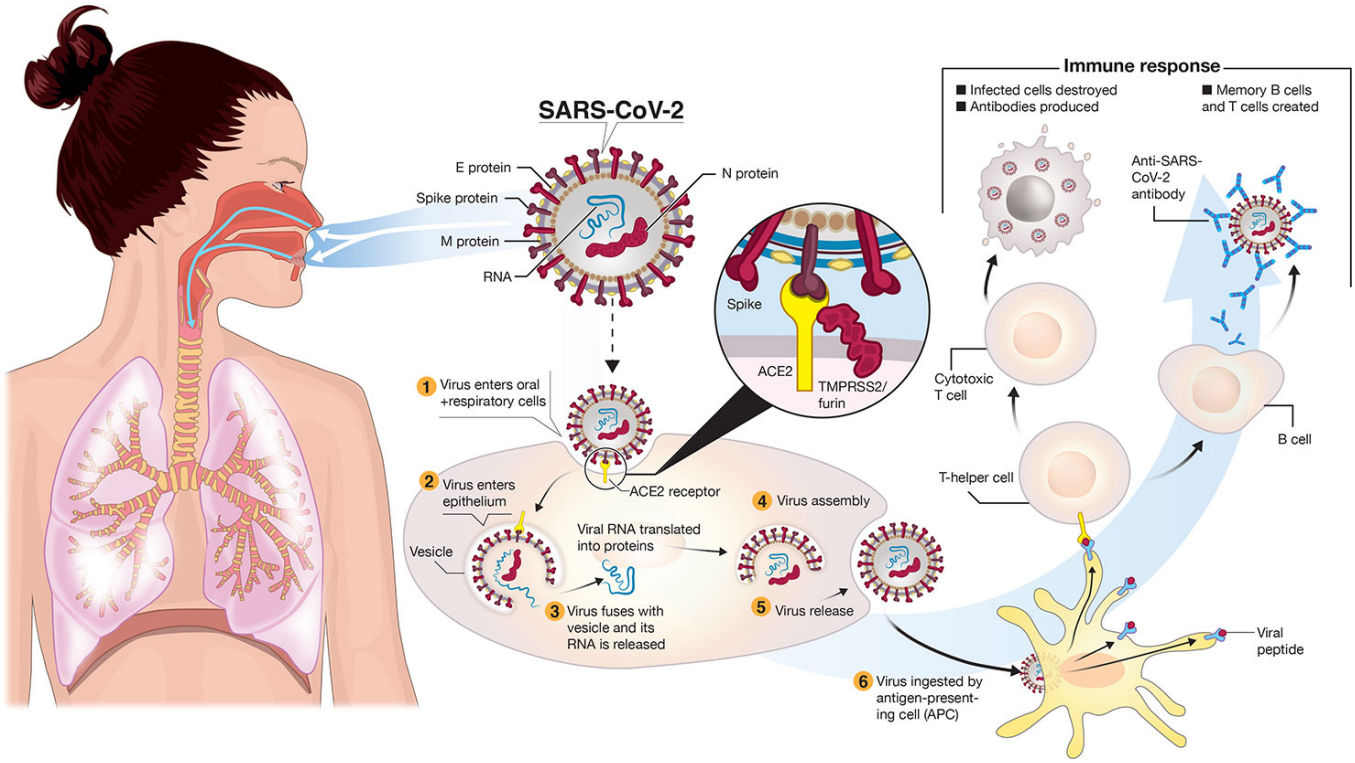
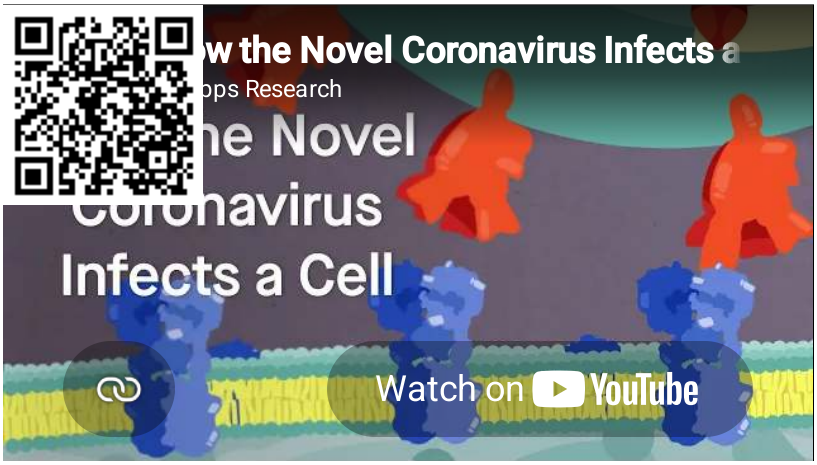


Figure 13.2.1 Transmission and Life-Cycle of SARS-CoV-2 Causing COVID-19. SARS-CoV-2 is transmitted via respiratory droplets of infected cases to oral and respiratory mucosal cells. The virus, possessing a single-stranded RNA genome wrapped in nucleocapsid (N) protein and three major surface proteins: membrane (M), envelope (E) and Spike, replicates and passes to the lower airways potentially leading to severe pneumonia. The gateway to host cell entry (magnified view) is via Spike-converting enzyme 2 (ACE2) interaction with cleavage of Spike in the prefusion state by proteases TMPRSS-2/furin. A simplified depiction of the life cycle of the virus is shown along with potential immune responses elicited. [Long description]

Take a look at the video below, *How the Novel Coronavirus Infects a Cell: Science, Simplified*, by Scripps Research (2020) on YouTube, describing how the novel coronavirus that causes COVID-19 enters the body and infects cells. Illustrated by a Scripps Research scientist, this installment of *Science, Simplified* gives an overview of the entire infection process.



Media Attribution

- **Figure 13.2.1**Fphar-11-00937-g001 by Colin D. Funk, Craig Laferrière, and Ali Ardakani (2020), CC BY 4.0, via Wikimedia Commons

Reference

Funk, C. D., Laferrière, C., Ardakani, A. (2020, June 19). A snapshot of the global race for vaccines targeting SARS-CoV-2 and the COVID-19 pandemic. *Frontiers in Pharmacology11*. https://doi.org/10.3389/fphar.2020.00937

Scripps Research. (2020, July 13). How the novel coronavirus infects a cell (video file). YouTube. https://youtu.be/dA70ZdYhhCg

Long Description

- **Figure 13.2.1** Transmission and Life-Cycle of SARS-CoV-2 Causing COVID-19. SARS-CoV-2 is transmitted via respiratory droplets of infected cases to oral and respiratory mucosal cells. The virus, possessing a single-stranded RNA genome wrapped in nucleocapsid (N) protein and three major surface proteins: membrane (M), envelope (E) and Spike, replicates and passes to the lower airways potentially leading to severe pneumonia. The gateway to host cell entry (magnified view) is via Spike-converting enzyme 2 (ACE2) interaction with cleavage of Spike in the prefusion state by proteases TMPRSS-2/furin. A simplified depiction of the life cycle of the virus is shown along with potential immune responses elicited. [Back to Figure 13.2.1]

This page titled 13.8.2: How Does SARS-Cov-2 Infect Humans? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.2: How Does SARS-Cov-2 Infect Humans?** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

13.8.3: Genes Implicated in Severe COVID-19 Infection in Humans

Overview

The information in this section *(Overview)* was extracted from Elhabyan, A. et al. (2020), The role of host genetics in susceptibility to severe viral infections in humans and insights into host genetics of severe COVID-19: A systematic review.

Susceptibility to severe viral infections was reported to be associated with genetic variants in immune response genes. This article systematically reviewed the genes related to viral susceptibility that were reported in human genetic studies (case-reports and genome wide association studies) to understand the role of host viral interactions and to provide insights into the pathogenesis of severe COVID-19. Approximately 15% of cases are severe and some of them are accompanied by a dysregulated immune system and cytokine storm. There is increasing evidence that severe manifestations of COVID-19 might be attributed to human genetic variants in genes related to immune deficiency and/or inflammasome activation (cytokine storm). Forty genes were found to be associated with viral susceptibility and 21 of them were associated with severe SARSCoV disease and severe COVID-19. Some of those genes were implicated in toll-like receptor pathways, others in C-lectin pathways, and others were related to inflammasome activation (cytokine storm).

Table 13.1 *Summary of Review (Elhabyan et al., 2020) – Showing Genes Associated with Clinical Manifestation of SARS-CoV/2*

Clinical Manifestation	Genes Associated
Susceptibility to SARS-CoV infection	CD14, HLA-B, FCGR2A, CCL2, CCL5, MxA, ABO, MBL, OAS-1, ICAM3, DC-SIGN
Susceptibility to SARS-CoV2 infection	ALOXE3, TMEM181, BRF2, ERAP2, LC6A20, LZTFL1, CCR9, FYCO1, CXCR6, XCR1, ABO, ApoE

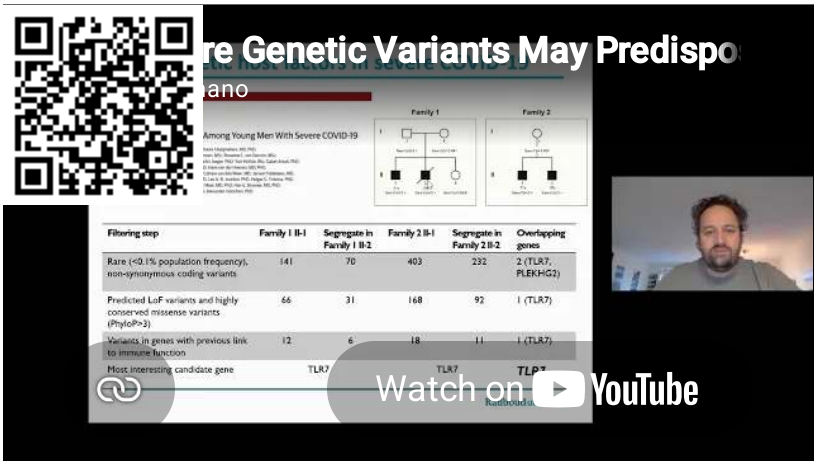
A Closer Look at Popular Theories

As researchers work tirelessly to uncover the genetic basis of COVID-19 severity and susceptibility, the following outlines some popular opinions based upon the science, as it stands when this was written. As the science advances, so will our theories and understanding.

- Emily Willingham's (2020) article, "Genes May Influence COVID-19 Risk, New Studies Hint" at *Scientific American*, and Jocelyn Kaiser's (2020) article, "Found: genes that sway the course of the coronavirus" at *Science* from American Association for the Advancement of Science. speak to early theories concerning**Chromosome 3 and ABO blood groups.**
- Another article from from Ann Gibbons (2020) at *Science*, "Neanderthal gene found in many people may open cells to coronavirus and increase COVID-19 severity" discusses the theory related to **Chromosome 2, specifically the DPP4 gene known as the Neanderthal gene** found in many people which may open cells to coronavirus and increase COVID-19 severity.
- An article from the National Institutes of Health (NIH, 2020), "Scientists discover genetic and immunologic underpinnings of some cases of severe COVID-19" speaks to a molecular explanation for **why more men than women die from COVID-19**.
- An article from John Hewitt (2020), *Medical Xpress*, "Unique susceptibility to unique Sars-CoV-2 variants and vaccines" gives a brief overview of how**individuals with different genetic variants in their immune system** components often have very different immune responses to SARS-CoV-2, and explains that they also will have different responses to vaccines.

Take a look at the video below by Dr. Alex Hoischen, Radboud University (Bionano Genomics, 2021), where he discusses his published results on genomic variants found in families with severe COVID-19. In two families with severely ill brothers, mutations were found in the Toll-Like Receptor 7 gene (TLR7), which affects the production of interferons, signaling molecules used to control the immune response. Several other studies have since made similar findings in other genes of the TLR family. Dr. Hoischen discussed how individual patients each may carry individually rare variants, that are collectively common and point to important pathways involved in the disease. His interest in the consortium is based on his understanding that larger SVs have a greater chance to be rare and disruptive, and genome-wide studies have lacked so far in their assessment.

The video below, *Rare Genetic Variants May Predispose to Severe COVID-19*, by Bionano Genomics (2021) on YouTube, discusses the links between genes and incidence of severe COVID-19 infection.



References

Bionano Genomics. (2021, January 22). Rare genetic variants may predispose to severe COVID-19 (video file). YouTube. https://www.youtube.com/watch?v=RaustWij4yg

Elhabyan, A. et. al. (2020). The role of host genetics in susceptibility to severe viral infections in humans and insights into host genetics of severe COVID-19: A systematic review. *Virus Research 289*(2020), 198163. https://doi.org/10.1016/j.virusres.2020.198163

Gibbons, A. (2020, December 18). Neanderthal gene found in many people may open cells to coronavirus and increase COVID-19 severity. *Science.* www.science.org/content/article/neanderthal-gene-found-many-people-may-open-cells-coronavirus-and-increase-covid-19

Hewitt, J. (2020, December 30). Unique susceptibility to unique Sars-CoV-2 variants and vaccines. *Medical Xpress.*https://medicalxpress.com/news/2020-12-unique-susceptibility-sars-cov-variants-vaccines.html

Kaiser, J. (2020, October 13). Found: genes that sway the course of the coronavirus. *Science 370*(6514), 275-276. www.science.org/doi/10.1126/science.370.6514.275

National Institutes of Health (NIH). (2020, September 24). Scientists discover genetic and immunologic underpinnings of some cases of severe COVID-19. U.S. Department of Health and Human Services. https://www.nih.gov/news-events/news-releases/scientists-discover-genetic-immunologic-underpinnings-some-cases-severe-covid-19

Willingham, E. (2020, July 21). Genes may influence COVID-19 risk, new studies hint. *Scientific American.*www.scientificamerican.com/article/genes-may-influence-covid-19-risk-new-studies-hint/

This page titled 13.8.3: Genes Implicated in Severe COVID-19 Infection in Humans is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.3: Genes Implicated in Severe COVID-19 Infection in Humans** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

13.8.4: Approaches for Vaccine Development Against SARS-Cov-2

The elucidation of the genome organization and functional domains of S protein for SARS-CoV **(Figure 13.4.1),** achieved through the work of scientists and geneticists all over the globe, has facilitated a deep understanding of the mode of action of this virus which has led to the development of a myriad of vaccine and drug candidates in an effort to mitigate the spread of this virus, as well as to assist in the diagnosis and management of infected patients.

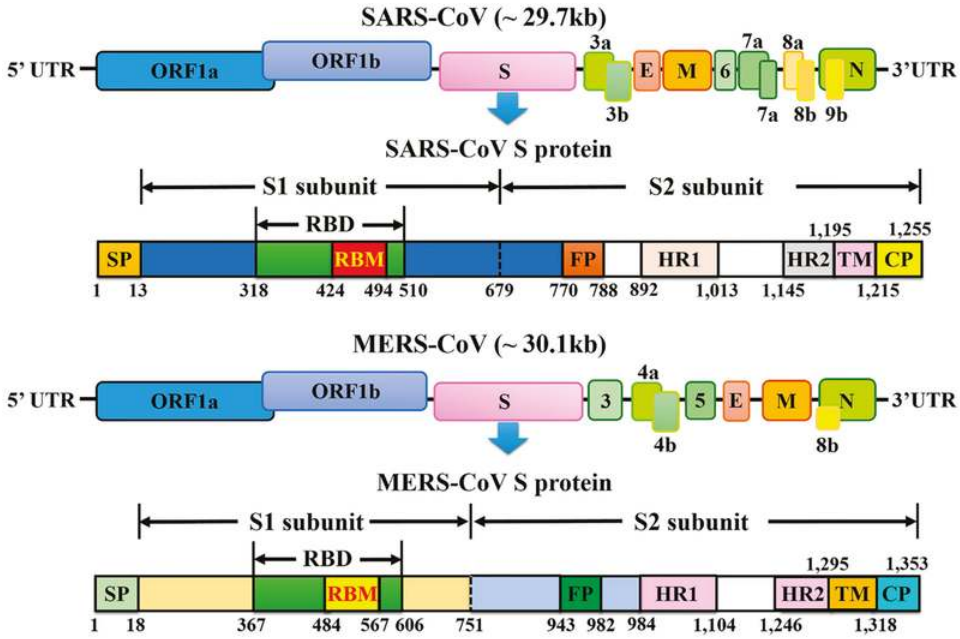


Figure 13.4.1 Schematic Representation of the Genome Organization and Functional Domains of S Protein for SARS-CoV and MERS-CoV. [Long description]

Figure 13.4.1 is a schematic representation of the genome organization and functional domains of S protein for SARS-CoV and MERS-CoV. The single-stranded RNA genomes of SARS-CoV and MERS-CoV encode two large genes, the ORF1a and ORF1b genes, which encode 16 non-structural proteins (nsp1–nsp16) that are highly conserved throughout coronaviruses. The structural genes encode the structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N), which are common features to all coronaviruses. The accessory genes (shades of green) are unique to different coronaviruses in terms of number, genomic organization, sequence, and function. The structure of each S protein is shown beneath the genome organization. The S protein mainly contains the S1 and S2 subunits. The residue numbers in each region represent their positions in the S protein of SARS and MERS, respectively. The S1/S2 cleavage sites are highlighted by dotted lines. SARS-CoV, severe acute respiratory syndrome coronavirus; MERS-CoV, Middle East respiratory syndrome coronavirus; CP, cytoplasm domain; FP, fusion peptide; HR, heptad repeat; RBD, receptor-binding domain; RBM, receptor-binding motif; SP, signal peptide; TM, transmembrane domain.

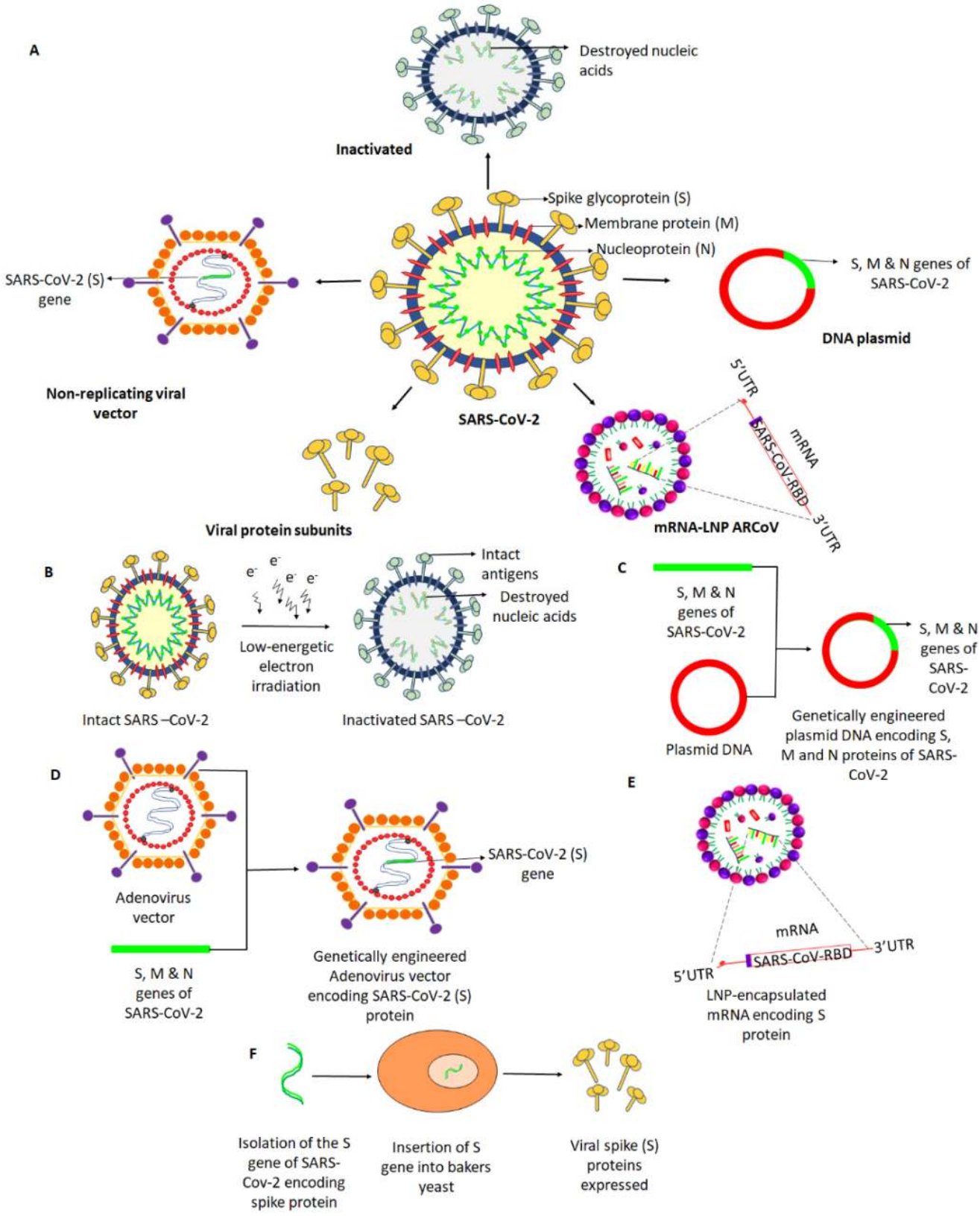
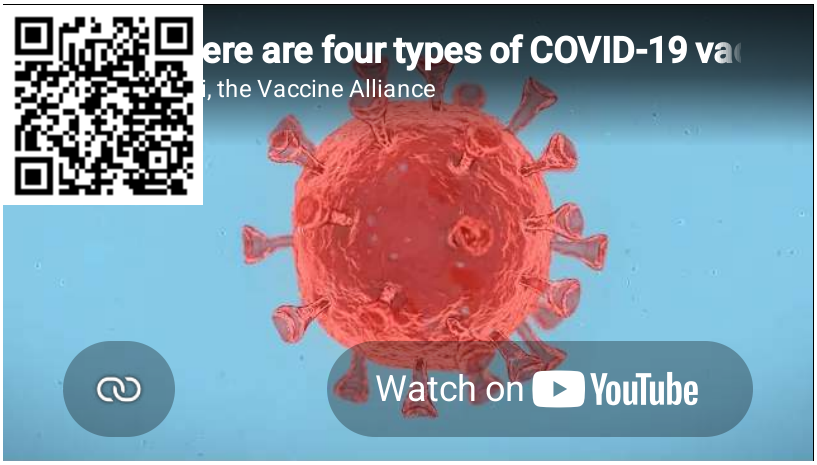


Figure 13.4.2 Different Approaches for the Development of Vaccine Candidates Against SARS-Cov-2 [Long description]

Different approaches for the development of vaccine candidates against SARS-Cov-2 have, and are being, developed and utilized. **Figure 13.4.2** summarizes these methods:

A. Potential vaccines under development, involve five leading platforms (inactivated virus, protein subunit, DNA, RNA, and non-replicating viral vector), as depicted.
B. Intact SARS-CoV-2 is neutralized by treatment with radiation to cease its ability to infect and replicate, while preserving the induction of an immune response.
C. A plasmid DNA is genetically engineered with the S, M, and N genes of SARS-CoV-2 encoding the respective proteins that may facilitate an immune response.
D. A replication-defective Adenovirus (Ad) vector is genetically engineered to express SARS-Cov-2 spike (S) protein.
E. An mRNA (replication-defective) that encodes the S protein of SARS-CoV-2 is encapsulated in a lipid nanoparticle (LNP), which, when injected, induces the body cells to produce the spike protein and direct the immune response.
F. Spike protein-encoding (S) gene of SARS-CoV-2 was isolated and genetically engineered into a baker's yeast, producing the spike protein antigens when grown. The produced S antigens can then be collected and purified.

