

BIO 300: INTRODUCTION TO GENETICS



Cheneir Neely
cneely

Introduction to Genetics (Neely)

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 borne 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 were 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 adaptations contact info@LibreTexts.org or visit our main website at <https://LibreTexts.org>.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 09/15/2026

TABLE OF CONTENTS

Licensing

1: Chemistry to Chromosomes

- 1.1: The Structure of DNA
- 1.2: Chromosomes and chromatin
- 1.3: Mitosis**
 - 1.3.1: DNA replication
- 1.4: In vitro applications of DNA replication
- 1.5: Genomes and variation
- 1.6: Chapter 1 Review Questions

2: Central Dogma

- 2.1: Overview of Transcription
- 2.2: RNA processing
- 2.3: Genetic Code and Translation
- 2.4: Chapter 2 Review Questions

3: Mutation

- 3.1: Origins of Mutations
- 3.2: Consequences of mutations
- 3.3: Example - Mutations and Cystic Fibrosis
- 3.4: Ames Test
- 3.5: DNA Repair
- 3.6: Chapter 3 Review Questions

4: Inheritance

4.1: Meiosis

- 4.1.1: Homologous recombination
- 4.1.2: CRISPR

4.2: Mendelian Genetics

- 4.2.1: Monohybrid Crosses and Segregation
- 4.2.2: Dihybrid Crosses and Independent Assortment

4.3: Pedigrees

4.4: Exceptions to autosomal inheritance

- 4.4.1: Inheritance patterns for X-linked and Y-linked genes
- 4.4.2: Organellar Inheritance

4.5: Linkage

- 4.5.1: Linkage and Mapping
- 4.5.2: GWAS

4.6: Exceptions to simple dominance

4.7: Gene Interactions

4.7.1: Complementation Test

4.8: Chapter 4 Review Questions (draft)

5: Chromosome variation

5.1: Changes in Chromosome Number

5.2: Changes in Chromosome Structure

5.3: Chapter 5 Review Questions (Draft)

6: Regulation of Gene Expression

6.1: Prokaryotic gene regulation

6.1.1: The Use of Mutants to Study the lac Operon

6.1.2: The Lux Operon

6.2: Eukaryotic Gene Regulation

6.2.1: Gene Expression in Evolution

6.2.2: Regulating Transcription at the Chromatin Level

6.2.3: microRNAs

6.3: Epigenetics

6.4: Chapter 6 Review Questions (Draft)

6.5: Chapter 6 Review Questions Draft

7: Manipulating and Analyzing Genes and Genomes

7.1: Analysis of STRs

7.2: Transgenic Organisms

8: Cancer Genetics

8.1: Classification of Cancers

8.2: Hallmarks of Cancer

8.3: Mutagens and Carcinogens

8.4: Oncogenes

8.5: Tumor Suppressor Genes

8.6: Genetics and Targeted Cancer Treatment – The Story of Gleevec™ (Imatinib)

8.7: Cancer Genetics (Exercises)

8.8: Chapter 8 Review Questions

9: Homework

9.1: 1. Instructions

9.2: Homework 2

Index

Index

Glossary

Detailed Licensing

Licensing

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

1: CHEMISTRY TO CHROMOSOMES

1.1: The Structure of DNA

1.2: Chromosomes and chromatin

1.3: Mitosis

1.3.1: DNA replication

1.4: In vitro applications of DNA replication

1.5: Genomes and variation

1.6: Chapter 1 Review Questions

Template:HideTOC

This page titled [1: Chemistry to Chromosomes](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [1: Chemistry to Chromosomes](#) by [Stefanie West Leacock](#) has no license indicated.

1.1: The Structure of DNA

Learning Objectives

- Identify the sugar, phosphate, nitrogenous base, 5' and 3' carbons in a nucleotide and the key difference between DNA and RNA.
- Explain the structure of the double helix, including the role of hydrogen bonds and covalent (phosphodiester) bonds.
- Explain why the abundance of A is roughly equal to T and G is roughly equal to C in DNA.
- Using the rules of base pairing, predict the complementary (or anti-parallel) strand of DNA for a given sequence.

Nucleic acids are the most important macromolecules for the continuity of life. They carry the genetic blueprint of a cell and carry instructions for the functioning of the cell.

DNA and RNA

The two main types of nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is the genetic material found in all living organisms, ranging from single-celled bacteria to multicellular mammals. DNA is found in the nucleus of eukaryotes and in the organelles, chloroplasts, and mitochondria. In prokaryotes, the DNA is not enclosed in a membranous envelope.

The entire genetic content of a cell is known as its genome, and the study of genomes is genomics. In eukaryotic cells but not in prokaryotes, DNA forms a complex with histone proteins to form chromatin, the substance of eukaryotic chromosomes. A chromosome may contain tens of thousands of genes. Many genes contain the information to make protein products; other genes code for RNA products. DNA controls all of the cellular activities by turning the genes “on” or “off.”

The other type of nucleic acid, RNA, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus but instead use an intermediary to communicate with the rest of the cell. This intermediary is the messenger RNA (mRNA). Other types of RNA—like rRNA, tRNA, and microRNA—are involved in protein synthesis and its regulation.

DNA and RNA are made up of monomers known as nucleotides. The nucleotides combine with each other to form a polynucleotide, DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group (Figure 1.1.1). Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to one or more phosphate groups.

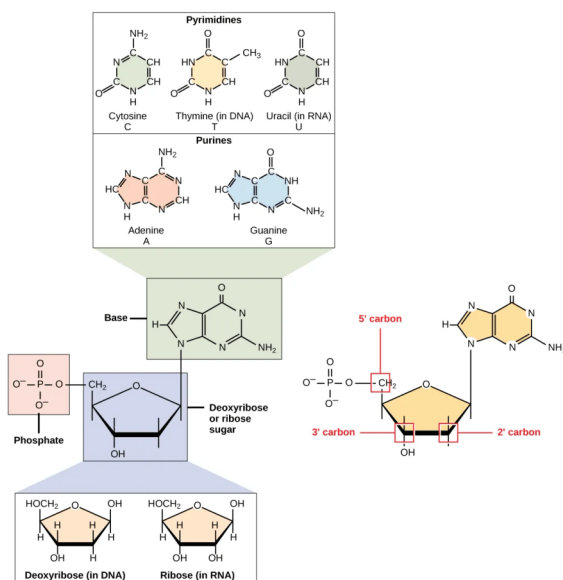


Figure 1.1.1: A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and one or more phosphate groups. Carbon residues in the pentose are numbered 1' through 5' (the prime distinguishes these residues from those in the base, which are numbered without using a prime notation). The base is attached to the 1' position of the ribose, and the phosphate is attached to the 5' position. The pentose sugar is deoxyribose in DNA and ribose in RNA. Deoxyribose has an H instead of an OH at the 2' position as in ribose. Bases can be divided into two categories: purines and pyrimidines. Purines have a double ring structure, and pyrimidines have a single ring. (Copyright OpenStax)

The nitrogenous bases, important components of nucleotides, are organic molecules and are so named because they contain carbon and nitrogen. They are bases because they contain an amino group that has the potential of binding an extra hydrogen, and thus, decreasing the hydrogen ion concentration in the environment, making it more basic. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G), cytosine (C), and thymine (T).

Adenine and guanine are classified as purines. The primary structure of a purine is two carbon-nitrogen rings. Cytosine, thymine, and uracil are classified as pyrimidines which have a single carbon-nitrogen ring as their primary structure (Figure 1.1.1). Each of these basic carbon-nitrogen rings has different functional groups attached to it. In molecular biology shorthand, the nitrogenous bases are simply known by their symbols A, T, G, C, and U. DNA contains A, T, G, and C whereas RNA contains A, U, G, and C.