Watch the video below called, *There are four types of COVID-19 vaccines: here's how they work*, by Gavi, the Vaccine Alliance(2020) on YouTube, which gives a brief overview of the main types of COVID-19 vaccines and the mechanism by which they bring about immunity in a patient.



Media Attributions

- **Figure 13.4.1** SARS-CoV MERS-CoV genome organization and S-protein domains by Song, et al. (2019), CC BY 4.0, via Wikimedia Commons
- **Figure 13.4.2**Different approaches for the development of vaccine candidates against SARS-Cov-2 by Faheem et al. (2021), CC BY 4.0, via Wikimedia Commons

References

Faheem, M. S. et al. (2021). Platforms exploited for SARS-CoV-2 vaccine development. *Vaccines* *9*(1), 11. https://doi.org/10.3390/vaccines9010011

Gavi, the Vaccine Alliance. (2020, December 18). There are four types of COVID-19 vaccines: here's how they work (video file). YouTube. https://www.youtube.com/watch?v=lFjIVIIcCvc

Song Z et al. (2019, January 14). From SARS to MERS, Thrusting coronaviruses into the spotlight. *Viruses 2019*, 11(1), 59. https://doi.org/10.3390/v11010059

Long Descriptions

- **Figure 13.4.1** Schematic representation of the genome organization and functional domains of S protein for SARS-CoV and MERS-CoV. The single-stranded RNA genomes of SARS-CoV and MERS-CoV encode two large genes, the ORF1a and ORF1b genes, which encode 16 non-structural proteins (nsp1–nsp16) that are highly conserved throughout coronaviruses. The structural genes encode the structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N), which are common features to all coronaviruses. [Back to Figure 13.4.1]

- **Figure 13.4.2** Different Approaches for the Development of Vaccine Candidates Against SARS-Cov-2. (A) Potential vaccines under development involve five leading platforms (inactivated virus, protein subunit, DNA, RNA, and non-replicating viral vector), as depicted. (B) Intact SARS-CoV-2 is neutralized by treatment with radiation to cease its ability to infect and replicate, while preserving the induction of an immune response. (C) A plasmid DNA is genetically engineered with the S, M, and N genes of SARS-CoV-2 encoding the respective proteins that may facilitate an immune response. (D) A replication-defective Adenovirus (Ad) vector is genetically engineered to express SARS-Cov-2 spike (S) protein. (E) An mRNA (replication-defective) that encodes the S protein of SARS-CoV-2 is encapsulated in a lipid nanoparticle (LNP), which, when injected, induces the body cells to produce the spike protein and direct the immune response. (F) Spike protein-encoding (S) gene of SARS-CoV-2 was isolated and genetically engineered into a baker's yeast, producing the spike protein antigens when grown. The produced S antigens can then be collected and purified. [Back to Figure 13.4.2]

This page titled 13.8.4: Approaches for Vaccine Development Against SARS-Cov-2 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.4: Approaches for Vaccine Development Against SARS-Cov-2** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

13.8.5: Chapter 13 Summary

The topics covered in this chapter can be summarized as follows:

- An introduction to the general group of RNA viruses known as coronaviruses with emphasis on the general structure of such viruses
- A summary of how the virus' surface spike protein mediates SARS-CoV-2 entry into human cells by binding to the ACE2 (Angiotensin-converting enzyme 2) receptor.
- A list of some of the genes Implicated in Severe COVID-19 Infection in Humans
- A discussion of the approaches for Vaccine Development Against SARS-Cov-2

This page titled 13.8.5: Chapter 13 Summary is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.5: Chapter 13 Summary** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

13.8.6: Key Terms in Chapter 13

Key Terms – Genes and COVID-19 Susceptibility in Humans

- Angiotensin-converting enzyme 2
- C-lectin pathway
- Coronavirus
- Cytokine storm
- spike protein
- Toll-like receptor
- Vaccine

This page titled 13.8.6: Key Terms in Chapter 13 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.6: Key Terms in Chapter 13** by Natasha Ramroop Singh is licensed CC BY-NC-SA 4.0. Original source: https://opengenetics.pressbooks.tru.ca.

13.E: Cancer Genetics (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions:

13.1 Why do oncogenes tend to be dominant, but mutations in tumor suppressors tend to be recessive?

13.2 What tumor suppressing functions are controlled by p53? How can a single gene affect so many different biological pathways?

13.3 Are all carcinogens mutagens? Are all mutagens carcinogens? Explain why or why not.

13.4 Imagine that a laboratory reports that feeding a chocolate to laboratory rats increases the incidence of cancer. What other details would you want to know before you stopped eating chocolate?

13.5 Do all women with HPV get cancer? Why or why not? Do all women with mutations in *BRCA1* get cancer? Why or why not?

This page titled 13.E: Cancer Genetics (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

13.S: Cancer Genetics (Summary)

- Cancer is the name given to a class of different diseases that share common properties.
- Most cancers require accumulation of mutations in several different genes.
- Most cancer causing mutations are sporadic, rather than inherited, and most are caused by environmental carcinogens, including virus, radiation, and certain chemicals.
- Oncogenes are hyperactivated regulators of cell division, and are often derived from gain-of-function mutations in proto-oncogenes.
- Tumor suppressor genes normal help to repair DNA damage, arrest cell division, or to kill over proliferating cells. Loss-of-function of these genes contributes to the progression of cancer.
- Genetic research into cancer can provide enzyme targets for drug investigation and potential treatment. E.g. Gleevec™

Key Terms:

metastasis	*ras*
dysplasia	apoptosis
carcinogen	*BRC1A*
HPV	p53
oncogene	tumor suppressor
ionizing	phosphorylation
epidemiology	CML
proto-oncogene	Gleevec™
receptor	*bcr-abl*
signal transduction	

This page titled 13.S: Cancer Genetics (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **13.S: Cancer Genetics (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

14: Genomics and Systems Biology

Imagine that you could identify and quantify every molecule within a cell (Figure 11.1) in a single assay. You could use this ability to better understand almost any aspect of biology. For example, by comparing the molecular profiles of plants that differed in their resistance to drought, you might discover which combination of genes or proteins makes a crop drought tolerant. Although it is not currently possible to study literally every molecule in a cell in a single experiment, recent advances in molecular biology have made it possible to study many genes (or their products) in parallel.

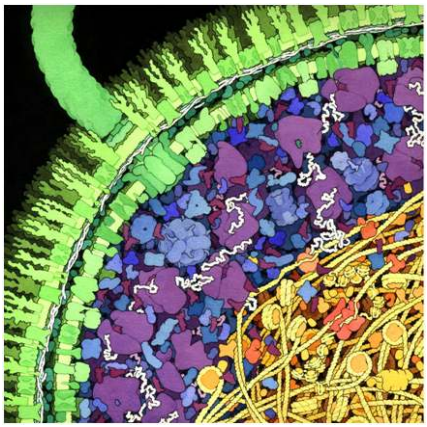


Figure 11.1: *An artist's depiction of part of an E.coli cell, showing many different types of molecules in their typical abundance. mRNA appears as white lines associated with purple ribosomes, while DNA and proteins such as histones are yellow. (Goodsell, Scripps-EDU)*

14.1: 'Omics Technologies

14.2: DNA Sequencing

14.3: Whole Genome Sequencing

14.4: Functional Genomics – Determining Function(s)

14.E: Genomics and Systems Biology (Exercises)

14.S: Genomics and Systems Biology (Summary)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 14: Genomics and Systems Biology is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

14.1: ‘Omics Technologies

The complete set of DNA within an organism is called its **genome**. **Genomics** is therefore the large-scale description, using techniques of molecular biology of many genes or even whole genomes at once. This type of research is facilitated by technologies that increase throughput (i.e. rate of analysis), and decrease cost. The –**omics** suffix has been used to indicate high-throughput analysis of many types of molecules, including transcripts (**transcriptomics**), proteins (**proteomics**), and the products of enzymatic reactions, or metabolites (**metabolomics**; Figure 14.1.1). Interpretation of the large data sets generated by –omics research depends on a combination of computational, biological, and statistical knowledge provided by experts in **bioinformatics**. Attempts to combine information from different types of –‘omics studies is sometimes called **systems biology**.



Figure 14.1.1: A word cloud listing some of the different –omics technologies. Terms that are more widely used are written in the largest characters. There is no significance to the color of each word. (Original-Deyholos-CC:AN)

This page titled 14.1: ‘Omics Technologies is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **11.1: ‘Omics Technologies** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

14.2: DNA Sequencing

DNA sequencing determines the order of nucleotide bases within a given fragment of DNA. This information can be used to infer the RNA or protein sequence encoded by the gene, from which further inferences may be made about the gene's function and its relationship to other genes and gene products. DNA sequence information is also useful in studying the regulation of gene expression. If DNA sequencing is applied to the study of many genes, or even a whole genome, it is considered an example of genomics.

Dideoxy sequencing

Recall that DNA polymerases incorporate nucleotides (dNTPs) into a growing strand of DNA, based on the sequence of a template strand. DNA polymerases add a new base only to the 3'-OH group of an existing strand of DNA; this is why primers are required in natural DNA synthesis and in techniques such as PCR. Most of the currently used DNA sequencing techniques rely on the random incorporation of modified nucleotides called terminators. Examples of terminators are the **dideoxy nucleotides** (**ddNTPs**), which lack a 3'-OH group and therefore cannot serve as an attachment site for the addition of new bases to a growing strand of DNA (Figure 14.2.1). After a ddNTP is incorporated into a strand of DNA, no further elongation can occur. Terminators are labeled with one of four fluorescent dyes, each specific for one the four nucleotide bases.

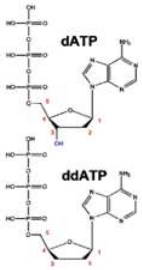


Figure 14.2.1: *ddNTPs (Original-Deyholos-CC:AN)*

To sequence a DNA fragment, you need many copies of that fragment (Figure 14.2.2). Unlike PCR, DNA sequencing does not amplify the target sequence and only one primer is used. This primer is hybridized to the denatured template DNA, and determines where on the template strand the sequencing reaction will begin. A mixture of dNTPs, fluorescently labeled terminators, and DNA polymerase is added to a tube containing the primer-template hybrid. The DNA polymerase will then synthesize a new strand of DNA until a fluorescently labeled nucleotide is incorporated, at which point extension is terminated. Because the reaction contains millions of template molecules, a corresponding number of shorter molecules is synthesized, each ending in a fluorescent label that corresponds to the last base incorporated.

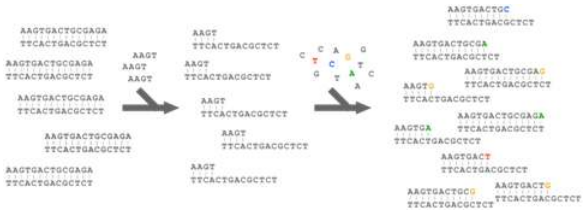


Figure 14.2.2: A sequencing reactions begins with many identical copies of a template DNA fragment. The template is denatured, then primers are annealed to the template. Following the addition of polymerase, regular dNTPS, and fluorescently labeled terminators, extension begins at the primer site. Elongation proceeds until a fluorescently labeled terminator (shown here in color) is incorporated. (Original-Deyholos-CC:AN)

The newly synthesized strands can be denatured from the template, and then separated electrophoretically based on their length (Figure 14.2.3). Since each band differs in length by one nucleotide, and the identity of that nucleotide is known from its fluorescence, the DNA sequence can be read simply from the order of the colors in successive bands. In practice, the maximum length of sequence that can be read from a single sequencing reaction is about 700 bp.

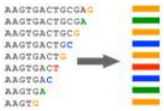


Figure 14.2.3: Fluorescently labeled products can be separated electrophoretically based on their length. (Original-Deyholos-CC:AN)

A particularly sensitive electrophoresis method used in the analysis of DNA sequencing reactions is called capillary electrophoresis (Figure 14.2.6). In this method, a current pulls the sequencing products through a gel-like matrix that is encased in a fine tube of clear plastic. As in conventional electrophoresis, the smallest fragments move through the capillary the fastest. As they pass through a point near the end of the capillary, the fluorescent intensity of each dye is read. This produces a graph called a chromatogram. The sequence is determined by identifying the highest peak (i.e. the dye with the most intense fluorescent signal) at each position.

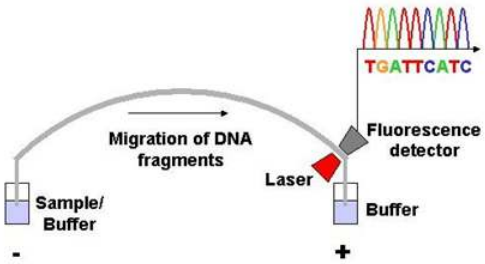


Figure 14.2.4: Fluorescently labeled products can be separated by capillary electrophoresis, generating a chromatogram from which the sequence can be read.(Wikipedia-Abizar Lakdawalla-PD)

Next-generation sequencing

Advances in technology over the past two decades have increased the speed and quality of sequencing, while decreasing the cost. This has become especially true with the most recently developed methods called next-generation sequencing. Not all of these new methods rely on terminators, but one that does is a method used in instruments sold by a company called **Illumina**. Illumina sequencers use a special variant of PCR called bridge PCR to make many thousands of copies of a short (45bp) template fragment. Each of these short template fragments is attached in a cluster in a small spot on a reaction surface. Millions of other clusters, each made by different template fragment, are located at other positions on the reaction surface. DNA synthesis at each template strand then proceeds using dye-labeled terminators that are used are reversible. Synthesis is therefore terminated (temporarily) after the incorporation of each nucleotide. Thus, after the first nucleotide is incorporated in each strand, a camera records the color of fluorescence emitted from each cluster. The terminators are then modified, and a second nucleotide is incorporated in each strand, and again the reaction surface is photographed. This cycle is repeated a total of 45 times. Because millions of 45 bp templates are sequenced in parallel in a single process, Illumina sequencing is very efficient compared to other sequencing techniques. However, the short length of the templates currently limits the application of this technology.

This page titled 14.2: DNA Sequencing is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **11.2: DNA Sequencing** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

14.3: Whole Genome Sequencing

The need for assembly

Given that the length of a single, individual sequencing read is somewhere between 45bp and 700bp, we are faced with a problem determining the sequence of longer fragments, such as the chromosomes in an entire genome of humans (3×10^9 bp). Obviously, we need to break the genome into smaller fragments. There are two different strategies for doing this:

1. clone-by-clone sequencing, which relies on the creation of a physical map first then sequencing, and
2. whole genome shotgun sequencing, which sequences first and does not require a physical map.

Physical mapping

A physical map is a representation of a genome, comprised of cloned fragments of DNA. The map is therefore made from physical entities (pieces of DNA) rather than abstract concepts such as the linkage frequencies and genes that make up a genetic map (Figure 14.3.1). It is usually possible to correlate genetic and physical maps, for example by identifying the clone that contains a particular molecular marker. The connection between physical and genetic maps allows the genes underlying particular mutations to be identified through a process call map-based cloning.

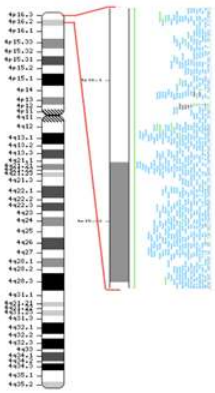


Figure 14.3.1: A portion of the physical map for human chromosome 4. The entire chromosome is shown at left. The physical map is made up of small blue lines, each of which represents a cloned piece of DNA approximately 100kb in length. (NCBI-unknown-PD)

To create a physical map, large fragments of the genome are cloned into plasmid vectors, or into larger vectors called bacterial artificial chromosomes (BACs). BACs can contain approximately 100kb fragments. The set of BACs produced in a cloning reaction will be redundant, meaning that different clones will contain DNA from the same part of the genome. Because of this redundancy, it is useful to select the minimum set of clones that represent the entire genome, and to order these clones respective to the sequence of the original chromosome. Note that this is all to be done without knowing the complete sequence of each BAC. Making a physical map may therefore rely on techniques related to Southern blotting: DNA from the ends of one BAC is used as a probe to find clones that contain the same sequence. These clones are then assumed to overlap each other. A set of overlapping clones is called a **contig**.

Clone-by-clone sequencing

Physical mapping of cloned sequences was once considered a pre-requisite for genome sequencing. The process would begin by breaking the genome into BAC-sized pieces, arranging these BACs into a map, then breaking each BAC up into a series of smaller clones, which were usually then also mapped. Eventually, a minimum set of smaller clones would be identified, each of which was small enough to be sequenced (Figure 14.3.8). Because the order of clones relative to the complete chromosome was known prior to sequencing, the resulting sequence information could be easily assembled into one complete chromosome at the end of the project. Clone-by-clone sequencing therefore minimizes the number of sequencing reactions that must be performed, and makes sequence assembly straightforward and reliable. However, a drawback of this strategy is the tedious process of building physical map prior to any sequencing.

Whole genome shotgun sequencing

This strategy breaks the genome into fragments that are small enough to be sequenced, then reassembles them simply by looking for overlaps in the sequence of each fragment. It avoids the laborious process of making a physical map (Figure 14.3.2). However, it requires many more sequencing reactions than the clone-by-clone method, because, in the shotgun approach, there is no way to avoid sequencing redundant fragments. There is also a question of the feasibility of assembling complete chromosomes based simply on the sequence overlaps of many small fragments. This is particularly a problem when the size of the fragments is smaller than the length of a repetitive region of DNA. Nevertheless, this method has now been successfully demonstrated in the nearly complete sequencing of many large genomes (rice, human, and many others). It is the current standard methodology.

However, shotgun assemblies are rarely able to complete entire genomes. The human genome, for example, relied on a combination of shotgun sequence and physical mapping to produce contiguous sequence for the length of each arm of each chromosome. Note that because of the highly repetitive nature of centromeric and telomeric DNA, sequencing projects rarely include these heterochromatic, gene poor regions.