The pentose sugar in DNA is deoxyribose, and in RNA, the sugar is ribose (Figure 1.1.1). The difference between the sugars is the presence of the hydroxyl group on the second carbon of the ribose and hydrogen on the second carbon of the deoxyribose (so deoxyribose is "missing" an -OH group). The carbon atoms of the sugar molecule are numbered as 1', 2', 3', 4', and 5' (1' is read as "one prime").

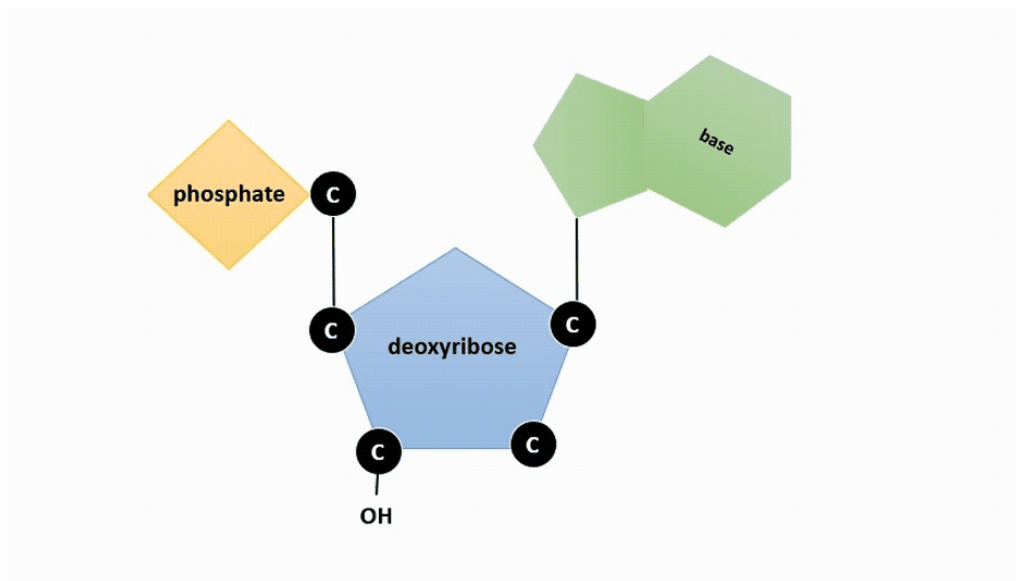


Figure 1.1.2: Animation highlighting carbons 1' through 5' on dexoyribose. (SWLeacock)

The phosphate residue is attached to the hydroxyl group of the 5' carbon of one sugar and the hydroxyl group of the 3' carbon of the sugar of the next nucleotide, which forms a 5'–3' phosphodiester linkage. The phosphodiester linkage is not formed by simple dehydration reaction like the other linkages connecting monomers in macromolecules: its formation involves the removal of two phosphate groups. A polynucleotide may have thousands of such phosphodiester linkages.

Chargaff's Rules

When **Watson and Crick** set out in the 1940's to determine the structure of DNA, it was already known that DNA is made up of a series four different types of molecules, called bases or nucleotides: adenine (A), cytosine (C), thymine (T), guanine (G). Watson and Crick also knew of **Chargaff's Rules**, which were a set of observations about the relative amount of each nucleotide that was present in almost any extract of DNA. Chargaff had observed that for any given species, the abundance of A was the same as T, and G was the same as C. This was essential to Watson & Crick's model.

Example 1.1.1

Chargaff determined the composition of nucleic acids in samples from a variety of species, including prokaryotes and eukaryotes. In one bacterial sample, the proportion of adenine was 15.5% (data adapted from Vischer et al, 1949). What proportion of guanine would have been present in this sample and why?

Solution

Because A pairs with T, the amount of T should be roughly equal to A, or approximately 15.5% percent. Thus, $A + T = 15.5 + 15.5 = 31\%$.

The percent of $G + C = 100\% - 31\% = 69\%$. Because G pairs with C, the amount of each of these should be roughly equal, so approximately 34.5% each.