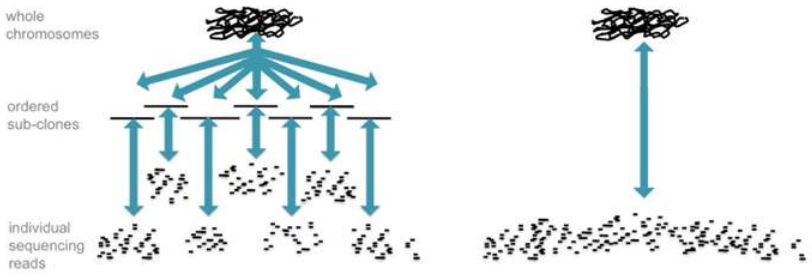


Figure 14.3.2: Genome sequencing strategies. A clone-by-clone strategy (left) in which the genome is divided into progressively smaller units (clones) before sequencing. Whole genome shotgun strategy (right) assembles the sequence from all the smaller reads. (Origianl-Deyholos-CC:AN)

Genome analysis

An assembled genome is a string of millions of A's,C's,G's,T's. Which of these represent nucleotides that encode proteins, and which of these represent other features of genes and their regulatory elements? The process of **genome annotation** relies on computers to define features such a start and stop codons, introns, exons, and splice sites. However, few of the predictions made by these programs is entirely accurate, and most must be verified experimentally for any gene of particular importance or interest.

This page titled 14.3: Whole Genome Sequencing is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **11.3: Whole Genome Sequencing** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

14.4: Functional Genomics – Determining Function(s)

Having identified putative genes within a genome sequence, how do we determine their function? Techniques of functional genomics are an experimental approach to address this question. One widely used technique in functional genomics is called **microarray analysis** (Figure 14.4.1). Microarrays can measure the abundance of mRNA for hundreds or thousands of genes at once. The abundance of mRNA of a particular gene is usually correlated with the activity of that gene. For example, genes that are involved in neuronal development likely produce more mRNA in brain tissue than in heart tissue. We can therefore learn about the relationship between particular genes and particular processes by comparing transcript abundance under different conditions. This can identify tissue specific expression (e.g. the nerve/heart example above), as well as differences in temporal expression (development), or exposure to external agents (eg, disease, hormones, drugs, etc.).

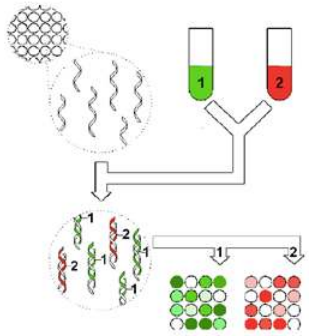


Figure 14.4.1: An example of a type of DNA microarray. Fluorescently labeled molecules derived from the transcripts of two tissue samples are hybridized to immobilized DNA molecules on the surface of an array. The labeled molecules bind in proportion to their abundance in the original tissue samples. The amount of green or red fluorescence at each spot can be measured using filters for the appropriate wavelengths. Thus, spots on the micoarray that are more green than red represent genes that are more abundant in the tissue sample from which green-labeled molecules were derived. (Original-Deyholos-CC:AN)

This page titled 14.4: Functional Genomics – Determining Function(s) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **11.4: Functional Genomics – Determining Function(s)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

14.E: Genomics and Systems Biology (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions

11.1 What are the advantages of high-throughput –omics techniques compared to studying a single gene or protein at a time? What are the disadvantages

11.2 What would the chromatogram from a capillary sequencer look like if you accidentally added only template, primers, polymerase, and fluorescent terminators to the sequencing reaction?

11.3 What are the advantages and disadvantages of clone-by-clone vs. whole genome shotgun sequencing?

11.4 How could you use DNA sequencing to identify new species of marine microorganisms?

11.5 Explain how you could use a microarray to identify wheat genes that have altered expression during drought?

11.6 A microarray identified 100 genes whose transcripts are abundant in tumors, but absent in normal tissues. Do any or all of these transcripts cause cancer? Explain your answer.

11.7 How could you ensure that each spot printed on a microarray contains DNA for only one gene?

11.8 What would the spots look like on a microarray after hybridization, if each spot contained a random mixture of genes?

11.9 What would the spots look like if the hybridization of green and red labeled DNA was done at low stringency?

This page titled 14.E: Genomics and Systems Biology (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

14.S: Genomics and Systems Biology (Summary)

- Genomics and related technologies differ from other techniques of molecular biology largely because of their scale; they allow many different genes (or gene products) to be studied in parallel.
- DNA sequencing can be applied to either a single gene, or in the case of genomics, to a large number of genes.
- Most DNA sequencing relies on the incorporation of dye-labeled terminator molecules, which create products that differ in length and end in a known nucleotide. The products can then be separated based on length, and the identity of the last based in each fragment can be determined based on fluorescence.
- Next-generation sequencing technologies have further reduced costs of sequencing, through miniaturization and parallelization.
- Physical maps are ordered sets of clones containing overlapping pieces of DNA, which together represent large pieces of chromosomes.
- Whole genomes may be sequenced using either a clone-by-clone approach, which required a physical map, or by a shotgun approach, in which small fragments are randomly sequenced.
- Genome analysis does not end after sequence acquisition; various features of the genome including genes (and their introns, exons, etc.) must be identified through a process called annotation.
- Functional genomics techniques including microarray analysis correlate transcript abundance with particular tissue samples. Genes whose transcripts are highly abundant under certain biological conditions may cause or respond to that condition.

Key Terms

genome	Ilumina
genomics	physical map
proteomics	BAC
transcriptomics	clone-by-clone sequencing
ddNTP	whole genome shotgun
terminator nucleotide	genome annotation
capillary electrophoresis	functional genomics
chromatogram	microarray
next-generation sequencing	

This page titled 14.S: Genomics and Systems Biology (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **11.S: Genomics and Systems Biology (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

15: Molecular Markers and Quantitative Traits



Figure 10.1: *Many traits, such as body mass, show continuous variation, rather than discrete variation. Although environment obviously also affects this trait, some of the variation observed between individuals is heritable, and is dependent on the interactions of multiple alleles at multiple loci. The study of quantitative traits is one of many applications of molecular markers. (Flickr-Jamie Golombek-CC:AND)*

15.1: Some Variations in the Genome Affect Complex Traits

15.2: Origins of Molecular Polymorphisms

15.3: Classification and Detection of Molecular Markers

15.4: Applications of Molecular Markers

15.5: Quantitative Trait Locus (QTL) Analysis

15.E: Molecular Markers and Quantitative Traits (Exercises)

15.S: Molecular Markers and Quantitative Traits (Summary)

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 15: Molecular Markers and Quantitative Traits is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

15.1: Some Variations in the Genome Affect Complex Traits

Imagine that you could compare the complete genomic DNA sequence of any two people you meet today. Although their sequences would be very similar on the whole, they would certainly not be identical at each of the 3 billion base pair positions you examined (unless, perhaps, your subjects were identical twins – but even they may have some somatic differences). In fact, the genomic sequences of almost any two unrelated people differ at millions of nucleotide positions. Some of these differences would be found in the regions of genes that code for proteins. Others might affect the amount of transcript that is made for a particular gene. A person's health, appearance, behavior, and other characteristics depend in part on these polymorphisms.

Most difference, however, have no effect at all. They have no effect on gene sequences or expression, because they occur within regions of DNA that neither encode proteins, nor regulate the expression of genes. These polymorphisms are nevertheless very useful because they can be used as **molecular markers** in medicine, forensics, ecology, agriculture, and many other fields. In most situations, molecular markers obey the same rules of inheritance that we have already described for other types of loci, and so can be used to create genetic maps and to identify linked genes.

This page titled 15.1: Some Variations in the Genome Affect Complex Traits is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.1: Some Variations in the Genome Affect Complex Traits** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

15.2: Origins of Molecular Polymorphisms

Mutations of DNA sequences can arise in many ways (Chapter 4). Some of these changes occur during DNA replication processes, resulting in an insertion, deletion, or substitution of one or a few nucleotides. Larger mutations can be caused by mobile genetic elements such as transposons, which are inserted more or less randomly into chromosomal DNA, sometimes occurring in clusters. In these and other types of **repetitive DNA** sequences, the number of repeated units is highly prone to change through unequal crossovers and other replication events.

This page titled 15.2: Origins of Molecular Polymorphisms is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.2: Origins of Molecular Polymorphisms** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

15.3: Classification and Detection of Molecular Markers

Regardless of their origins, molecular markers can be classified as polymorphisms that either vary in the length of a DNA sequence, or vary only in the identity of nucleotides at a particular position on a chromosome (Figure 15.3.1). In both cases, because two or more alternative versions of the DNA sequence exist, we can treat each variant as a different allele of a single locus. Each allele gives a different **molecular phenotype**. For example, polymorphisms of **SSRs** (short sequence repeats) can be distinguished based on the length of PCR products: one allele of a particular SSR locus might produce a 100bp band, while the same primers used with a different allele as a template might produce a 120bp band (Figure 15.3.2). A different type of marker, called a **SNP** (single nucleotide polymorphism), is an example of polymorphism that varies in nucleotide identity, but not length. SNPs are the most common of any molecular markers, and the genotypes of thousands of SNP loci can be determined in parallel, using new, hybridization based instruments. Note that the alleles of most molecular markers are co-dominant, since it is possible to distinguish the molecular phenotype of a heterozygote from either homozygote.

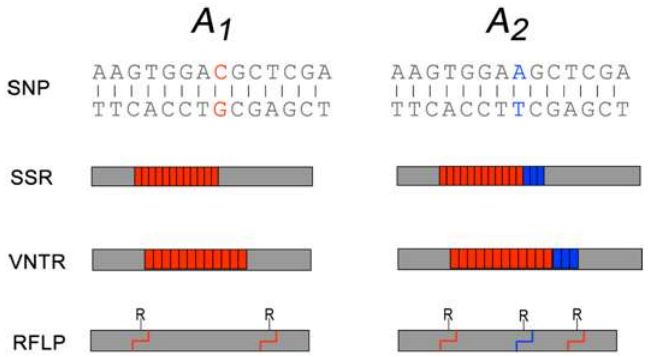


Figure 15.3.1: Some examples of DNA polymorphisms. The variant region is marked in blue, and each variant sequence is arbitrarily assigned one of two allele labels. Abbreviations: SNP (Single Nucleotide Polymorphism); SSR (Simple Sequence Repeat) = SSLP (Simple Sequence Length Polymorphism); **VNTR** (Variable Number of Tandem Repeats); **RFLP** (Restriction Fragment Length Polymorphisms. VNTRs and SSRs differ in the size of the repeat unit; VNTRs are larger than SSRs. (Original-Deyholos-CC:AN)

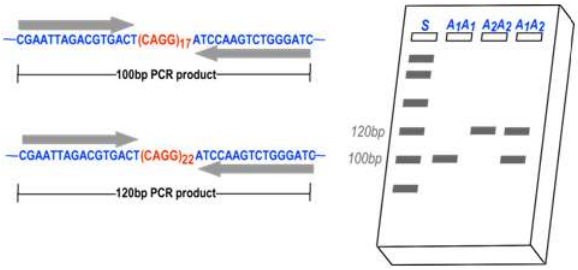


Figure 15.3.2: Determining the genotype of an individual at a single SSR locus using a specific pair of PCR primers and agarose gel electrophoresis. S= size standard (Original-Deyholos-CC:AN)

Mutations that do not affect the function of protein sequences or gene expression are likely to persist in a population as polymorphisms, since there will be no selection either in favor or against them (i.e. they are **neutral**). Note that the although the rate of spontaneous mutation in natural populations is sufficiently high so as to generate millions of polymorphisms that accumulate over thousands of generations, the rate of mutation is on the other hand sufficiently low that existing polymorphisms are stable throughout the few generations we study in a typical genetic experiment.

This page titled 15.3: Classification and Detection of Molecular Markers is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.3: Classification and Detection of Molecular Markers** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

15.4: Applications of Molecular Markers

Several characteristics of molecular markers make them useful to geneticists. First, because of the way DNA polymorphisms arise and are retained, they are frequent throughout the genome. Second, because they are phenotypically neutral, it is relatively easy to find markers that differ between two individuals. Third, their neutrality also makes it possible to study hundreds of loci without worrying about gene interactions or other influences that make it difficult to infer genotype from phenotype. Lastly, unlike visible traits such as eye color or petal color, the phenotype of a molecular marker can be detected in any tissue or developmental stage, and the same type of assay can be used to score molecular phenotypes at millions of different loci. Thus, the neutrality, high density, high degree of polymorphism, co-dominance, and ease of detection of molecular markers has lead to their wide adoption in many areas of research.

It is worth emphasizing again that DNA polymorphisms are a natural part of most genomes. Geneticists discover these polymorphisms in various ways, including comparison of random DNA sequence fragments from several individuals in a population. Once molecular markers have been identified, they can be used in many ways, including:

DNA fingerprinting

By comparing the allelic genotypes at multiple molecular marker loci, it is possible to determine the likelihood of similarity between two DNA samples. If markers differ, then clearly the DNA is from different sources. If they don't differ, then one can estimate the unlikelihood of them coming from different sources – eg they are from the same source. For example, a forensic scientist can demonstrate that a blood sample found on a weapon came from a particular suspect. Similarly, that leaves in the back of a suspect's pick-up truck came from a particular tree at a crime scene. DNA fingerprinting is also useful in paternity testing (Figure 15.4.1) and in commercial applications such as verification of species of origin of certain foods and herbal products.

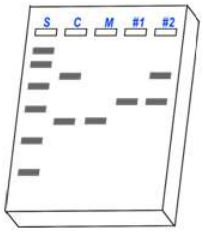


Figure 15.4.1: Paternity testing. Given the molecular phenotype of the child (C) and mother (M), only one of the possible fathers (#2) has alleles that are consistent with the child's phenotype.(Original-Deyholos-CC:AN)

Construction of genetic linkage maps

By calculating the recombination frequency between pairs of molecular markers, a map of each chromosome can be generated for almost any organism (Figure 15.4.2). These maps are calculated using the same mapping techniques described for genes in Chapter 7, however, the high density and ease with which molecular markers can be genotyped makes them more useful than other phenotypes for constructing genetic maps. These maps are useful in further studies, including map-based cloning of protein coding genes that were identified by mutation.

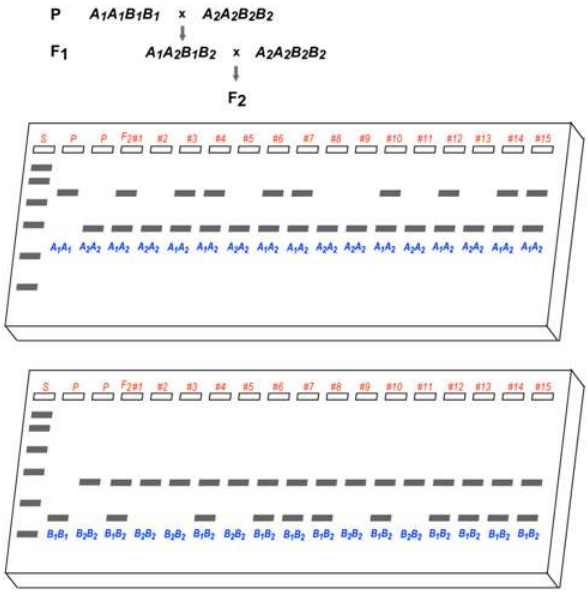


Figure 15.4.2: CMeasuring recombination frequency between two molecular marker loci, A and B. A different pair of primers is used to amplify DNA from either parent (P) and 15 of the F_2 offspring from the cross shown. Recombinant progeny will have the genotype $A_1A_2B_2B_2$ or $A_2A_2B_1B_2$. Individuals #3, #8, #13 are recombinant, so the recombination frequency is 3/15=20%. (Original-Deyholos-CC:AN)

Population studies

As described in Chapter 5, the observed frequency of alleles, including alleles of molecular markers, can be compared to frequencies expected for populations in Hardy-Weinberg equilibrium to determine whether the population is in equilibrium. By monitoring molecular markers, ecologists and wildlife biologists can make inferences about migration, selection, diversity, and other population-level parameters.

Molecular markers can also be used by anthropologists to study migration events in human ancestry. There is a large commercial business available that will genotype people and determine their deep genetic heritage for ~$100. This can be examined through the maternal line via sequencing their mitochondrial genome and through the paternal line via genotyping their Y-chromosome.

For example, about 8% of the men in parts of Asia (about 0.5% of the men in the world) have a Y-chromosomal lineage belonging to Genghis Khan (the haplogroup C3) and his decendents.

Identification of linked traits

It is often possible to correlate, or link, an allele of a molecular marker with a particular disease or other trait of interest. One way to make this correlation is to obtain genomic DNA samples from hundreds of individuals with a particular disease, as well as samples from a control population of healthy individuals. The genotype of each individual is scored at hundreds or thousands of molecular marker loci (e.g. SNPs), to find alleles that are usually present in persons with the disease, but not in healthy subjects. The molecular marker is presumed to be tightly linked to the gene that causes the disease, although this protein-coding gene may itself be as yet unknown. The presence of a particular molecular polymorphism may therefore be used to diagnose a disease, or to advise an individual of susceptibility to a disease.

Molecular markers may also be used in a similar way in agriculture to track desired traits. For example, markers can be identified by screening both the traits and molecular marker genotypes of hundreds of individuals. Markers that are linked to desirable traits can then be used during breeding to select varieties with economically useful combinations of traits, even when the genes underlying the traits are not known.

Quantitative trait locus (QTL) mapping

Molecular markers can be used to identify multiple different regions of chromosomes that contain genes that act together to produce complex traits. This process involves finding combinations of alleles of molecular markers that are correlated with a quantitative phenotype such as body mass, height, or intelligence. QTL mapping is described in more detail in the following section.

This page titled 15.4: Applications of Molecular Markers is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.4: Applications of Molecular Markers** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

15.5: Quantitative Trait Locus (QTL) Analysis

Most of the phenotypic traits commonly used in introductory genetics are qualitative, meaning that the phenotype exists in only two (or possibly a few more) discrete, alternative forms, such as either purple or white flowers, or red or white eyes. These qualitative traits are therefore said to exhibit **discrete variation**. On the other hand, many interesting and important traits exhibit **continuous variation**; these exhibit a continuous range of phenotypes that are usually measured quantitatively, such as intelligence, body mass, blood pressure in animals (including humans), and yield, water use, or vitamin content in crops. Traits with continuous variation are often complex, and do not show the simple Mendelian segregation ratios (e.g. 3:1) observed with some qualitative traits. Many complex traits are also influenced heavily by the environment. Nevertheless, complex traits can often be shown to have a component that is heritable, and which must therefore involve one or more genes.

How can genes, which are inherited (in the case of a diploid) as at most two variants each, explain the wide range of continuous variation observed for many traits? The lack of an immediately obvious explanation to this question was one of the early objections to Mendel's explanation of the mechanisms of heredity. However, upon further consideration, it becomes clear that the more loci that contribute to trait, the more phenotypic classes may be observed for that trait (Figure 15.5.1).

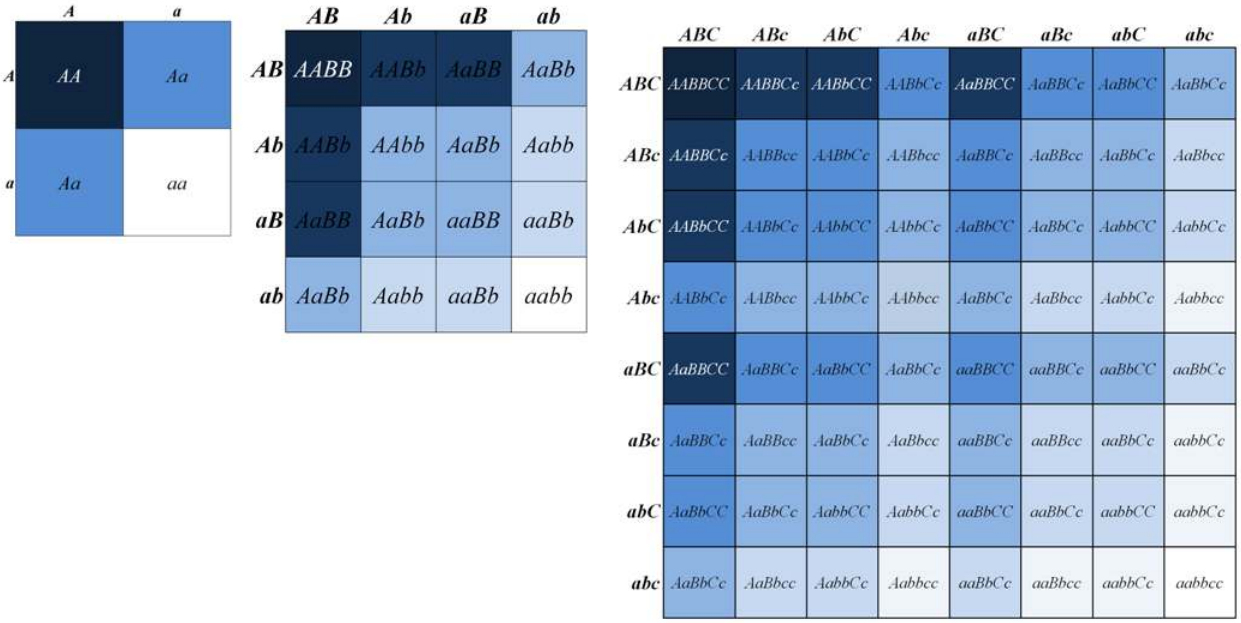


Figure 15.5.1: Punnett Squares for one, two, or three loci. We are using a simplified example of up to three semi-dominant genes, and in each case the effect on the phenotype is additive, meaning the more “upper case” alleles present, the stronger the phenotype. Comparison of the Punnett Squares and the associated phenotypes shows that under these conditions, the larger the number of genes that affect a trait, the more intermediate phenotypic classes that will be expected. (Original-Deyholos-CC:AN)

If the number of phenotypic classes is sufficiently large (as with three or more loci), individual classes may become indistinguishable from each other (particularly when environmental effects are included), and the phenotype appears as a continuous variation (Figure 15.5.2). Thus, quantitative traits are sometimes called **polygenic traits**, because it is assumed that their phenotypes are controlled by the combined activity of many genes. Note that this does not imply that each of the individual genes has an equal influence on a polygenic trait – some may have major effect, while others only minor. Furthermore, any single gene may influence more than one trait, whether these traits are quantitative or qualitative traits.

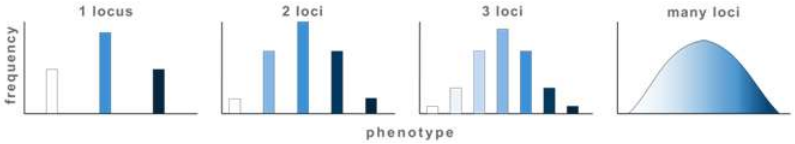


Figure 15.5.2: The more loci that affect a trait, the larger the number of phenotypic classes that can be expected. For some traits, the number of contributing loci is so large that the phenotypic classes blend together in apparently continuous variation. (Original-Deyholos-CC:AN)

We can use molecular markers to identify at least some of the genes (those with a major influence) that affect a given quantitative trait. This is essentially an extension of the mapping techniques we have already considered for discrete traits. A QTL mapping experiment will ideally start with two pure-breeding lines that differ greatly from each other in respect to one or more quantitative traits (Figure 15.5.3). The parents and all of their progeny should be raised under as close to the same environmental conditions as possible, to ensure that observed variation is due to genetic rather than external environmental factors. These parental lines must also be polymorphic for a large number of molecular loci, meaning that they must have different alleles from each other at hundreds of loci. The parental lines are crossed, and then this F_1 individual, in which recombination between parental chromosomes has occurred is self-fertilized (or back-crossed). Because of recombination (both crossing over and independent assortment), each of the F_2 individuals will contain a different combination of molecular markers, and also a different combination of alleles for the genes that control the quantitative trait of interest (Table 15.5.1).

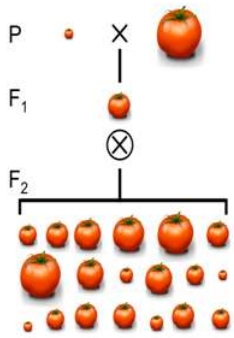


Figure 15.5.3: Strategy for a typical QTL mapping experiment. Two parents that differ in a quantitative trait (e.g. fruit mass) are crossed, and the F_1 is self-fertilized (as shown by the cross-in-circle symbol). The F_2 progeny will show a range of quantitative values for the trait. The task is then to identify alleles of markers from one parent that are strongly correlated with the quantitative trait. For example, markers from the large-fruit parent that are always present in large-fruit F_2 individuals (but never in small-fruit individuals) are likely linked to loci that control fruit mass.

Table 15.5.1 Genotypes and quantitative data for some individuals from the crosses shown in Figure 15.5.8

	genotype	fruit mass
P	$A_1A_1B_1B_1C_1C_1D_1D_1E_1E_1F_1F_1G_1G_1H_1H_1J_1J_1K_1K_1$	10g
P	$A_2A_2B_2B_2C_2C_2D_2D_2E_2E_2F_2F_2G_2G_2H_2H_2J_2J_2K_2K_2$	90g
F_1	$A_1A_2B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	50g
F_2 #001	$A_1A_1B_1B_2C_1C_1D_2D_2E_1E_2F_1F_2G_1G_2H_1H_1J_1J_2K_1K_2$	80g
F_2 #002	$A_1A_2B_1B_2C_1C_2D_1D_1E_1E_2F_1F_2G_2G_2H_1H_2J_2J_2K_1K_1$	10g
F_2 #003	$A_2A_2B_1B_2C_2C_2D_1D_2E_1E_2F_1F_2G_1G_1H_1H_2J_1J_2K_1K_2$	50g
F_2 #004	$A_1A_2B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_2K_2$	60g
F_2 #005	$A_1A_2B_1B_1C_1C_2D_2D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_1K_2K_2$	90g
F_2 #006	$A_1A_2B_2B_2C_1C_2D_1D_2E_1E_2F_1F_2G_2G_2H_1H_2J_1J_2K_1K_2$	60g
F_2 #007	$A_2A_2B_1B_1C_1C_2D_2D_2E_1E_2F_1F_2G_1G_1H_1H_2J_1J_1K_1K_2$	80g
F_2 #008	$A_1A_1B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	50g
F_2 #009	$A_1A_2B_1B_2C_2C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	50g
F_2 #010	$A_1A_2B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_2K_2$	30g
F_2 #011	$A_1A_2B_1B_2C_1C_2D_2D_2E_1E_1F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	80g
F_2 #012	$A_1A_1B_1B_2C_1C_2D_1D_1E_1E_2F_2F_2G_1G_2H_1H_2J_1J_2K_2K_2$	30g
F_2 #013	$A_2A_2B_1B_1C_1C_2D_1D_1E_1E_2F_1F_1G_1G_2H_2H_2J_1J_1K_1K_1$	10g
F_2 #014	$A_2A_2B_1B_1C_1C_1D_2D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_1K_1K_1$	70g
F_2 #015	$A_2A_2B_2B_2C_1C_2D_1D_2E_1E_2F_2F_2G_1G_2H_1H_1J_2J_2K_1K_2$	40g
F_2 #016	$A_1A_2B_2B_2C_1C_2D_1D_1E_1E_2F_1F_1G_2G_2H_1H_1J_1J_2K_1K_1$	10g
F_2 #017	$A_1A_2B_2B_2C_1C_2D_2D_2E_2E_2F_1F_1G_2G_2H_1H_2J_1J_2K_2K_2$	90g
F_2 #018	$A_1A_2B_2B_2C_1C_2D_1D_2E_1E_2F_1F_1G_2G_2H_1H_2J_1J_2K_1K_1$	40g
F_2 #019	$A_1A_1B_1B_2C_1C_2D_1D_1E_1E_2F_2F_2G_1G_1H_1H_1J_1J_2K_1K_2$	20g
F_2 #100	$A_1A_1B_1B_2C_1C_2D_2D_2E_1E_2F_1F_2G_2G_2H_1H_2J_2J_2K_1K_2$	80g

By comparing the molecular marker genotypes of several hundred F_2 individuals with their quantitative phenotypes, a researcher can identify molecular markers for which the presence of particular alleles is always associated with extreme values of the trait. In this way, regions of chromosomes that contain genes that contribute to quantitative traits can be identified. (Figure 15.5.4) It then

takes much more work (further mapping and other experimentation) to identify the individual genes in each of the regions that control the quantitative trait.

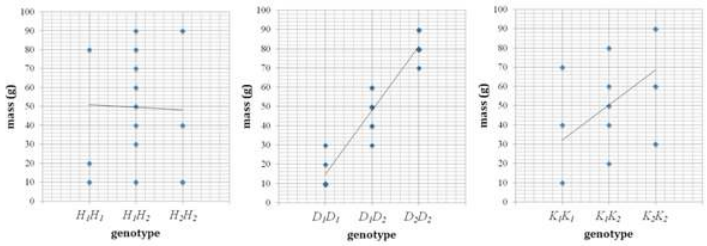


Figure 15.5.4: Plots of fruit mass and genotype for selected loci from Table 15.5.1. For most loci (e.g. H), the genotype shows no significant correlation with fruit weight. However, for some molecular markers, the genotype will be highly correlated with fruit weight. Both D and K influence fruit weight, but the effect of genotype at locus D is larger than at locus K. (Original-Deyholos-CC:AN)

This page titled 15.5: Quantitative Trait Locus (QTL) Analysis is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.5: Quantitative Trait Locus (QTL) Analysis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

15.E: Molecular Markers and Quantitative Traits (Exercises)

These are homework exercises to accompany Nickle and Barrette-Ng's "Online Open Genetics" TextMap. Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Study Questions:

10.1 Three different polymorphisms have been identified at a particular molecular marker locus. A single pair of PCR primers will amplify either a 50bp fragment (*B2*), a 60bp fragment (*B3*), or a 100bp fragment (*B4*).

Draw the PCR bands that would be expected if these primers were used to amplify DNA from individuals with each of the following genotypes:

a) B_2B_2

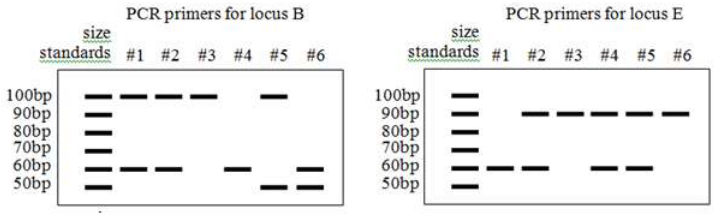
b) B_4B_4

c) B_2B_3

d) B_2B_4

10.2 In addition to the primers used to genotype locus *B* (described above), a separate pair of primers can amplify another polymorphic SSR locus *E*, with either a 60bp product (E_1) or a 90bp (E_2) product. DNA was extracted from six individuals (#1-#6), and DNA from each individual was used as a template in separate PCR reactions with primers for either locus *B* or primers for locus *E*, and the PCR products were visualized on electrophoretic gels as shown below.

Based on the following PCR banding patterns, what is the full genotype of each of the six individuals?



10.3 Based on the genotypes you recorded in Question 10.2, can you determine which of the individuals could be a parent of individual #1?

10.4 Here is part of the DNA sequence of a chromosome:

```
TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAAGTTGTGA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATTTCCTTAGTTAATGAAGACACACACACACACACACACACACAAGAATCAACAAATTCAAAATTCAACACT
```

Identify the following features on the sequence:

a) the region of the fragment that is most likely to be polymorphic

b) any simple sequence repeats

c) the best target sites for PCR primers that could be used to detect polymorphisms in the length of the simple sequence repeat region in different individuals

10.5 In a particular diploid plant, seed color is a polygenic trait. If true-breeding plants that produce red seeds are crossed with true breeding plants that produce white seeds, the F_1 produces seeds that are intermediate in color (i.e. pink). When an F_1 plant self-fertilizes, white seeds are observed in the next generation. How many genes are involved in seed color for each of the following frequencies of white seeds in the F_2 generation?

a) 1/4 white seeds

b) 1/16 white seeds

c) 1/64 white seeds

d) 1/256 white seeds

10.6 If height in humans is a polygenic trait, explain why it occasionally happens that two tall parents have a child who grows up to be much shorter than either of them.

10.7 In quantitative trait (QTL) mapping, researchers cross two parents that differ in expression of some quantitative trait, then allow chromosomes from these parents to recombine randomly, and after several generations of inbreeding, produce a large number of offspring ("recombinant inbred lines"). Because the position of crossovers is random, each of the offspring contain a different combination of chromosomal regions from each of the two parents. The researchers then use molecular markers to determine which chromosomal regions have the greatest influence on the quantitative trait, e.g. in tall offspring, which chromosomal regions always come from the tall parent?

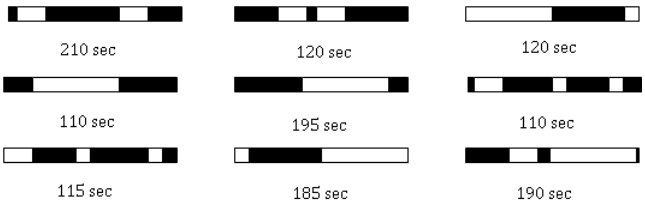
Imagine that two mice strains have been identified that differ in the time required to complete a maze, which may be an indication of intelligence. The time for maze completion is heritable and these parental strains "breed true" for the same completion time in each generation. Imagine also that their chromosomes are different colors and we can track the inheritance of chromosomal regions from each parent based on this color.

Based on the following diagrams of one chromosome from each individual in a pedigree, identify a chromosomal region that may contain a gene that affects time to complete a maze. The time for each individual is shown below each chromosome. Assume that all individuals are homozygous for all loci.

Parents:



Selected individuals from among F8 progeny of the above parents:



10.8 In a more realistic situation (as compared to question 7), where you could not distinguish the parental origin of different chromosomal regions just by appearance of chromosomes, explain how you could identify which parent was the source of a particular region of a chromosome in recombinant offspring.

This page titled 15.E: Molecular Markers and Quantitative Traits (Exercises) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

15.S: Molecular Markers and Quantitative Traits (Summary)

- Natural variations in the length or identity of DNA sequences occur at millions of locations throughout most genomes.
- DNA polymorphisms are often neutral, but because of linkage may be used as molecular markers to identify regions of genomes that contain genes of interest.
- Molecular markers are useful because of their neutrality, co-dominance, density, allele frequencies, ease of detection, and expression in all tissues.
- Molecular markers can be used for any application in which the identity of two DNA samples is to be compared, or when a particular region of a chromosome is to be correlated with inheritance of a trait.
- Many important traits show continuous, rather than discrete variation.These are also called quantitative traits.
- Many quantitative traits are influenced by a combination of environment and genetics.
- The heritable component of quantitative traits can best be studied under controlled conditions, with pure-breeding parents that are polymorphic for both a quantitative trait and a large number of molecular markers.
- Molecular markers can be identified for which specific alleles are tightly correlated with the quantitative value of a particular phenotype.The genes that are linked to these markers can be identified through subsequent research.

This page titled 15.S: Molecular Markers and Quantitative Traits (Summary) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **10.S: Molecular Markers and Quantitative Traits (Summary)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

16: Appendix 1 – Simple Nomenclature

You have probably encountered symbols in high school genetics and perhaps in previous chapters assigned to you in this course. Chapter 1 points out that a position on a chromosome is a gene locus, which simply means the location at which a gene is found. At that position, the gene could be one of several variants that we call alleles. The set of alleles comprise an organism's genotype, which might affect the organism's phenotype through the action of the genotype.

When you are solving genetics problems you have to keep track of different alleles of one or more genes. This can become confusing, so understanding how a gene works, how alleles vary, how they are expressed, and the influence they have on phenotype is important. A consistent nomenclature system – a set of rules dictating how you name alleles – becomes immensely helpful.

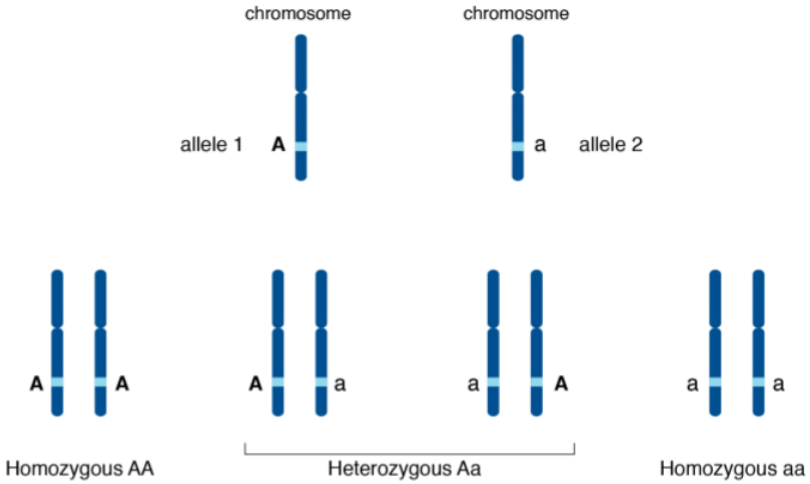


Figure A1.1: *Two chromosomes with the same gene but different alleles, A and a. (Image used with permission of www.genome.gov provided it includes the link http://xoax.net/mediaGallery/Allele-dat342)*

16.1: Textbook Introduction
16.2: Why Use Nomenclature?
16.3: Biochemistry of Gene Behavior
16.4: Basic Nomenclature
16.5: Three-Letter Symbols
16.6: Linked Genes
16.7: Summary

Contributors

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.

This page titled 16: Appendix 1 – Simple Nomenclature is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

16.1: Textbook Introduction

The first edition of this book was produced in January, 2009 as instructional material for students in Biology 207 at the University of Alberta, and is released to the public for non-commercial use under the Creative Commons License (See below). Users are encouraged to make modifications and improvements to the book.

All text in the original edition was written by Michael Deyholos, Ph.D. Chapter 9 and portions of Chapters 2 and 3 in this edition were written by Mike Harrington, Ph.D. Additional content and editing by John Locke, Ph.D. and Mark Wolansky, M.Sc. Photos and some diagrams were obtained from various, non-copyrighted sources, including Flickr, Wikipedia, Public Library of Science, and Wikimedia Commons. Photo attributions are listed in the caption with each image. This work is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 2.5 Canada License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/2.5/ca/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Under the terms of the license, you are free:

- **to Share** — to copy, distribute and transmit the work
- **to Remix** — to adapt the work

Under the following conditions:

- **Attribution.** You must include the name of the original author in books or excerpts derived from this work.
- **Non-commercial.** You may not use this work for commercial purposes.
- **Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.
- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this license.

Image, Photo and Figure License Details:

The source of each figure is given at the end of the legend and includes the following:

- CC: AD Creative Commons Attribution-No Derivative Works 2.0 Generic
- CC: AN Creative Commons Attribution-Noncommercial 2.0 Generic
- CC: AND Creative Commons Attribution-Noncommercial-No Derivative Works 2.0 Generic
- CC: AS Creative Commons Attribution-Share Alike 2.0 Generic
- EDU Educational use explicitly allowed by author, who retains copyright
- GFDL Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License
- PD Public Domain because it was created by a US government agency or because the author has explicitly released it into the public domain.

Bibliography

1. Alberts, B. et al. 2004. "Molecular Biology of the Cell, fourth edition". Garland Science, New York
2. Felix Ratjen, Gerd Döring, Cystic fibrosis, The Lancet, Volume 361, Issue 9358, 22 February 2003, Pages 681-689
3. Francis, Richard C. “Epigenetics”. 2011, Norton, NewYork.
4. Griffiths, A. J. F. et al. 2005. "Introduction to Genetic Analysis, eighth edition." W. H. Freeman and Company, New York
5. King, R. C. and W. D. Stansfield. 1997. "A Dictionary of Genetics, fifth edition". Oxford University Press, Toronto
6. Lap-Chee Tsui and Ruslan Dorfman The Cystic Fibrosis Gene: A Molecular Genetic Perspective Cold Spring Harb Perspect, February 2013;3
7. MendelWeb.
8. R. B. Blumberg, {August 1, 2012}. World Wide Web URL: www.mendelweb.org/
9. Online Mendelian Inheritance in Man, OMIM®.

10. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), {August 1, 2012}. World Wide Web URL: http://omim.org/
11. Watson, J. D. et al. 2008. "Molecular Biology of the Gene, sixth edition" Pearson Education, Inc., San Franciso
12. Your Genes, Your Health
13. Dolan DNA Learning Center / Cold Spring Harbor Laboratory, {August 1, 2012}. World Wide Web URL: www.ygyh.org/

This page titled 16.1: Textbook Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **Textbook Introduction** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.2: Why Use Nomenclature?

Word and practical problems in biology can get confusing in a hurry, particularly if you're distracted by something like exam stress! When you are manipulating several ideas, it is good practice to be thoughtful and follow rules that keep you consistent in your interpretation. As an instructor, I have seen work in which the student clearly got flustered and forgot that the mutation he or she was working on was dominant. This often leads to an answer that is inconsistent with the data.

There are a few simple rules we can use for nomenclature. This first appendix uses a simplified system to communicate allele characteristics. The next appendix will show a more complicated system that carries even more information in the gene symbols.