Query 1.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



DNA Double-Helix Structure

Using proportional metal models of the individual nucleotides, Watson and Crick deduced a structure for DNA that was consistent with Chargaff's Rules and with x-ray crystallography data that was obtained (with some controversy) from another researcher named Rosalind Franklin. In Watson and Crick's famous **double helix**, each of the two strands contains DNA bases connected through covalent (phosphodiester) bonds to a sugar-phosphate backbone. Because one side of each sugar molecule is always connected to the opposite side of the next sugar molecule, each strand of DNA has polarity: these are called the 5' (5-prime) end and the 3' (3-prime) end, in accordance with the nomenclature of the carbons in the sugars. The two strands of the double helix run in **anti-parallel** (i.e. opposite) directions, with the 5' end of one strand adjacent to the 3' end of the other strand. The double helix has a **right-handed** twist, (rather than the left-handed twist that is often represented incorrectly in popular media). The DNA bases extend from the backbone towards the center of the helix, with a pair of bases from each strand forming hydrogen bonds that help to hold the two strands together. Under most conditions, the two strands are slightly offset, which creates a major groove on one face of the double helix, and a minor groove on the other.

Rules of base pairing

- A pairs with T (or U if RNA)
- G pairs with C

Because of the structure of the bases, A can only form hydrogen bonds with T, and G can only form hydrogen bonds with C (remember Chargaff's Rules). Each strand is therefore said to be complementary to the other, and so each strand also contains enough information to act as a template for the synthesis of the other. This complementary redundancy is important in DNA replication and repair. If the sequence of one strand is AATTGGCC, the complementary strand would have the sequence TTAACCGG. During DNA replication, each strand is copied, resulting in a daughter DNA double helix containing one parental DNA strand and a newly synthesized strand.

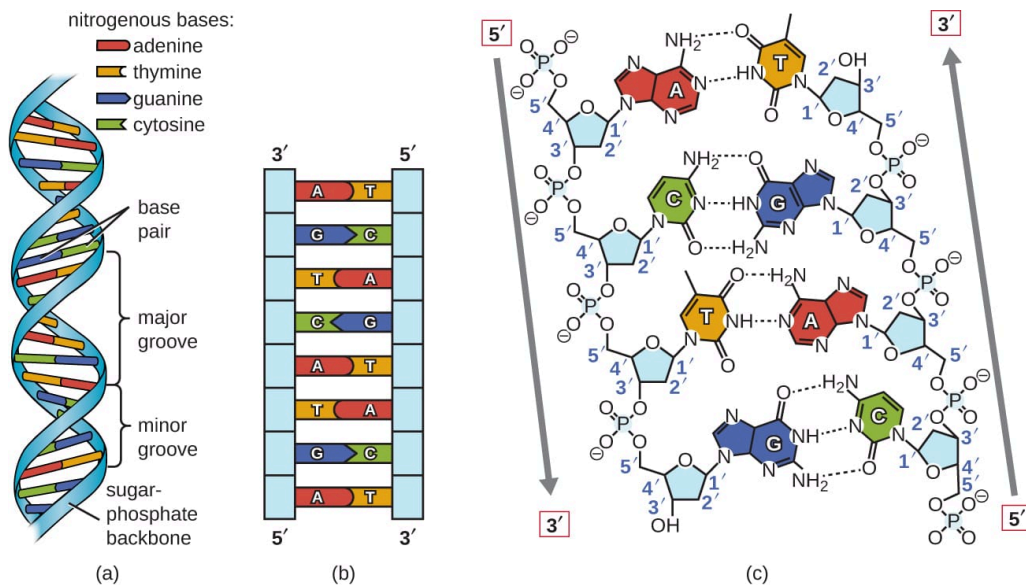


Figure 1.1.2: Watson and Crick proposed the double helix model for DNA. (a) The sugar-phosphate backbones are on the outside of the double helix and purines and pyrimidines form the “rungs” of the DNA helix ladder. (b) The two DNA strands are antiparallel to each other. (c) The direction of each strand is identified by numbering the carbons (1 through 5) in each sugar molecule. The 5' end is the one where carbon #5 is not bound to another nucleotide; the 3' end is the one where carbon #3 is not bound to another nucleotide and has a free hydroxyl group. (OpenStax-CNX)