This page titled 16.2: Why Use Nomenclature? is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **14.1: Why Use Nomenclature?** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.3: Biochemistry of Gene Behavior

The Mechanics of Gene Expression

We often think of genes as “made of DNA”; they reside in the nucleus and endosymbiotic organelles of eukaryotes or the nucleoid region of prokaryotes. As described in Chapter 13, they are transcribed into an RNA message by RNA polymerase then interpreted by ribosomes that assemble particular amino acids into a polypeptide strand (also known as a protein) based on the sequence of nucleotides. In a cell, proteins can act as enzymes, structural features, pigments, and a host of other functions, including regulating the expression of other genes. This expression of genes leads to how an organism looks – its phenotype.

Dominant and Recessive; Homozygous, Heterozygous, and Hemizygous

Alleles themselves do not directly exert an effect on the phenotype of an organism. You will recall from Chapter 3 that genes are instructions, often for proteins. It is the effect of the protein that causes an organism to take on particular traits. In this chapter we will look at how to design symbols appropriate to communicate the characteristics of alleles you want to investigate, but keep in mind that the allele instructs what kind of protein to make.

Chapter 1 points out that organisms usually fall into the classes of being diploid or haploid. Humans and eukaryotic genetic systems usually assume the organism is diploid, which means that most chromosomes are represented as pairs. Each pair has homologous loci: the term *homologous* means “the same information” and refers to the gene at each locus. Note that although the loci, or genes, are the same, the alleles that comprise them may be different!

If an organism is **diploid** and **homozygous** for an allele (Figure A1.2, left and right), the gene at the same position of the homologous chromosomes is the same allele. Only one type of protein is made. If an organism is diploid and **heterozygous** (Figure A1.2, middle), and the protein from one allele influences the phenotype more than the protein from the other allele, we use the terms **dominant** and **recessive**, respectively.

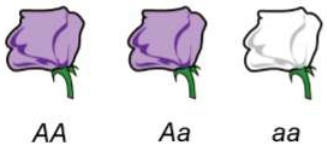


Figure A1.2: *Relationship between genotype and phenotype for an allele that is completely dominant to another allele. (Original-M. Deyholos -CC:AN – from chapter 13)*

If one copy of an allele makes enough protein to compensate for the absence of protein from the other allele in a heterozygote, it will influence the phenotype. If this phenotype looks identical to that of an organism homozygous for the “functional” allele, we consider the “functional” allele to be dominant to the “null”. We could also say the “null” is recessive to the “functional” allele. Keep in mind that the alleles themselves aren’t doing anything, but it’s common practice to label the alleles as dominant or recessive, although in reality we are talking about the expression of those alleles.

Why this is important is how the proteins from two alleles interact. If both proteins are identical (from a homozygous genotype) the phenotype that results will be that of the action of one “type” of protein, even if though there are two copies of the gene – they are both the same allele. There’s no real interaction because the proteins do the same thing. It’s in heterozygotes that we can see whether a particular allele is dominant over another or otherwise influences the phenotype in an interesting way.

Here’s another point about the “normal” allele: it was honed by natural selection over a long period of time. Genes are instructions for the protein tools of an organism’s cells. For this reason we often call the “normal” allele of a gene the “wild type” allele. This would be the allele most common in the wild, presumably because it provides a benefit to the organism. Thus, most mutations are likely to reduce the effectiveness of the wild type allele, although the process of evolution allows (and, in fact, requires) an occasional beneficial allele to permeate a population if it provides a selective advantage.

Note that we do not use the terms homozygous or heterozygous in haploid organisms. If they are haploid, their phenotype will reflect the genotype of the only allele present. The proper term for their genotype is hemizygous to reflect only one copy of each gene.

There are Many Kinds of Alleles for a Gene

A mutation is a change in nucleotide sequence. Chapter 11 goes into more detail of what this means . What's important now is that you understand that the amino acids of a protein can be different if we compare different alleles of a gene and they may behave differently – often one protein will "work better" than the other. If the promoter of a "functional" allele of a gene is damaged, the allele that is created might not even create an mRNA so no protein will be encoded by that allele. This is called an amorphic or "null" mutant (See Chapter 1).

Keep also in mind that a gene can be mutated in different parts of its sequence to create different alleles. A diploid organism can have a maximum of two alleles (aside from gene duplication or abnormal chromosome structure, but ignore that for now). In a population, though, there can be many, many different alleles. Perhaps allele one decreases protein function and allele 2 is even less effective. This describes an **allelic series** as follows: wild--type > allele 1 > allele 2 > any null allele. Wild--type alleles in this case encode the most effective protein. Null alleles represent catastrophic mutations that eliminate transcription or produces proteins that can't function at all. See section 3.2 for another example of an allelic series.

Gender and Gene Interactions

Finally, here are a couple points about other interactions before we move on to gene symbols. Some organisms have a pair of sex chromosomes that dictate gender (see Chapter 4). We'll ignore incorporating ways to communicate if a chromosome is a sex chromosome in this chapter; Appendix 2 will show how we can indicate those using a symbol.

This page titled 16.3: Biochemistry of Gene Behavior is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **14.2: Biochemistry of Gene Behavior** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.4: Basic Nomenclature

Single Letter System

Sometimes what you want to do is a little rough work for investigating your genetic model. A genetic model is a diagram of the logic that you propose for a particular type of inheritance. For instance, if you cross a true-breeding purple plant with a true-breeding white plant (Figure A1.2), you will get a heterozygote (the middle plant). If we name the gene after the mutation (*a* is the first letter in "albino"), we know that the heterozygote will have one capital letter "*A*" and a lower case "*a*". The heterozygote is the F_1 generation ("first filial", which means it's the first child from parents that are crossed). The F_1 is purple, which means the "*a*" allele is recessive; only one copy of the "*A*" allele is needed for enough purple pigment to make it identical to one true-breeding parent. This is complete dominance.

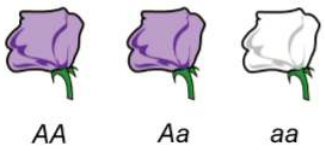


Figure A1.2

We cannot know from the information given which allele is wild type or mutant. One hypothesis, though, is that purple pigments are required to attract pollinators and therefore would help the plant in the wild. Albino plants could be a mutant and might not generate as many seeds for lack of pollinators. If the context of your genetics problem doesn't indicate which allele is wild-type, it's good practice to name your allele based on the recessive trait. Often the recessive allele is the mutation. We might consider that the "*a*" allele is null and makes no pigment at all. One or two "*A*" alleles make enough protein to cause the plant to be purple.

We can't know from the information given which allele is wild type or mutant. One hypothesis, though, is that purple pigments are required to attract pollinators and therefore would help the plant in the wild. Albino plants could be a mutant and might not generate as many seeds for lack of pollinators. If the context of your genetics problem doesn't indicate which allele is wild-type, it's good practice to name your allele based on the recessive trait. Often the recessive allele is the mutation. We might consider that the "a" allele is null and makes no pigment at all. One or two "A" alleles make enough protein to cause the plant to be purple.

Figure A1.2 already assumes that the capital letter (*A*) stands for an allele that encodes a protein for purple pigment and the recessive allele (*a*) doesn't make pigment. Thus the *Aa* heterozygote is sufficient evidence to adopt upper- and lower-case letter "*A*"s to communicate the characteristics of purple and white alleles.

A note of caution. When you're writing down gene symbols for homework or on an exam, be sure to make the characters distinct. A typewritten "*y*" is easy to distinguish from the upper case "*Y*" but not as easy when writing it down. Instructors who ask you to show your work need to be able to follow your logic train. More important than that is YOU have to be able to follow your own reasoning. Students often switch symbols and come up with an answer that is inconsistent with the data given because of this. Consider underlining your capitals or putting a line through one of them to make it distinct (e.g. $\underline{Y}$ for the dominant allele; *y* for the recessive).

Name the gene after the mutant phenotype

Some instructors would accept "*P*" for "*purple*" for the previous cross. However, the better answer is to follow an established system. During "exam fog", it's easy to get lost if you are inconsistent with how you develop your symbols. During your study period and when you're practicing genetics problems, be thoughtful about the gene names you choose.

Let's always choose a letter based on the mutant phenotype for our gene symbol. If we are presented with a ladybug mutant that is small, we might choose "*d*" for "*dwarf*". Geneticists sometimes set up a research program based on unusual phenotypes of the organism they are studying. The fact that a mutant phenotype that is heritable exists tells us that there is a genetic control for the trait and that it might be isolated in the lab. When you look at your classmates, you don't necessarily note that none of them has an arm growing from the tops of their heads. If one student had this trait, however, you couldn't help but notice it. If you found out it showed up in that student's ancestry in a predictable fashion, you might reasonably suggest that there is a genetic basis for that. If it happened to be controlled by a single gene, you might call the gene "*extra arm*" or "*arm head*". If it happened to be a dominant trait, you might use the letter "$\not{X}$" or "*A*" for the mutant allele. The wild type allele would be "*x*" or "*a*", respectively.

you don't know which is mutant, use the recessive trait for the gene name

What if you don't know which allele is mutant? What if you're presented with two true-- breeding frogs: one that is gold and one that is yellow. If you don't know what the predominant colour in nature is you can't know which one is mutant. If you crossed them and all the progeny are gold, then you know the dominant allele encodes a protein to make it *gold*. The recessive, therefore, is "*yellow*" and you should name the gene "*y*" after the recessive phenotype. This means the dominant allele would be "*Y*". Your offspring would therefore be *Yy* and the gold parent would be *YY*. The yellow parent would be *yy*.

Apply these ideas at the Online Open Genetics exercises.

1.3 One Letter System

Superscripts

Sometimes a letter is used as the name of a gene, and superscripts can modify it to indicate the different alleles. One common single letter code for an allelic series is "*I*". Red blood cells can have their cell membranes modified by sugar tags that give rise to our blood type. One allele of *I* gives rise to blood type A and is therefore called I^A. An enzyme encoded by I^B modifies sugars to create blood type B. A heterozygote I^AI^B demonstrates *both* sugar tags because those alleles are expressed – they are codominant. People with blood type O only possess alleles for the *I* gene that don't work and are therefore recessive – they don't modify the extracellular sugar tags. Because it is recessive, individuals are homozygous for *i*: they are *ii*. Chapter 13, Section 7 has more detail on this allelic series.

Sometimes a superscript "plus sign" is used to denote the wild type allele. One might use the symbol W^+ to indicate a wild--type allele that promotes wing growth. Note that the generic "wing" gene name isn't a best practice – name the gene after the mutant phenotype! A wingless mutant would be W^-. You should never use a "+" and shift the case of the letter unless you are dealing with a special case such as the codominance in the blood type example above. The capital "*I*" letter indicates it is dominant to "*i*". The superscript A and B for the codominant alleles indicate the dominant alleles are different from each other.

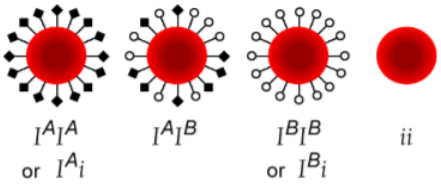


Figure A1.3: Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The I^A allele is completely dominant to the i allele. The I^B allele is completely dominant to the i allele. (Original-Deholos -CC:AN – from Chapter 1)

Superscripts can be symbols, a single letter, or many letters. They modify the gene name only in the superscripted symbols: the regular--sized letters are identical between them (see Table 1). This means that Abc^+ and *abc* would be different genes (*i.e.* not allelic); Abc^+ and *Abc* are alleles, as are abc^+ and *abc*. Note that a superscript is not mandatory for all alleles of that gene, depending on the convention.

Alleles of bacterial genes are typically indicated with a superscript + or -. For example, a bacterial allele that creates an enzyme that makes methionine would be met^+, and a defective allele of that gene is met^-.

Table 1: Examples of genes using a superscript modifier.

white^{apr} *or white*^{apricot}	An allele of the *white* gene which has an “apricot” phenotype
Abc^{+} *Abc*	Two alleles for the *Abc* gene (wild-type and mutant, respectively). Note the mutant allele is dominant.
w^{+}; w^{a}; w	Three alleles in a series for the w gene. The first is wild type; the second two are different mutant alleles.
bio^{+}; bio^{-}	A wild-type allele of a biotin gene and its recessive counterpart. This is likely a bacterial gene due to the convention.

This page titled 16.4: Basic Nomenclature is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **14.3: Basic Nomenclature** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.5: Three-Letter Symbols

It's perfectly acceptable to use a single letter or even two letters. Sometimes, though, multiple traits spelled with the same first letter can get confusing. Using three letters for a gene symbol can make it easier to remember what the letters stand for. In fact, for some model systems those who study them adopt a defined nomenclature system. A plant often used for genetic studies, called *Arabidopsis thaliana*, has a three-letter code (https://www.arabidopsis.org/portals/nomenclature/namerule.jsp).

Just as we saw for the one-letter symbols, the dominant allele has the first letter capitalized and the last two letters are lower-case. Recessive alleles are all lower-case. With three letters, you can make gene names that are easier to keep track of. For example, you might see a fly with an extra set of wings. Instead of calling it "*w*" for "wings" (which is a poor choice because it doesn't represent the mutant phenotype), you can instead call it "*exw*" for "*ex*tra *w*ings". Then, when you see it, you can sound out the abbreviation and remember that it stands for the mutation. If the wild-type allele for this is dominant, then you would write that one "*Exw*". The first letter indicates that it is the dominant allele. What you can't tell just from these examples is that wild-type allele is dominant! Appendix 2 will go into an extension of this system so you can add that information to the gene symbol. *Hint*: remember in Section 3.2 that the "+" superscript indicates the wild-type allele.



1.4 Three Letter System

This page titled 16.5: Three-Letter Symbols is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **14.4: Three-Letter Symbols** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.6: Linked Genes

Mendel was lucky. He studied a variety of traits in pea plants and his data were consistent with his idea of traits being encoded by pairs of discrete heritable units. He did not call these “genes” and had no idea about their chromosomal origins or chemical makeup. It turns out that the genes he studied were either on different chromosomes and so assorted independently (Chapter 9), or so far apart on the same chromosome that linkage could not be detected (Chapter 10).

Symbols for a gene can be drawn on a page to communicate their position on a chromosome. To do this, we use a forward slash (/) to demonstrate what is on each chromosome. Figure A1.4 shows how we might conceptualize the position of a gene on two chromosomes by collapsing the chromosomes into a single line.

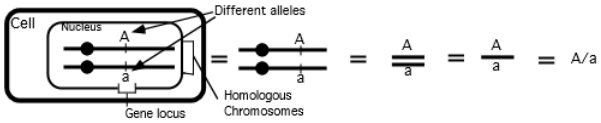


Figure A1.4: A diagram of how chromosomes, loci and alleles look in the cell, and how we depict them in written form. Note that only one gene is represented in this figure and chromosomes have not replicated. *(Original-J.Locke- CC BY-NC 3.0 from Chapter 10)*

There’s no question about where the gene is located when only one trait is under investigation: it will be at the same position on each homolog. Two genes, however, can be one of three possibilities. Each possibility has implications for gene mapping and predicting ratios from a dihybrid cross. Figure A1.5 shows the positions of genes for an unlinked situation as well as linked genes in coupling and repulsion configurations. If genes are unlinked, put the allele symbols for one gene on either side of one slash followed by a semicolon (indicating that it’s unlinked) and the other gene with the alleles separated by a second slash (*A/a*; *B/b*). When genes are linked, only one slash is used: remember, the slash stands for a pair of homologous chromosomes. Genes in coupling would have the dominant genes together on one side of the slash and recessives on the other side (*AB/ab*). Repulsion would represent the other arrangement (*Ab/aB*).

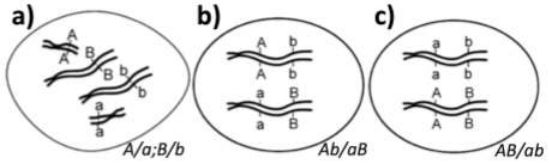


Figure A1.5: Three gene arrangements for cells of genotype AaBb. Chromosomes are replicated (shown with sister chromatids). a) demonstrates unlinked genes: A/a; B/b. Linked genes are shown in b) repulsion Ab/aB and c) coupling AB/ab. *(Modified by T. Nickle from an original by J.Locke- CC BY-NC 3.0 from Chapter 18)*

Practice your skills with identifying linked and unlinked genes online in module 1.5. Some examples of different forms of gene symbols are shown in Table 2. Keep in mind that sometimes you have flexibility in which system of nomenclature you use, but sometimes it is dictated to you, for example in publications or other formal submissions. You are discouraged from inventing your own system or mixing up different systems because it will confuse your readers (or graders!).

Table 2: Examples of symbols used to represent genes and alleles.

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used in publications because not all alleles show complete dominance and many genes have more than two alleles. It's quick and easy for you to use when working out genetics problems when you are sure each gene involves only two alleles.
a^+ and *a*	Superscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +. The mutant allele of gene a would be recessive.
met^+ and met^-	This is typical of a prokaryote gene symbol. It could be referring to wild-type (functional) and mutant (nonfunctional) alleles of a gene that makes a protein in the methionine synthesis pathway.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene, but on homologous chromosomes. Both representations in this row are identical: it represents a homozygous dominant.
Aa/Aa or *Aa/aa*	Note that this example shows two alleles of the gene *Aa*. We know that the gene symbol is two letters because the slash separates the allele found on each of the homologous chromosomes. We cannot tell if the mutant phenotype is recessive because there's no indication which is wild type.
Grn^+ *shr/Grn shr*	The three-letter system is used here. "*Grn*" might mean that the phenotype is "green", but we can't be sure. What we do know is that the mutant allele codes for a protein leading to a dominant phenotype. The wild-type allele must be recessive to the mutant allele. Maybe "*shr*" means "shrunken" or "short", but we know that the mutant phenotype can only be seen in the homozygous recessive configuration. The phenotype for this organism is mutant for both *Grn* and *shr* traits. Final note: the genes are on the same chromosome based on the position of the slash.
bob^+*/bob*; *mia/mia*	This also uses the three-letter system. The organism is heterozygous for bob but shows the wild-type trait in its phenotype. It is homozygous recessive for *mia* and therefore shows that mutant phenotype. The genes are unlinked.

A more advanced system of nomenclature is outlined in Appendix 2. New rules are introduced to help you identify sex linked genes and predict phenotypes from gene symbols alone.

This page titled 16.6: Linked Genes is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **14.5: Linked Genes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.7: Summary

- Symbols are used to denote the alleles, or genotype, of a locus.
- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- In a diploid organism, the alleles can be homozygous, heterozygous or hemizygous.
- Allelic interactions at a locus can be described as dominant vs. recessive, incomplete dominance, or co-dominance.
- Two different genes can be on the same chromosome (linked) or on different ones (unlinked)
- Steps:
  - Can you identify the mutant? Name the gene after its phenotype
  - If you cannot tell which allele is mutant, name the gene after the recessive allele
  - Capitalize the first letter for the dominant allele; use lower-case for the recessive allele
  - If genes are linked, write the gene symbols together on each side of the slash. If genes are unlinked, they appear on different sides of a semicolon

This page titled 16.7: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Ying Liu via source content that was edited to the style and standards of the LibreTexts platform.