3D structure of a DNA double helix

Spin the double helix to see the orientation of the sugars and phosphates in the backbone (ribbon in the model), the base pairs, major and minor grooves! (PDB ID = 1bna <https://www.rcsb.org/3d-view/1BNA>)

Implications of DNA structure

As for most biological molecules, the structure is important to the function, and the function of DNA is to contain information. Important properties that are derived from the DNA structure are:

- A complementary strand can always be synthesized from a single strand, due to the arrangement of hydrogen bonds between GC and AT bases.
- Hydrogen bonds stabilize the double helix, but can be broken when DNA needs to be accessed.
- The order of bases contains the information needed to code for amino acids in proteins during translation.
- Even sequences of DNA that do not encode amino acids can still provide information by interacting with proteins that function in DNA packaging and regulation. The major and minor grooves of DNA may determine which sequences are visible to DNA interacting proteins.

GET GROOVY with DNA

IF YOU LOOK THROUGH THE WINDOWS, WHAT DO YOU SEE

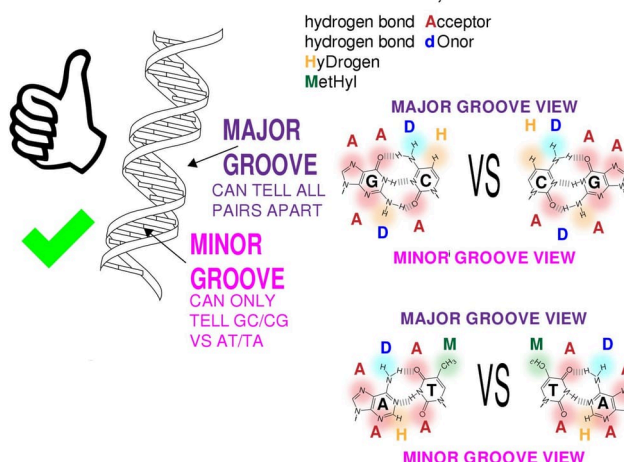


Figure 1.1.3: The significance of major and minor grooves in a DNA double helix. A DNA double helix twists in a right-handed fashion, just as the fingers on the right hand are "pointing" to the right when the right hand forms a "thumbs up." The predominant structure of a double helix results in major and minor grooves. The bases within the double helix interact with each other via hydrogen bonds, but the different bases pairs have different combinations atoms exposed in the major grooves. Proteins that recognize DNA sequences often do so by interacting with particular combinations of base pairs in major grooves based on these exposed atoms. In the minor grooves AT and TA base pairs appear the same, likewise GC and CG look the same in the minor groove. (Copyright By Biochemlife - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=73107644>)

Thinking ahead exercise 1.1.1

A mutation occurs, and cytosine is replaced with adenine. What impact do you think this will have on the DNA structure?

Answer available at the end of this section.

Thinking ahead exercise 1.1.2

If the two DNA strands were connected by covalent bonds rather than hydrogen bonds, what problems might occur, if any?

Answer available at the end of this section.

Contributors and Attributions

- Connie Rye (East Mississippi Community College), Robert Wise (University of Wisconsin, Oshkosh), Vladimir Jurukovski (Suffolk County Community College), Jean DeSaix (University of North Carolina at Chapel Hill), Jung Choi (Georgia Institute of Technology), Yael Avissar (Rhode Island College) among other contributing authors. Original content by OpenStax (CC BY 4.0; Download for free at <http://cnx.org/contents/185cbf87-c72...f21b5eabd@9.87>).