- **14.6: Summary** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

Index

A

abnormalities
10: Changes in Chromosome Number and Structure

ABO blood group
4.7: Relationships Between Genes, Genotypes and Phenotypes

activator
8.7.3: Prokaryotic Gene Regulation

adenocarcinoma
13.1: Classification of Cancers

adenocarcinomas
13: Cancer Genetics

agricultural genetics
14: Genomics and Systems Biology

allele
4.1: Introduction
4.5: Laws of Inheritance
4.6: Mendel's First Law
4.11: Black fur color - a dominant trait

allele_frequencies
11: Pedigrees and Populations

alleles
5: Genetic Analysis of Multiple Genes
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
9: Mutation and Variation
9.1: Mutation and Polymorphism
9.3: Genetic screening for mutations- Forward genetics
9.4: Types of Mutations
9.8: Linkage and Mapping
9.8.1: Linkage
9.8.2: Recombination
9.8.3: Linkage Reduces Recombination Frequency
9.8.5: Inferring Recombination From Genetic Data
9.8.7: Mapping With Three-Point Crosses
9.8.E: Linkage and Mapping (Exercises)
9.8.S: Linkage and Mapping (Summary)

allelic mutations
9.6: Complementation tests and Allelism

ALS
11.4: Sporadic and Non-Heritable Diseases

Alzheimer's disease
9.3: Genetic screening for mutations- Forward genetics

amino acid
9.5: Some mutations may not have detectable phenotypes

aminoacyl tRNA synthetase
8.6: Ribosomes and Protein Synthesis

amniocentesis
10.4: Diagnosing Human Chromosome Abnormalities

amplification
14.2: DNA Sequencing

analysis
11.2: Pedigree Analysis
14.1: 'Omics Technologies

anaphase
10.1: Changes in Chromosome Number

aneuploid
3.6: Errors in Meiosis
10.1: Changes in Chromosome Number

aneuploidy
2.5: Karyotypes Describe Chromosome Number and Structure
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
10.3: Chromosome Abnormalities in Humans

angiogenesis
13.3: Hallmarks of Cancer

anticodon
8.5: RNA Processing in Eukaryotes

antimorph
9.4: Types of Mutations

apoptosis
13.3: Hallmarks of Cancer
13.6: Tumor Suppressor Genes

Arabidopsis thaliana
1.6: Model Organisms Facilitate Genetic Advances

assembly
14.3: Whole Genome Sequencing

autosomal dominant
11.3: Inferring the Mode of Inheritance

autosomal recessive
11.5: Calculating Probabilities

autosome
3.6: Errors in Meiosis

autosomes
6: Sex Chromosomes Sex Linkage
6.1: Introduction
6.2: Autosomes and Sex Chromosomes
6.4: Sex Linkage- An Exception to Mendel's First Law
6.6: Chapter 10 Summary

B

BAC
14.3: Whole Genome Sequencing

bacterial colonies
1.2: DNA is the Genetic Material

bacterial strains
1.2: DNA is the Genetic Material

basal cell
13.1: Classification of Cancers

base pair changes
9.5: Some mutations may not have detectable phenotypes

benign tumors
13.2: Cancer Cell Biology

biochemical pathway
9.6: Complementation tests and Allelism

bioinformatics
14.1: 'Omics Technologies

biological processes
9.6: Complementation tests and Allelism

biological theories
3: Meiosis - Sexual Reproduction

biology
6.1: Introduction

biomolecular structure
1.3: DNA Structure

birds
6.2: Autosomes and Sex Chromosomes

blending inheritance
1.4: Genes are the Basic Units of Inheritance

BRCA1
13.6: Tumor Suppressor Genes

breeding
9.8.6: Genetic Mapping

brightfield microscopy
10.4: Diagnosing Human Chromosome Abnormalities

C

CAAT box
8.4: Eukaryotic Transcription

Caenorhabditis elegans
1.6: Model Organisms Facilitate Genetic Advances

cAMP binding protein (CAP)
8.11.1: The lac Operon

cancer
8.7.8: Cancer and Gene Regulation
10.2: Changes in Chromosome Structure
13: Cancer Genetics
13.1: Classification of Cancers
13.2: Cancer Cell Biology
13.3: Hallmarks of Cancer
13.4: Mutagens and Carcinogens
13.5: Oncogenes
13.6: Tumor Suppressor Genes
13.7: The "Poster Boy" of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)

cancer cell biology
13: Cancer Genetics

cancer classification
13: Cancer Genetics

cancer research
2.4: The Cell Cycle and Changes in DNA Content

cancers
9: Mutation and Variation

CAP
8.7.3: Prokaryotic Gene Regulation

carcinogen
13.4: Mutagens and Carcinogens

carcinoma
13.1: Classification of Cancers

carcinomas
13: Cancer Genetics

carriers
11.3: Inferring the Mode of Inheritance

catabolite activator protein
8.7.3: Prokaryotic Gene Regulation

cell biology
14: Genomics and Systems Biology

cell cycle
2.4: The Cell Cycle and Changes in DNA Content

cell division
1: DNA, Genes, Chromosomes and Genome
3.5: Comparing Meiosis and Mitosis
10: Changes in Chromosome Number and Structure
10.1: Changes in Chromosome Number
13.2: Cancer Cell Biology
13.4: Mutagens and Carcinogens
13.5: Oncogenes

cell theory
2.2: Studying Cells
cell_cycle
13.6: Tumor Suppressor Genes
centimorgans
9.8.6: Genetic Mapping
central dogma
8.2: The Genetic Code
CF
9.7: Example of Human Mutations
chapter summary
6: Sex Chromosomes Sex Linkage
chemical carcinogens
13.4: Mutagens and Carcinogens
chimera
4.13: Sex-Linkage- An Exception to Mendel's First Law
chromatin
2.3: DNA is Packaged into Chromatin
chromosomal
9.8.6: Genetic Mapping
9.8.S: Linkage and Mapping (Summary)
chromosomal abnormality
10.4: Diagnosing Human Chromosome Abnormalities
chromosomal disorder
6.5: Y-Linked Genes
chromosome
1.3: DNA Structure
9.8.3: Linkage Reduces Recombination Frequency
9.8.E: Linkage and Mapping (Exercises)
10: Changes in Chromosome Number and Structure
10.2: Changes in Chromosome Structure
10.3: Chromosome Abnormalities in Humans
chromosome abnormalities
10.1: Changes in Chromosome Number
chromosome inversion
3.6: Errors in Meiosis
chromosomes
1: DNA, Genes, Chromosomes and Genome
1.1: Introduction
2.5: Karyotypes Describe Chromosome Number and Structure
3.3: Meiosis II
3.5: Comparing Meiosis and Mitosis
6.2: Autosomes and Sex Chromosomes
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
9.8.1: Linkage
9.8.2: Recombination
9.8.4: Crossovers Allow Recombination of Linked Loci
chronic inflammation
13.4: Mutagens and Carcinogens
Chronic Myelogenous Leukemia
13.7: The "Poster Boy" of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
cloning
4.4: Punnett Squares and Test Crosses
CML
13.7: The "Poster Boy" of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
coat color
5.2: Epistasis and Other Gene Interactions
codominance
4.7: Relationships Between Genes, Genotypes and Phenotypes
4.16: Extensions of the Laws of Inheritance
codon
8.2: The Genetic Code
cohesin
3.2: Meiosis I
colinear
8.2: The Genetic Code
complementation tests
9.6: Complementation tests and Allelism
complex inheritance
11.4: Sporadic and Non-Heritable Diseases
consensus
8.3: Prokaryotic Transcription
conservation biology
1.1: Introduction
constitutive mutants
8.11.2: The Use of Mutants to Study the lac Operon
contig
14.3: Whole Genome Sequencing
continuous variation
4.3: Mendel's Experiments
core enzyme
8.3: Prokaryotic Transcription
counseling
11.2: Pedigree Analysis
cri du chat
10.4: Diagnosing Human Chromosome Abnormalities
crossing over
3.2: Meiosis I
9.8.4: Crossovers Allow Recombination of Linked Loci
crossover
3.3: Meiosis II
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
9.8.3: Linkage Reduces Recombination Frequency
crossover_hot_spots
9.8.6: Genetic Mapping
crossovers
9.8.S: Linkage and Mapping (Summary)
cystic fibrosis
9: Mutation and Variation
CysticFibrosis
9.7: Example of Human Mutations
cytokinesis
3.3: Meiosis II

D

Danio rerio
1.6: Model Organisms Facilitate Genetic Advances
degeneracy
8.2: The Genetic Code
deletion
9.S: Mutation and Variation (Summary)
deoxyribonucleic acid
1: DNA, Genes, Chromosomes and Genome
detection
13.2: Cancer Cell Biology
Dicer
8.7.6: Eukaryotic Post-transcriptional Gene Regulation
dihybrid
4.5: Laws of Inheritance
dihybrid crosses
5: Genetic Analysis of Multiple Genes
5.2: Epistasis and Other Gene Interactions
dihybrids
5.1: Dihybrid Crosses
diploid
3.1: Overview of Meiosis
3.3: Meiosis II
3.5: Comparing Meiosis and Mitosis
diploids
6.2: Autosomes and Sex Chromosomes
discontinuous variation
4.3: Mendel's Experiments
disease
13.2: Cancer Cell Biology
14.4: Functional Genomics – Determining Function(s)
diseases
9: Mutation and Variation
11.2: Pedigree Analysis
11.3: Inferring the Mode of Inheritance
11.4: Sporadic and Non-Heritable Diseases
13: Cancer Genetics
dispersive model
7.4: Basics of DNA Replication
DNA
1: DNA, Genes, Chromosomes and Genome
1.1: Introduction
1.3: DNA Structure
1.5: Genes and Genomes
2.5: Karyotypes Describe Chromosome Number and Structure
7.2: Historical Basis of Modern Understanding
7.3: DNA Structure and Sequencing
9: Mutation and Variation
9.1: Mutation and Polymorphism
9.2: Origins of Mutations
9.E: Mutation and Variation (Exercises)
10: Changes in Chromosome Number and Structure
10.4: Diagnosing Human Chromosome Abnormalities
13.4: Mutagens and Carcinogens
14.2: DNA Sequencing
14.3: Whole Genome Sequencing
DNA ligation
12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
DNA methylation
8.7.8: Cancer and Gene Regulation
DNA polymerase II
7.7: DNA Repair
DNA replication
1.3: DNA Structure
7.4: Basics of DNA Replication
7.6: DNA Replication in Eukaryotes
DNA replication error
9.S: Mutation and Variation (Summary)
DNA sequencing
1.1: Introduction
14: Genomics and Systems Biology
DNA structure
7.3: DNA Structure and Sequencing

DNA synthesis
2.4: The Cell Cycle and Changes in DNA Content
DNA_repair
13.6: Tumor Suppressor Genes
dominance
4.10: Biochemical Basis of Dominance
dominant
4.3: Mendel's Experiments
4.7: Relationships Between Genes, Genotypes and Phenotypes
4.11: Black fur color - a dominant trait
dosage compensation
6.6: Chapter 10 Summary
double crossovers
9.8.7: Mapping With Three-Point Crosses
double helix
1.3: DNA Structure
Down Syndrome
10.3: Chromosome Abnormalities in Humans
10.4: Diagnosing Human Chromosome Abnormalities
downstream
8.3: Prokaryotic Transcription
Drosophila
6.4: Sex Linkage- An Exception to Mendel's First Law
6.6: Chapter 10 Summary
9.3: Genetic screening for mutations- Forward genetics
Drosophila melanogaster
1.6: Model Organisms Facilitate Genetic Advances
6.1: Introduction
drought tolerance
14: Genomics and Systems Biology
drugdevelopment
9.7: Example of Human Mutations
dysplasia
13.2: Cancer Cell Biology

E

electron microscope
2.2: Studying Cells
electrophoresis
7.3: DNA Structure and Sequencing
14.2: DNA Sequencing
enhancer
8.7.5: Eukaryotic Transcription Gene Regulation
environmental factors
11.4: Sporadic and Non-Heritable Diseases
environmental influence
9.E: Mutation and Variation (Exercises)
epidemiology
1.1: Introduction
epigenetic
8.7.2: Regulation of Gene Expression
Epigenetic mechanisms
8.7.4: Eukaryotic Epigenetic Gene Regulation
epigenetics
8.11.6: Epigenetics
episome
8.11.2: The Use of Mutants to Study the lac Operon
epistasis
4.16: Extensions of the Laws of Inheritance
5.2: Epistasis and Other Gene Interactions
epithelial
13.1: Classification of Cancers
errors
9.2: Origins of Mutations
Escherichia coli
1.6: Model Organisms Facilitate Genetic Advances
essential genes
9.5: Some mutations may not have detectable phenotypes
ethidium bromide
12.6: DNA Analysis - Gel Electrophoresis
euchromatin
2.3: DNA is Packaged into Chromatin
eukaryotes
7.6: DNA Replication in Eukaryotes
eukaryotic
11: Pedigrees and Populations
eukaryotic cells
1.5: Genes and Genomes
2.4: The Cell Cycle and Changes in DNA Content
3.5: Comparing Meiosis and Mitosis
eukaryotic organisms
3: Meiosis - Sexual Reproduction
euploid
3.6: Errors in Meiosis
10.1: Changes in Chromosome Number
evolution
1.5: Genes and Genomes
3: Meiosis - Sexual Reproduction
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
9: Mutation and Variation
9.2: Origins of Mutations
9.8: Linkage and Mapping
9.8.1: Linkage
10.2: Changes in Chromosome Structure
11: Pedigrees and Populations
exon
8.5: RNA Processing in Eukaryotes
Expressivity
4.14: Phenotypes May Not Be As Expected from the Genotype
external agents
14.4: Functional Genomics – Determining Function(s)

F

F1
4.3: Mendel's Experiments
F1 generation
1.4: Genes are the Basic Units of Inheritance
F2
4.3: Mendel's Experiments
F2 generation
1.4: Genes are the Basic Units of Inheritance
facilitates chromatin transcription
8.4: Eukaryotic Transcription
FACT
8.4: Eukaryotic Transcription
family
11.2: Pedigree Analysis
family trees
11.3: Inferring the Mode of Inheritance
fertility
10.2: Changes in Chromosome Structure
fertilization
3: Meiosis - Sexual Reproduction
3.1: Overview of Meiosis
fidelity
9.1: Mutation and Polymorphism
fish
10.4: Diagnosing Human Chromosome Abnormalities
floral scent
9.E: Mutation and Variation (Exercises)
flower color
1.4: Genes are the Basic Units of Inheritance
fluorescence in situ hybridization
10.4: Diagnosing Human Chromosome Abnormalities
forensic science
1.1: Introduction
frameshift mutation
7.7: DNA Repair
Frederick Griffith
7.2: Historical Basis of Modern Understanding
Friedrich Miescher
7.2: Historical Basis of Modern Understanding
fruit color
5.2: Epistasis and Other Gene Interactions
fruit fly
9.E: Mutation and Variation (Exercises)
functional genomics
14: Genomics and Systems Biology
14.4: Functional Genomics – Determining Function(s)

G

G1 phase
2.4: The Cell Cycle and Changes in DNA Content
G2 phase
2.4: The Cell Cycle and Changes in DNA Content
gamete
4.1: Introduction
9.8.3: Linkage Reduces Recombination Frequency
10.1: Changes in Chromosome Number
gametes
3.3: Meiosis II
9.8.2: Recombination
9.8.5: Inferring Recombination From Genetic Data
9.8.E: Linkage and Mapping (Exercises)
9.8.S: Linkage and Mapping (Summary)
10.2: Changes in Chromosome Structure
gastrointestinaldisease
9.7: Example of Human Mutations
GDP
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
gel electrophoresis
12.6: DNA Analysis - Gel Electrophoresis
gene
9.7: Example of Human Mutations
14.2: DNA Sequencing

gene expression
1.5: Genes and Genomes
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
8.7.2: Regulation of Gene Expression
14: Genomics and Systems Biology
14.4: Functional Genomics – Determining Function(s)
gene interactions
5: Genetic Analysis of Multiple Genes
5.2: Epistasis and Other Gene Interactions
gene linkage
9.8.1: Linkage
gene loci
6.1: Introduction
gene locus
11.4: Sporadic and Non-Heritable Diseases
gene naming
9.5: Some mutations may not have detectable phenotypes
gene regulation
1.3: DNA Structure
8.7.4: Eukaryotic Epigenetic Gene Regulation
8.7.8: Cancer and Gene Regulation
9.4: Types of Mutations
gene therapy
12.8: Transgenic organisms
genes
1: DNA, Genes, Chromosomes and Genome
1.1: Introduction
1.5: Genes and Genomes
9.8.7: Mapping With Three-Point Crosses
9.8.E: Linkage and Mapping (Exercises)
9.E: Mutation and Variation (Exercises)
10: Changes in Chromosome Number and Structure
11.1: Prelude to Pedigrees and Populations
13.5: Oncogenes
13.6: Tumor Suppressor Genes
genetic
9.2: Origins of Mutations
9.8.6: Genetic Mapping
10.2: Changes in Chromosome Structure
11.2: Pedigree Analysis
genetic analysis
5: Genetic Analysis of Multiple Genes
genetic causes
11.4: Sporadic and Non-Heritable Diseases
genetic counseling
11.5: Calculating Probabilities
genetic disease
13: Cancer Genetics
genetic diversity
9: Mutation and Variation
genetic factors
1.4: Genes are the Basic Units of Inheritance
genetic identity
3.5: Comparing Meiosis and Mitosis
genetic information
1.1: Introduction
genetic inheritance
1: DNA, Genes, Chromosomes and Genome
9.8.3: Linkage Reduces Recombination Frequency
genetic mapping
5: Genetic Analysis of Multiple Genes
9.8: Linkage and Mapping
9.8.1: Linkage
9.8.7: Mapping With Three-Point Crosses
genetic maps
9.8.S: Linkage and Mapping (Summary)
genetic material
1.3: DNA Structure
genetic ratios
5.2: Epistasis and Other Gene Interactions
genetic redundancy
9.5: Some mutations may not have detectable phenotypes
genetic screens
9.3: Genetic screening for mutations- Forward genetics
genetic transformation
1.2: DNA is the Genetic Material
genetic variation
3: Meiosis - Sexual Reproduction
genetic_diseases
11: Pedigrees and Populations
genetic_maps
9.8.6: Genetic Mapping
geneticists
9.6: Complementation tests and Allelism
genetics
1.4: Genes are the Basic Units of Inheritance
1.5: Genes and Genomes
1.6: Model Organisms Facilitate Genetic Advances
2.5: Karyotypes Describe Chromosome Number and Structure
3: Meiosis - Sexual Reproduction
6.1: Introduction
6.2: Autosomes and Sex Chromosomes
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
6.4: Sex Linkage- An Exception to Mendel's First Law
6.6: Chapter 10 Summary
9: Mutation and Variation
9.8: Linkage and Mapping
9.8.1: Linkage
9.8.2: Recombination
9.8.4: Crossovers Allow Recombination of Linked Loci
9.8.E: Linkage and Mapping (Exercises)
9.E: Mutation and Variation (Exercises)
10: Changes in Chromosome Number and Structure
10.3: Chromosome Abnormalities in Humans
11: Pedigrees and Populations
11.1: Prelude to Pedigrees and Populations
11.3: Inferring the Mode of Inheritance
11.5: Calculating Probabilities
genetics research
9.1: Mutation and Polymorphism
geneticvariation
3.3: Meiosis II
genome
1: DNA, Genes, Chromosomes and Genome
1.1: Introduction
9.2: Origins of Mutations
14.1: ‘Omics Technologies
14.3: Whole Genome Sequencing
genome annotation
14.3: Whole Genome Sequencing
genomes
1.5: Genes and Genomes
genomics
1.1: Introduction
14: Genomics and Systems Biology
14.1: ‘Omics Technologies
14.2: DNA Sequencing
genotype
4.1: Introduction
4.5: Laws of Inheritance
9.8.2: Recombination
11.5: Calculating Probabilities
genotypes
9.8.5: Inferring Recombination From Genetic Data
11.2: Pedigree Analysis
Giemsa dye
10.4: Diagnosing Human Chromosome Abnormalities
Gleevec
13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
Griffith experiment
1.2: DNA is the Genetic Material
growth factors
13.5: Oncogenes
growth signals
13.3: Hallmarks of Cancer
GTP
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
guanine diphosphate
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
guanine triphosphate
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
gynandromorph
4.13: Sex-Linkage- An Exception to Mendel’s First Law

H

hairpin
8.3: Prokaryotic Transcription
hallmarks
13.3: Hallmarks of Cancer
haploid
3.1: Overview of Meiosis
3.3: Meiosis II
3.5: Comparing Meiosis and Mitosis
4.1: Introduction
haploinsufficient
4.10: Biochemical Basis of Dominance
haplosufficient
4.10: Biochemical Basis of Dominance
Hayflick limit
2.3: DNA is Packaged into Chromatin
health problems
10.3: Chromosome Abnormalities in Humans
healthcare
9.7: Example of Human Mutations
helicase
7.5: DNA Replication in Prokaryotes

hemizygous
4.6: Mendel's First Law
4.16: Extensions of the Laws of Inheritance
heredity
1: DNA, Genes, Chromosomes and Genome
1.4: Genes are the Basic Units of Inheritance
9.8: Linkage and Mapping
9.8.1: Linkage
9.8.2: Recombination
9.8.E: Linkage and Mapping (Exercises)
9.E: Mutation and Variation (Exercises)
heritable component
11.4: Sporadic and Non-Heritable Diseases
heterochromatin
2.3: DNA is Packaged into Chromatin
heterogametic
6.2: Autosomes and Sex Chromosomes
heterozygous
4.5: Laws of Inheritance
4.6: Mendel's First Law
9.8.5: Inferring Recombination From Genetic Data
highthroughput
14.1: 'Omics Technologies
histone
2.3: DNA is Packaged into Chromatin
histone acetylation
8.7.8: Cancer and Gene Regulation
holocentric
2.3: DNA is Packaged into Chromatin
holoenzyme
8.3: Prokaryotic Transcription
homogametic
6.2: Autosomes and Sex Chromosomes
homologous chromosomes
3.1: Overview of Meiosis
3.5: Comparing Meiosis and Mitosis
4.1: Introduction
homozygous
4.5: Laws of Inheritance
9.3: Genetic screening for mutations- Forward genetics
9.8.5: Inferring Recombination From Genetic Data
hormones
14.4: Functional Genomics – Determining Function(s)
HPV
13.4: Mutagens and Carcinogens
human traits
11.1: Prelude to Pedigrees and Populations
11.4: Sporadic and Non-Heritable Diseases
hybridization
4.3: Mendel's Experiments
hypermorph
9.4: Types of Mutations
hypomorph
9.4: Types of Mutations

I

imatinib
13.7: The "Poster Boy" of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
immune evasion
13.3: Hallmarks of Cancer
incomplete dominance
4.7: Relationships Between Genes, Genotypes and Phenotypes
4.16: Extensions of the Laws of Inheritance
independent assortment
3.2: Meiosis I
5: Genetic Analysis of Multiple Genes
9.8: Linkage and Mapping
9.8.1: Linkage
induced mutation
7.7: DNA Repair
inducible operon
8.7.3: Prokaryotic Gene Regulation
infertility
9.7: Example of Human Mutations
inflammation
13.3: Hallmarks of Cancer
inheritance
3: Meiosis - Sexual Reproduction
5: Genetic Analysis of Multiple Genes
6.1: Introduction
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
6.4: Sex Linkage- An Exception to Mendel's First Law
6.5: Y-Linked Genes
6.6: Chapter 10 Summary
9.8.4: Crossovers Allow Recombination of Linked Loci
9.8.5: Inferring Recombination From Genetic Data
11: Pedigrees and Populations
11.1: Prelude to Pedigrees and Populations
11.2: Pedigree Analysis
11.3: Inferring the Mode of Inheritance
11.5: Calculating Probabilities
inheritance patterns
5.2: Epistasis and Other Gene Interactions
initiation complex
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
initiation site
8.3: Prokaryotic Transcription
initiator tRNA
8.6: Ribosomes and Protein Synthesis
insertion
9.S: Mutation and Variation (Summary)
intellectual disability
10.3: Chromosome Abnormalities in Humans
interphase
2.4: The Cell Cycle and Changes in DNA Content
3.3: Meiosis II
introduction
6: Sex Chromosomes Sex Linkage
intron
8.5: RNA Processing in Eukaryotes
iontransport
9.7: Example of Human Mutations
Isolating Genomic DNA
12.2: Isolating Genomic DNA

K

karyogram
2.5: Karyotypes Describe Chromosome Number and Structure
3.6: Errors in Meiosis
karyotype
2.5: Karyotypes Describe Chromosome Number and Structure
3.6: Errors in Meiosis
10.3: Chromosome Abnormalities in Humans
key terms
6: Sex Chromosomes Sex Linkage
Kozak's rules
8.6: Ribosomes and Protein Synthesis

L

laboratory research
1.6: Model Organisms Facilitate Genetic Advances
Labrador Retriever
5.2: Epistasis and Other Gene Interactions
lac operon
8.7.3: Prokaryotic Gene Regulation
8.11.1: The lac Operon
lacI
8.11.1: The lac Operon
lagging strand
7.5: DNA Replication in Prokaryotes
large 60S ribosomal subunit
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
law of dominance
4.5: Laws of Inheritance
Law of Equal Segregation
4.6: Mendel's First Law
Law of Independent Assortment
4.5: Laws of Inheritance
4.9: Pedigrees and Punnett Squares
5.1: Dihybrid Crosses
law of segregation
4.5: Laws of Inheritance
4.9: Pedigrees and Punnett Squares
leading strand
7.5: DNA Replication in Prokaryotes
learning
9.3: Genetic screening for mutations- Forward genetics
lethal alleles
9.5: Some mutations may not have detectable phenotypes
ligase
7.5: DNA Replication in Prokaryotes
light microscope
2.2: Studying Cells
linkage
4.16: Extensions of the Laws of Inheritance
9.8: Linkage and Mapping
9.8.3: Linkage Reduces Recombination Frequency
9.8.4: Crossovers Allow Recombination of Linked Loci
9.8.6: Genetic Mapping
9.8.S: Linkage and Mapping (Summary)
linked
9.8.S: Linkage and Mapping (Summary)
linked genes
5.1: Dihybrid Crosses

loci
9.8.3: Linkage Reduces Recombination Frequency
9.8.4: Crossovers Allow Recombination of Linked Loci
9.8.7: Mapping With Three-Point Crosses
9.8.S: Linkage and Mapping (Summary)
locus
4.7: Relationships Between Genes, Genotypes and Phenotypes
loss of function
9.6: Complementation tests and Allelism
lungdisease
9.7: Example of Human Mutations

M

male development
6.5: Y-Linked Genes
male fertility
6.5: Y-Linked Genes
malignant tumors
13.2: Cancer Cell Biology
mammals
6.2: Autosomes and Sex Chromosomes
map distance
9.8.7: Mapping With Three-Point Crosses
meiocytes
3.4: Meiosis
meioses
9.8.S: Linkage and Mapping (Summary)
meiosis
2.4: The Cell Cycle and Changes in DNA Content
2.5: Karyotypes Describe Chromosome Number and Structure
3: Meiosis - Sexual Reproduction
3.3: Meiosis II
3.5: Comparing Meiosis and Mitosis
4.1: Introduction
6.2: Autosomes and Sex Chromosomes
6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes
6.4: Sex Linkage- An Exception to Mendel's First Law
9.8.2: Recombination
9.8.3: Linkage Reduces Recombination Frequency
9.8.4: Crossovers Allow Recombination of Linked Loci
9.8.5: Inferring Recombination From Genetic Data
10: Changes in Chromosome Number and Structure
10.1: Changes in Chromosome Number
10.2: Changes in Chromosome Structure
10.3: Chromosome Abnormalities in Humans
meiosis I
3.1: Overview of Meiosis
3.2: Meiosis I
meiosis II
3.1: Overview of Meiosis
melanoma
13.1: Classification of Cancers
memory
9.3: Genetic screening for mutations- Forward genetics
Mendel
1.4: Genes are the Basic Units of Inheritance
5: Genetic Analysis of Multiple Genes
9.8: Linkage and Mapping
9.8.1: Linkage
9.8.2: Recombination
Mendel’s First Law
4.6: Mendel’s First Law
Mendel’s Second Law
5.1: Dihybrid Crosses
mendel's first law
6: Sex Chromosomes Sex Linkage
Mendelian inheritance
6.6: Chapter 10 Summary
11.5: Calculating Probabilities
metabolic pathways
13.3: Hallmarks of Cancer
metabolomics
14.1: ‘Omics Technologies
metastasis
13.3: Hallmarks of Cancer
metastatic
13.2: Cancer Cell Biology
microarray analysis
14.4: Functional Genomics – Determining Function(s)
microbiology
1.2: DNA is the Genetic Material
microRNA
8.7.6: Eukaryotic Post-transcriptional Gene Regulation
microsatellites
9.2: Origins of Mutations
microscope
2.2: Studying Cells
miRNA
8.7.6: Eukaryotic Post-transcriptional Gene Regulation
mismatch repair
7.7: DNA Repair
mitosis
2.4: The Cell Cycle and Changes in DNA Content
3.3: Meiosis II
3.5: Comparing Meiosis and Mitosis
10: Changes in Chromosome Number and Structure
10.1: Changes in Chromosome Number
mobile genetic elements
9.2: Origins of Mutations
model organisms
1.6: Model Organisms Facilitate Genetic Advances
model system
4.3: Mendel’s Experiments
molecular biology
1: DNA, Genes, Chromosomes and Genome
1.5: Genes and Genomes
14: Genomics and Systems Biology
molecular genetics
12: Techniques of Molecular Genetics
molecular markers
9.1: Mutation and Polymorphism
molecularbiology
14.1: ‘Omics Technologies
monohybrid
4.5: Laws of Inheritance
monohybrid crosses
4.4: Punnett Squares and Test Crosses
5.1: Dihybrid Crosses
monosomy
2.5: Karyotypes Describe Chromosome Number and Structure
3.6: Errors in Meiosis
mosaic
4.13: Sex-Linkage- An Exception to Mendel’s First Law
Mus musculus
1.6: Model Organisms Facilitate Genetic Advances
mutagen
9.S: Mutation and Variation (Summary)
13.4: Mutagens and Carcinogens
mutant
9.1: Mutation and Polymorphism
9.S: Mutation and Variation (Summary)
mutant alleles
11.4: Sporadic and Non-Heritable Diseases
mutant phenotype
9.6: Complementation tests and Allelism
mutant screening
9.6: Complementation tests and Allelism
mutant screens
9.5: Some mutations may not have detectable phenotypes
mutants
6.4: Sex Linkage- An Exception to Mendel's First Law
mutation
1.5: Genes and Genomes
3: Meiosis - Sexual Reproduction
7.7: DNA Repair
9: Mutation and Variation
9.7: Example of Human Mutations
9.8.6: Genetic Mapping
9.S: Mutation and Variation (Summary)
13.5: Oncogenes
mutational analysis
9.5: Some mutations may not have detectable phenotypes
mutations
2.5: Karyotypes Describe Chromosome Number and Structure
9.1: Mutation and Polymorphism
9.2: Origins of Mutations
9.3: Genetic screening for mutations- Forward genetics
9.4: Types of Mutations
9.E: Mutation and Variation (Exercises)
13: Cancer Genetics
13.6: Tumor Suppressor Genes
myc
8.7.8: Cancer and Gene Regulation
myeloid cells
13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)

N

natural_selection
11: Pedigrees and Populations
negative regulator
8.7.3: Prokaryotic Gene Regulation
neomorph
9.4: Types of Mutations
neuronal development
14.4: Functional Genomics – Determining Function(s)

nondisjunction
3.6: Errors in Meiosis
10.1: Changes in Chromosome Number
10.3: Chromosome Abnormalities in Humans
nonsense codon
8.2: The Genetic Code
nontemplate strand
8.3: Prokaryotic Transcription
northern blot
12.7: DNA Analysis- Blotting and Hybridization
Nucleases
12.2: Isolating Genomic DNA
nucleosome
2.3: DNA is Packaged into Chromatin
nucleotide excision repair
7.7: DNA Repair
nucleotides
1.3: DNA Structure

O

Octamer box
8.4: Eukaryotic Transcription
OGG
1.E: Overview, DNA, and Genes (Exercises)
1.S: Overview, DNA, and Genes (Summary)
2.3: DNA is Packaged into Chromatin
3.4: Meiosis
4: Mendelian Genetics
4.4: Punnett Squares and Test Crosses
4.6: Mendel's First Law
4.7: Relationships Between Genes, Genotypes and Phenotypes
4.10: Biochemical Basis of Dominance
4.13: Sex-Linkage- An Exception to Mendel's First Law
4.14: Phenotypes May Not Be As Expected from the Genotype
4.15: Phenotypic Ratios May Not Be As Expected
4.E: Genetic Analysis of Single Genes (Exercises)
4.S: Genetic Analysis of Single Genes (Summary)
Okazaki fragment
7.5: DNA Replication in Prokaryotes
omics
14.1: 'Omics Technologies
oncogene
13.4: Mutagens and Carcinogens
13.5: Oncogenes
OOG
1.E: Overview, DNA, and Genes (Exercises)
1.S: Overview, DNA, and Genes (Summary)
2.3: DNA is Packaged into Chromatin
3.4: Meiosis
4.4: Punnett Squares and Test Crosses
4.6: Mendel's First Law
4.7: Relationships Between Genes, Genotypes and Phenotypes
4.10: Biochemical Basis of Dominance
4.13: Sex-Linkage- An Exception to Mendel's First Law
4.14: Phenotypes May Not Be As Expected from the Genotype
4.15: Phenotypic Ratios May Not Be As Expected
4.E: Genetic Analysis of Single Genes (Exercises)
4.S: Genetic Analysis of Single Genes (Summary)
operator
8.7.3: Prokaryotic Gene Regulation
operons
8.7.3: Prokaryotic Gene Regulation
organisms
11: Pedigrees and Populations
origin of replication
2.3: DNA is Packaged into Chromatin

P

P
4.3: Mendel's Experiments
p53
13.6: Tumor Suppressor Genes
parental genotypes
9.8.7: Mapping With Three-Point Crosses
particulate inheritance
1.4: Genes are the Basic Units of Inheritance
pathogenic bacteria
1.2: DNA is the Genetic Material
pathogenicity
1.2: DNA is the Genetic Material
pedigree
4.9: Pedigrees and Punnett Squares
11.2: Pedigree Analysis
11.3: Inferring the Mode of Inheritance
11.5: Calculating Probabilities
pedigrees
11: Pedigrees and Populations
11.1: Prelude to Pedigrees and Populations
penetrance
4.14: Phenotypes May Not Be As Expected from the Genotype
peptidyl transferase
8.6: Ribosomes and Protein Synthesis
phenotype
4.1: Introduction
4.5: Laws of Inheritance
9: Mutation and Variation
9.1: Mutation and Polymorphism
9.4: Types of Mutations
9.5: Some mutations may not have detectable phenotypes
9.8.5: Inferring Recombination From Genetic Data
10: Changes in Chromosome Number and Structure
11.5: Calculating Probabilities
phenotypes
6.1: Introduction
6.4: Sex Linkage- An Exception to Mendel's First Law
6.6: Chapter 10 Summary
9.3: Genetic screening for mutations- Forward genetics
9.8.E: Linkage and Mapping (Exercises)
9.E: Mutation and Variation (Exercises)
11.2: Pedigree Analysis
11.3: Inferring the Mode of Inheritance
phenotypic ratio
5: Genetic Analysis of Multiple Genes
5.2: Epistasis and Other Gene Interactions
phenotypic traits
1.5: Genes and Genomes
Philadelphia chromosome
13.7: The "Poster Boy" of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
physical mapping
14.3: Whole Genome Sequencing
plasmid
8.3: Prokaryotic Transcription
plasmids
12.5: Cloning DNA - Plasmid Vectors
ploidy
3.1: Overview of Meiosis
point mutation
7.7: DNA Repair
polydactyly
11.1: Prelude to Pedigrees and Populations
polymerase chain reaction (PCR)
12.3: Isolating or Detecting a Specific Sequence by PCR
polymorphism
9.1: Mutation and Polymorphism
9.8.6: Genetic Mapping
9.S: Mutation and Variation (Summary)
polymorphisms
9.E: Mutation and Variation (Exercises)
polyploid
3.6: Errors in Meiosis
polysome
8.6: Ribosomes and Protein Synthesis
population
11.1: Prelude to Pedigrees and Populations
population genetics
11.6: Population Genetics
population_genetics
11: Pedigrees and Populations
populations
10.2: Changes in Chromosome Structure
positive regulator
8.7.3: Prokaryotic Gene Regulation
preinitiation complex
8.4: Eukaryotic Transcription
primase
7.5: DNA Replication in Prokaryotes
primer
7.5: DNA Replication in Prokaryotes
12.3: Isolating or Detecting a Specific Sequence by PCR
primers
14.2: DNA Sequencing
probabilities
11.5: Calculating Probabilities
prokaryotes
7.5: DNA Replication in Prokaryotes
promoter
8.3: Prokaryotic Transcription
proofreading
7.7: DNA Repair
prophase I
9.8.4: Crossovers Allow Recombination of Linked Loci
proteasome
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
protein
14.2: DNA Sequencing
protein sequencing
14: Genomics and Systems Biology
proteins
1: DNA, Genes, Chromosomes and Genome

proteomics
14.1: ‘Omics Technologies
pseudoautosomal regions
6.5: Y-Linked Genes
Punnett square
4.4: Punnett Squares and Test Crosses
4.5: Laws of Inheritance
4.9: Pedigrees and Punnett Squares
11.5: Calculating Probabilities
purple flowers
9.6: Complementation tests and Allelism
pyrimidine dimers
7.7: DNA Repair

R

radiation
13.4: Mutagens and Carcinogens
reading frame
8.2: The Genetic Code
rearrangements
10.2: Changes in Chromosome Structure
receptors
13.5: Oncogenes
recessive
4.3: Mendel’s Experiments
4.7: Relationships Between Genes, Genotypes and Phenotypes
4.11: Black fur color - a dominant trait
9.3: Genetic screening for mutations- Forward genetics
recessive mutations
9.6: Complementation tests and Allelism
reciprocal cross
4.3: Mendel’s Experiments
reciprocal crosses
4.13: Sex-Linkage- An Exception to Mendel’s First Law
6.4: Sex Linkage- An Exception to Mendel's First Law
recombinant
3.2: Meiosis I
9.8.2: Recombination
recombination
4.16: Extensions of the Laws of Inheritance
9.8: Linkage and Mapping
9.8.3: Linkage Reduces Recombination Frequency
9.8.4: Crossovers Allow Recombination of Linked Loci
9.8.5: Inferring Recombination From Genetic Data
9.8.6: Genetic Mapping
9.8.E: Linkage and Mapping (Exercises)
9.8.S: Linkage and Mapping (Summary)
recombination frequency
9.8.7: Mapping With Three-Point Crosses
replication
9.2: Origins of Mutations
10: Changes in Chromosome Number and Structure
replication fork
7.5: DNA Replication in Prokaryotes
repressor
8.7.3: Prokaryotic Gene Regulation
reproduction
3: Meiosis - Sexual Reproduction
restriction endonucleases,
12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
restriction enzymes
12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
retrotransposon
9.S: Mutation and Variation (Summary)
RISC
8.7.6: Eukaryotic Post-transcriptional Gene Regulation
RNA
1.3: DNA Structure
14: Genomics and Systems Biology
14.2: DNA Sequencing
RNA editing
8.5: RNA Processing in Eukaryotes
RNA stability
8.7.6: Eukaryotic Post-transcriptional Gene Regulation

S

S phase
2.4: The Cell Cycle and Changes in DNA Content
Saccharomyces cerevisiae
1.6: Model Organisms Facilitate Genetic Advances
sampling effects
4.15: Phenotypic Ratios May Not Be As Expected
sarcoma
13.1: Classification of Cancers
sarcomas
13: Cancer Genetics
scaffold proteins
2.3: DNA is Packaged into Chromatin
segregation
9.8.2: Recombination
9.8.3: Linkage Reduces Recombination Frequency
10.1: Changes in Chromosome Number
semiconservative replication
1.3: DNA Structure
sequencing
14.2: DNA Sequencing
14.3: Whole Genome Sequencing
sex chromosomes
6: Sex Chromosomes Sex Linkage
6.1: Introduction
6.2: Autosomes and Sex Chromosomes
6.4: Sex Linkage- An Exception to Mendel's First Law
6.6: Chapter 10 Summary
sex determination
6.1: Introduction
6.5: Y-Linked Genes
sex linkage
6: Sex Chromosomes Sex Linkage
6.1: Introduction
signal sequence
8.6: Ribosomes and Protein Synthesis
signal transduction
13.5: Oncogenes
silent mutation
7.7: DNA Repair
silent mutations
9.5: Some mutations may not have detectable phenotypes
skin cancer
13.1: Classification of Cancers
sliding clamp
7.5: DNA Replication in Prokaryotes
small 40S ribosomal subunit
8.7.7: Eukaryotic Translational and Post-translational Gene Regulation
small nuclear RNA
8.4: Eukaryotic Transcription
somatic cell
3.1: Overview of Meiosis
Southern blot
12.7: DNA Analysis- Blotting and Hybridization
species
10.2: Changes in Chromosome Structure
spindle fibers
3.5: Comparing Meiosis and Mitosis
splicing
8.5: RNA Processing in Eukaryotes
spontaneous mutation
7.7: DNA Repair
squamous cell
13.1: Classification of Cancers
squash
5.2: Epistasis and Other Gene Interactions
SRY gene
6.5: Y-Linked Genes
STARLINK AFFAIR
12.3: Isolating or Detecting a Specific Sequence by PCR
start codon
8.6: Ribosomes and Protein Synthesis
Streptococcus pneumoniae
1.2: DNA is the Genetic Material
study
11.1: Prelude to Pedigrees and Populations
study questions
6: Sex Chromosomes Sex Linkage
substitution
9.S: Mutation and Variation (Summary)
synapsis
3.2: Meiosis I
synteny
9.8.4: Crossovers Allow Recombination of Linked Loci