Access for free at <https://openstax.org/books/biology/pages/1-introduction>

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

References

VISCHER E, ZAMENHOF S, CHARGAFF E. Microbial nucleic acids; the desoxypentose nucleic acids of avian tubercle bacilli and yeast. *J Biol Chem*. 1949;177(1):429-438.

Answers

Thinking ahead exercise 1.1.1

Adenine is larger than cytosine and will not be able to base pair properly with the guanine on the opposing strand, causing the double helix to bulge at that position. DNA repair enzymes may recognize the bulge and replace the incorrect nucleotide.

Thinking ahead exercise 1.1.2

Covalent bonds are much stronger than hydrogen bonds and not as easily broken. If the DNA strands were connected by covalent bonds, then it would be much more difficult to unwind the double helix. This would be problematic for processes that require unwinding of the DNA molecule, such as replication and transcription.

This page titled [1.1: The Structure of DNA](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [1.1: The Structure of DNA](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

1.2: Chromosomes and chromatin

Learning Objectives

- Understand that chromosomes contain genes, which are DNA sequences that encode products and describe how the positions of individual genes on a given chromosome are related to their positions on the homolog of that chromosome.
- Discuss how DNA is packaged in the chromosomes in terms of histones, nucleosomes, and chromatin (heterochromatin and euchromatin).
- Explain the meaning of ploidy (haploid, diploid) and how it relates to the number of homologues of each chromosome.
- Compare prokaryotic and eukaryotic chromosomes.
- Interpret a karyotype.

What is a chromosome?

Chromosomes are units of DNA stored within cells. In prokaryotes, these units are most often circular, whereas in eukaryotes the units are typically linear.

Genes are sequences within chromosomes that contain information in the sequence of nucleotide bases that encodes a product (RNA or a protein).

Features and Compaction of Circular Chromosomes

The bacterial chromosome is typically one molecule of double-stranded, helical DNA. In most bacteria, the two ends of the double-stranded DNA covalently bond together to form both a physical and genetic circle. The chromosome is generally around 1000 μm long and frequently contains as many as 3,500 genes. *E. coli*, a bacterium that is 2-3 μm in length, has a chromosome approximately 1400 μm long.

What does it mean to be haploid?

Prokaryotic cells usually have one only copy of their chromosome(s) and therefore one copy of each gene on the chromosome.

To enable a macromolecule this large to fit within the bacterium, proteins bind to the DNA, segregating the DNA molecule into around 50 chromosomal domains and making it more compact. Prokaryotes primarily compact chromosomes by supercoiling, the process of twisting a piece of DNA that causes it to "fold up" on itself. Think of an old-fashioned phone cord or piece of string that you keep twisting. Supercoiling is not random, but is controlled by enzymes (topoisomerases) that can add or remove "twists" in the double helix to loosen or tighten the chromosome compaction.

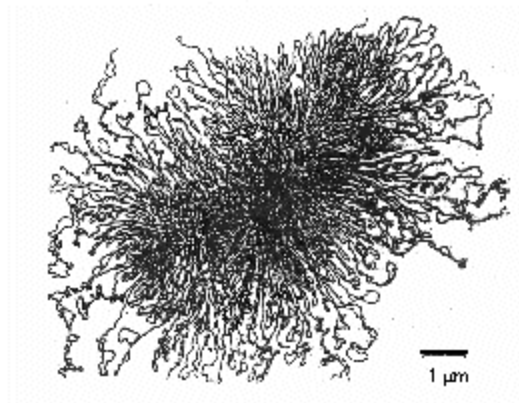


Figure 1.2.1 Electron Micrograph of a Bacterial Chromosome shows loops of DNA attached to a central core of proteins to compact the circular chromosome (From Microbiology (Kaiser))

Video: Representing Bacterial DNA

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.2.1: How we represent prokaryotic chromosomes and the location of genes (Map adapted from Duan et al, 2013 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058640>)

Features of Linear Chromosomes

Linear chromosomes contain structural features such as centromeres and telomeres. In most cases, each chromosome contains one **centromere**. These sequences are bound by proteins that will link the centromere to microtubules that facilitate chromosome movement during cell division. Under the microscope, centromeres of metaphase chromosomes can sometimes appear as constrictions in the body of the chromosome. If a centromere is located near the middle of a chromosome, it is **metacentric**, whereas a **telocentric** centromere is at, or near, the very end of the chromosome. Some organisms also do not have a single centromere but are **holocentric**. **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.