T

targeted therapy
13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)
TATA box
8.3: Prokaryotic Transcription
technology
14.2: DNA Sequencing
telomerase
7.6: DNA Replication in Eukaryotes
telomere
7.6: DNA Replication in Eukaryotes

telomeres
13.3: Hallmarks of Cancer
template strand
8.3: Prokaryotic Transcription
temporal expression
14.4: Functional Genomics – Determining Function(s)
test cross
4.5: Laws of Inheritance
testcross
9.8.5: Inferring Recombination From Genetic Data
tetrad
3.2: Meiosis I
tissue classification
13.1: Classification of Cancers
tissue specific expression
14.4: Functional Genomics – Determining Function(s)
topoisomerase
7.5: DNA Replication in Prokaryotes
trait
4.1: Introduction
4.3: Mendel's Experiments
traits
1.4: Genes are the Basic Units of Inheritance
5: Genetic Analysis of Multiple Genes
9.8: Linkage and Mapping
9.8.E: Linkage and Mapping (Exercises)
11.1: Prelude to Pedigrees and Populations
11.2: Pedigree Analysis
11.3: Inferring the Mode of Inheritance
transcript abundance
14.4: Functional Genomics – Determining Function(s)
transcription factor
8.7.4: Eukaryotic Epigenetic Gene Regulation
transcription factor binding site
8.7.5: Eukaryotic Transcription Gene Regulation
transcription factors
13.5: Oncogenes
transcription_factor
13.6: Tumor Suppressor Genes
transcriptomics
14.1: 'Omics Technologies
transfection
12.8: Transgenic organisms
transformation
7.2: Historical Basis of Modern Understanding
12.5: Cloning DNA - Plasmid Vectors
12.8: Transgenic organisms
transgene
12.8: Transgenic organisms
transgenesis
12.8: Transgenic organisms
Transgenic organism
12.8: Transgenic organisms
transgenic plant
12.8: Transgenic organisms
transition substitution
7.7: DNA Repair
transposable elements
9.2: Origins of Mutations
transposon
9.S: Mutation and Variation (Summary)
transversion substitution
7.7: DNA Repair
treatment
13.2: Cancer Cell Biology
trihybrid
9.8.7: Mapping With Three-Point Crosses
trisomy
2.5: Karyotypes Describe Chromosome Number and Structure
3.6: Errors in Meiosis
10.3: Chromosome Abnormalities in Humans
trisomy 21
10.4: Diagnosing Human Chromosome Abnormalities
Trp operon
8.7.3: Prokaryotic Gene Regulation
tryptophan
8.7.3: Prokaryotic Gene Regulation
tumor
13.2: Cancer Cell Biology
tumor_suppressors
13.6: Tumor Suppressor Genes
tyrosine kinase
13.7: The "Poster Boy" of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)

U
unified cell theory
2.2: Studying Cells
untranslated region
8.7.6: Eukaryotic Post-transcriptional Gene Regulation
upstream
8.3: Prokaryotic Transcription

V
vaccine development
13: Cancer Genetics
variation
9.8.E: Linkage and Mapping (Exercises)
vernalization
8.11.6: Epigenetics

W
Watson and Crick
7.3: DNA Structure and Sequencing
Western blot
12.7: DNA Analysis- Blotting and Hybridization
whole genome shotgun
14.3: Whole Genome Sequencing
wild type
4.16: Extensions of the Laws of Inheritance

X
X inactivation
3.6: Errors in Meiosis
XYY syndrome
6.5: Y-Linked Genes

Y
Y chromosome
6.5: Y-Linked Genes

Z
zygote
4.1: Introduction

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Introduction to Genetics

Webpages: 225

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-SA 4.0: 44% (99 pages)
- CC BY 4.0: 32% (72 pages)
- CC BY-NC-SA 4.0: 13.8% (31 pages)
- Undeclared: 10.2% (23 pages)

By Page

- Introduction to Genetics - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: DNA, Genes, Chromosomes and Genome - *CC BY 4.0*
    - 1.1: Introduction - *CC BY 4.0*
    - 1.2: DNA is the Genetic Material - *CC BY 4.0*
    - 1.3: DNA Structure - *CC BY 4.0*
    - 1.4: Genes are the Basic Units of Inheritance - *CC BY 4.0*
    - 1.5: Genes and Genomes - *CC BY 4.0*
    - 1.6: Model Organisms Facilitate Genetic Advances - *CC BY 4.0*
    - 1.E: Overview, DNA, and Genes (Exercises) - *CC BY 4.0*
    - 1.S: Overview, DNA, and Genes (Summary) - *CC BY 4.0*
  - 2: Cells and Mitosis - *CC BY 4.0*
    - 2.1: Introduction - *CC BY 4.0*
    - 2.2: Studying Cells - *CC BY 4.0*
    - 2.3: DNA is Packaged into Chromatin - *CC BY-SA 4.0*
    - 2.4: The Cell Cycle and Changes in DNA Content - *CC BY-SA 4.0*
    - 2.5: Karyotypes Describe Chromosome Number and Structure - *CC BY-SA 4.0*
    - 2.6: Key Words and Terms - *CC BY 4.0*
  - 3: Meiosis - Sexual Reproduction - *CC BY 4.0*
    - 3.1: Overview of Meiosis - *CC BY 4.0*
    - 3.2: Meiosis I - *CC BY 4.0*
    - 3.3: Meiosis II - *CC BY 4.0*
    - 3.4: Meiosis - *CC BY-SA 4.0*
    - 3.5: Comparing Meiosis and Mitosis - *CC BY 4.0*
    - 3.6: Errors in Meiosis - *CC BY 4.0*
  - 4: Mendelian Genetics - *CC BY 4.0*
    - 4.1: Introduction - *CC BY 4.0*
    - 4.2: Introduction - *CC BY-NC-SA 4.0*
    - 4.3: Mendel's Experiments - *CC BY 4.0*
    - 4.4: Punnett Squares and Test Crosses - *CC BY 4.0*
    - 4.5: Laws of Inheritance - *CC BY 4.0*
    - 4.6: Mendel's First Law - *CC BY 4.0*
    - 4.7: Relationships Between Genes, Genotypes and Phenotypes - *CC BY 4.0*
    - 4.8: Pedigree Analysis - *CC BY 4.0*
      - 4.8.1: Symbols used in Pedigree Charts - *CC BY 4.0*
      - 4.8.2: Modes of Inheritance - *CC BY 4.0*
      - 4.8.3: Sporadic and Non-Heritable Diseases - *CC BY 4.0*
      - 4.8.4: Calculating Probabilities Using Pedigree Charts - *CC BY 4.0*
      - 4.8.5: Chapter 4 Summary - *CC BY 4.0*
      - 4.8.6: Key Terms in Chapter 4 - *CC BY 4.0*
      - 4.8.7: Chapter 4 Study Questions - *CC BY 4.0*
      - 4.8.8: Chapter 4 Answers - *CC BY 4.0*
    - 4.9: Pedigrees and Punnett Squares - *CC BY 4.0*
    - 4.10: Biochemical Basis of Dominance - *CC BY 4.0*
    - 4.11: Black fur color - a dominant trait - *CC BY 4.0*
    - 4.12: Yellow fur color - a recessive trait - *CC BY 4.0*
    - 4.13: Sex-Linkage- An Exception to Mendel's First Law - *CC BY 4.0*
    - 4.14: Phenotypes May Not Be As Expected from the Genotype - *CC BY 4.0*
    - 4.15: Phenotypic Ratios May Not Be As Expected - *CC BY 4.0*

- 4.16: Extensions of the Laws of Inheritance - *CC BY 4.0*
- 4.E: Genetic Analysis of Single Genes (Exercises) - *CC BY 4.0*
- 4.S: Genetic Analysis of Single Genes (Summary) - *CC BY 4.0*

5: Genetic Analysis of Multiple Genes - *CC BY-SA 4.0*

- 5.1: Dihybrid Crosses - *CC BY-SA 4.0*
- 5.2: Epistasis and Other Gene Interactions - *CC BY-SA 4.0*
- 5.3: Example of Multiple Genes Affecting One Character - *CC BY-SA 4.0*
- 5.4: Alleles at a Single Locus - *CC BY-NC-SA 4.0*
  - 5.4.1: Introduction - *CC BY-NC-SA 4.0*
  - 5.4.2: Somatic vs. Germline Mutations - *CC BY-NC-SA 4.0*
  - 5.4.3: Alleles- Hetero-, Homo-, Hemizygosity - *CC BY-NC-SA 4.0*
  - 5.4.4: Pleiotropy vs. Polygenic Inheritance - *CC BY-NC-SA 4.0*
  - 5.4.5: Types of Dominance - *CC BY-NC-SA 4.0*
  - 5.4.6: Biochemical Basis of Dominance - *CC BY-NC-SA 4.0*
  - 5.4.7: Classification of Mutants - *CC BY-NC-SA 4.0*
  - 5.4.8: Muller's Morphs - *CC BY-NC-SA 4.0*
  - 5.4.9: Chapter 6 Summary - *CC BY-NC-SA 4.0*
  - 5.4.10: Key Terms in Chapter 6 - *CC BY-NC-SA 4.0*
  - 5.4.11: Chapter 6 Study Questions - *CC BY-NC-SA 4.0*
  - 5.4.12: Chapter 6 Answers - *CC BY-NC-SA 4.0*
- 5.E: Genetic Analysis of Multiple Genes (Exercises) - *CC BY-SA 4.0*
- 5.S: Genetic Analysis of Multiple Genes (Summary) - *CC BY-SA 4.0*

6: Sex Chromosomes Sex Linkage - *CC BY-NC-SA 4.0*

- 6.1: Introduction - *CC BY-NC-SA 4.0*
- 6.2: Autosomes and Sex Chromosomes - *CC BY-NC-SA 4.0*
- 6.3: Pseudo-Autosomal Regions on the X and Y Chromosomes - *CC BY-NC-SA 4.0*
- 6.4: Sex Linkage- An Exception to Mendel's First Law - *CC BY-NC-SA 4.0*
- 6.5: Y-Linked Genes - *CC BY-NC-SA 4.0*
- 6.6: Chapter 10 Summary - *CC BY-NC-SA 4.0*
- 6.7: Key Terms in Chapter 10 - *CC BY-NC-SA 4.0*
- 6.8: Chapter 10 Study Questions - *CC BY-NC-SA 4.0*
- 6.9: Chapter 10 Answers - *CC BY-NC-SA 4.0*

7: DNA Structure and Function - *CC BY 4.0*

- 7.1: Introduction - *CC BY 4.0*
- 7.2: Historical Basis of Modern Understanding - *CC BY 4.0*
- 7.3: DNA Structure and Sequencing - *CC BY 4.0*
- 7.4: Basics of DNA Replication - *CC BY 4.0*
- 7.5: DNA Replication in Prokaryotes - *CC BY 4.0*
- 7.6: DNA Replication in Eukaryotes - *CC BY 4.0*
- 7.7: DNA Repair - *CC BY 4.0*
- 7.8: Key Terms - *CC BY 4.0*
- 7.9: Chapter Summary - *Undeclared*
- 7.10: Visual Connection Questions - *Undeclared*
- 7.11: Review Questions - *Undeclared*
- 7.12: Critical Thinking Questions - *Undeclared*

8: Genes and Proteins - *CC BY 4.0*

- 8.1: Introduction - *CC BY 4.0*
- 8.2: The Genetic Code - *CC BY 4.0*
- 8.3: Prokaryotic Transcription - *CC BY 4.0*
- 8.4: Eukaryotic Transcription - *CC BY 4.0*
- 8.5: RNA Processing in Eukaryotes - *CC BY 4.0*
- 8.6: Ribosomes and Protein Synthesis - *CC BY 4.0*
- 8.7: Gene Expression - *CC BY 4.0*
  - 8.7.1: Introduction - *CC BY 4.0*
  - 8.7.2: Regulation of Gene Expression - *CC BY 4.0*
  - 8.7.3: Prokaryotic Gene Regulation - *CC BY 4.0*
  - 8.7.4: Eukaryotic Epigenetic Gene Regulation - *CC BY 4.0*
  - 8.7.5: Eukaryotic Transcription Gene Regulation - *CC BY 4.0*
  - 8.7.6: Eukaryotic Post-transcriptional Gene Regulation - *CC BY 4.0*
  - 8.7.7: Eukaryotic Translational and Post-translational Gene Regulation - *CC BY 4.0*
  - 8.7.8: Cancer and Gene Regulation - *CC BY 4.0*
  - 8.7.9: Key Terms - *CC BY 4.0*
  - 8.7.10: Chapter Summary - *Undeclared*
  - 8.7.11: Visual Connection Questions - *Undeclared*
  - 8.7.12: Review Questions - *Undeclared*
  - 8.7.13: Critical Thinking Questions - *Undeclared*
- 8.8: Key Terms - *CC BY 4.0*
- 8.9: Chapter Summary - *Undeclared*
- 8.10: Visual Connection Questions - *Undeclared*
- 8.11: Regulation of Gene Expression - *CC BY-SA 4.0*
  - 8.11.1: The lac Operon - *CC BY-SA 4.0*
  - 8.11.2: The Use of Mutants to Study the lac Operon - *CC BY-SA 4.0*
  - 8.11.3: Eukaryotic Gene Regulation - *CC BY-SA 4.0*

        - 8.11.4: Regulatory Elements in Evolution - *CC BY-SA 4.0*
        - 8.11.5: Additional Levels of Regulating Transcription - *CC BY-SA 4.0*
        - 8.11.6: Epigenetics - *CC BY-SA 4.0*
        - 8.11.7: Regulation of Gene Expression (Exercises) - *CC BY-SA 4.0*
        - 8.11.8: Gene Expression Essentials - *Undeclared*
        - 8.11.S: Regulation of Gene Expression (Summary) - *CC BY-SA 4.0*
    - 8.12: Review Questions - *Undeclared*
    - 8.13: Critical Thinking Questions - *Undeclared*
- 9: Mutation and Variation - *CC BY-SA 4.0*
    - 9.1: Mutation and Polymorphism - *CC BY-SA 4.0*
    - 9.2: Origins of Mutations - *CC BY-SA 4.0*
    - 9.3: Genetic screening for mutations- Forward genetics - *CC BY-SA 4.0*
    - 9.4: Types of Mutations - *CC BY-SA 4.0*
    - 9.5: Some mutations may not have detectable phenotypes - *CC BY-SA 4.0*
    - 9.6: Complementation tests and Allelism - *CC BY-SA 4.0*
    - 9.7: Example of Human Mutations - *CC BY-SA 4.0*
    - 9.8: Linkage and Mapping - *CC BY-SA 4.0*
        - 9.8.1: Linkage - *CC BY-SA 4.0*
        - 9.8.2: Recombination - *CC BY-SA 4.0*
        - 9.8.3: Linkage Reduces Recombination Frequency - *CC BY-SA 4.0*
        - 9.8.4: Crossovers Allow Recombination of Linked Loci - *CC BY-SA 4.0*
        - 9.8.5: Inferring Recombination From Genetic Data - *CC BY-SA 4.0*
        - 9.8.6: Genetic Mapping - *CC BY-SA 4.0*
        - 9.8.7: Mapping With Three-Point Crosses - *CC BY-SA 4.0*
        - 9.8.E: Linkage and Mapping (Exercises) - *CC BY-SA 4.0*
        - 9.8.S: Linkage and Mapping (Summary) - *CC BY-SA 4.0*
    - 9.E: Mutation and Variation (Exercises) - *CC BY-SA 4.0*
    - 9.S: Mutation and Variation (Summary) - *CC BY-SA 4.0*
- 10: Changes in Chromosome Number and Structure - *CC BY-SA 4.0*
    - 10.1: Changes in Chromosome Number - *CC BY-SA 4.0*
    - 10.2: Changes in Chromosome Structure - *CC BY-SA 4.0*
    - 10.3: Chromosome Abnormalities in Humans - *CC BY-SA 4.0*
    - 10.4: Diagnosing Human Chromosome Abnormalities - *CC BY-SA 4.0*
    - 10.E: Changes in Chromosome Number and Structure (Exercises) - *CC BY-SA 4.0*
    - 10.S: Changes in Chromosome Number and Structure (Summary) - *CC BY-SA 4.0*
- 11: Pedigrees and Populations - *CC BY-SA 4.0*
    - 11.1: Prelude to Pedigrees and Populations - *CC BY-SA 4.0*
    - 11.2: Pedigree Analysis - *CC BY-SA 4.0*
    - 11.3: Inferring the Mode of Inheritance - *CC BY-SA 4.0*
    - 11.4: Sporadic and Non-Heritable Diseases - *CC BY-SA 4.0*
    - 11.5: Calculating Probabilities - *CC BY-SA 4.0*
    - 11.6: Population Genetics - *CC BY-SA 4.0*
    - 11.E: Pedigrees and Populations (Exercises) - *CC BY-SA 4.0*
    - 11.S: Pedigrees and Populations (Summary) - *CC BY-SA 4.0*
- 12: Techniques of Molecular Genetics - *CC BY-SA 4.0*
    - 12.1: Prelude to Molecular Genetics - *CC BY-SA 4.0*
    - 12.2: Isolating Genomic DNA - *CC BY-SA 4.0*
    - 12.3: Isolating or Detecting a Specific Sequence by PCR - *CC BY-SA 4.0*
    - 12.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation - *CC BY-SA 4.0*
    - 12.5: Cloning DNA - Plasmid Vectors - *CC BY-SA 4.0*
    - 12.6: DNA Analysis - Gel Electrophoresis - *CC BY-SA 4.0*
    - 12.7: DNA Analysis- Blotting and Hybridization - *CC BY-SA 4.0*
    - 12.8: Transgenic organisms - *CC BY-SA 4.0*
    - 12.9: Techniques of Molecular Genetics (Exercises) - *CC BY-SA 4.0*
    - 12.S: Techniques of Molecular Genetics (Summary) - *CC BY-SA 4.0*
- 13: Cancer Genetics - *CC BY-SA 4.0*
    - 13.1: Classification of Cancers - *CC BY-SA 4.0*
    - 13.2: Cancer Cell Biology - *CC BY-SA 4.0*
    - 13.3: Hallmarks of Cancer - *CC BY-SA 4.0*
    - 13.4: Mutagens and Carcinogens - *CC BY-SA 4.0*
    - 13.5: Oncogenes - *CC BY-SA 4.0*
    - 13.6: Tumor Suppressor Genes - *CC BY-SA 4.0*
    - 13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib) - *CC BY-SA 4.0*

- 13.8: Genes and COVID-19 Susceptibility in Humans - *CC BY-NC-SA 4.0*
  - 13.8.1: Introduction - *CC BY-NC-SA 4.0*
  - 13.8.2: How Does SARS-Cov-2 Infect Humans? - *CC BY-NC-SA 4.0*
  - 13.8.3: Genes Implicated in Severe COVID-19 Infection in Humans - *CC BY-NC-SA 4.0*
  - 13.8.4: Approaches for Vaccine Development Against SARS-Cov-2 - *CC BY-NC-SA 4.0*
  - 13.8.5: Chapter 13 Summary - *CC BY-NC-SA 4.0*
  - 13.8.6: Key Terms in Chapter 13 - *CC BY-NC-SA 4.0*
- 13.E: Cancer Genetics (Exercises) - *CC BY-SA 4.0*
- 13.S: Cancer Genetics (Summary) - *CC BY-SA 4.0*

14: Genomics and Systems Biology - *CC BY-SA 4.0*
- 14.1: 'Omics Technologies - *CC BY-SA 4.0*
- 14.2: DNA Sequencing - *CC BY-SA 4.0*
- 14.3: Whole Genome Sequencing - *CC BY-SA 4.0*
- 14.4: Functional Genomics – Determining Function(s) - *CC BY-SA 4.0*
- 14.E: Genomics and Systems Biology (Exercises) - *CC BY-SA 4.0*
- 14.S: Genomics and Systems Biology (Summary) - *CC BY-SA 4.0*

15: Molecular Markers and Quantitative Traits - *CC BY-SA 4.0*
- 15.1: Some Variations in the Genome Affect Complex Traits - *CC BY-SA 4.0*
- 15.2: Origins of Molecular Polymorphisms - *CC BY-SA 4.0*
- 15.3: Classification and Detection of Molecular Markers - *CC BY-SA 4.0*
- 15.4: Applications of Molecular Markers - *CC BY-SA 4.0*
- 15.5: Quantitative Trait Locus (QTL) Analysis - *CC BY-SA 4.0*
- 15.E: Molecular Markers and Quantitative Traits (Exercises) - *CC BY-SA 4.0*
- 15.S: Molecular Markers and Quantitative Traits (Summary) - *CC BY-SA 4.0*

16: Appendix 1 – Simple Nomenclature - *CC BY-SA 4.0*
- 16.1: Textbook Introduction - *CC BY-SA 4.0*
- 16.2: Why Use Nomenclature? - *CC BY-SA 4.0*
- 16.3: Biochemistry of Gene Behavior - *CC BY-SA 4.0*
- 16.4: Basic Nomenclature - *CC BY-SA 4.0*
- 16.5: Three-Letter Symbols - *CC BY-SA 4.0*
- 16.6: Linked Genes - *CC BY-SA 4.0*
- 16.7: Summary - *CC BY-SA 4.0*

Back Matter - *Undeclared*
- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*