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 in the same order. **Non-homologous chromosomes** contain different gene loci, and are usually distinguishable based on cytological features such as length, centromere position, and banding patterns produced by staining.

What does it mean to be diploid?

Most eukaryotic organisms are **diploid**, meaning they have two sets of chromosomes. The diploid number is represented by **2n**.

Remember this means that each cell has:

- two homologs of each chromosome
- two copies of each gene

The number of non-homologous chromosomes varies by organism.

Levels of compaction in eukaryotes

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 (17,000 times smaller!). 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 on the level of gene activity required in any particular region of the chromosome.

There are several different levels of structural organization in eukaryotic chromosomes, with each successive level contributing to the further compaction of DNA. The compaction of DNA requires proteins and the combination of proteins and DNA is **chromatin**. 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**, which is composed of eight polypeptides, two copies of histone proteins H2A, H2B, H3, and H4. 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 fiber**. Subsequent levels of organization involve the addition of **scaffold proteins** that wind the 30nm fiber into coils, which are in turn wound around other scaffold proteins.

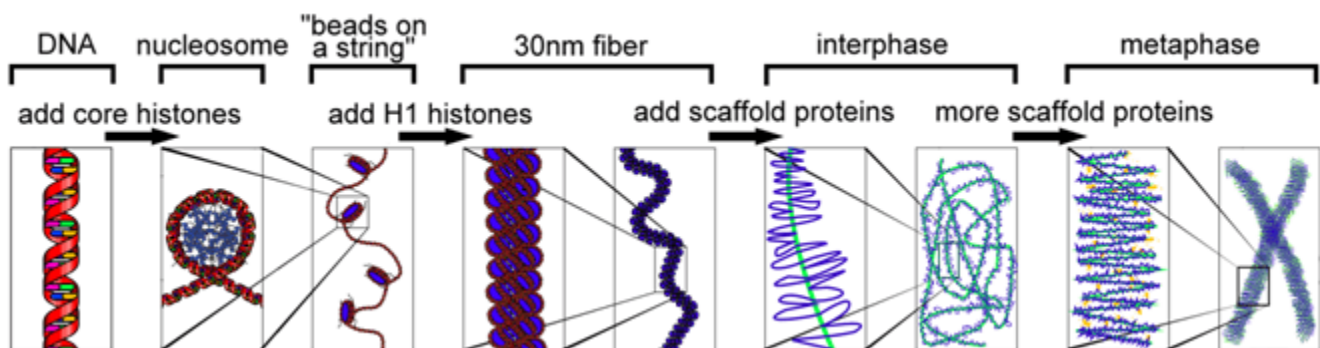


Figure 1.2.3: Successive stages of chromosome compaction depend on the introduction of additional proteins. (Wikipedia-R. Wheeler-GFDL)
Bottom: 3D structure of *Drosophila* nucleosome surrounding by a DNA double helix (PDB ID=2NQB <https://www.rcsb.org/3d-view/2NQB/1>)

Chromatin Packaging Varies Within a Chromosome: Euchromatin & Heterochromatin

Classically, there are two major types of chromatin, but these are more the ends of a continuous and varied spectrum. **Euchromatin** is more loosely packed, and tends to contain genes that are being transcribed (or actively being utilized by the cell). For example, there might be widely-spaced nucleosomes within a euchromatin region, leaving more of the DNA accessible for proteins that interact with that region. In contrast, **heterochromatin** usually contains densely-packed nucleosomes, is often rich in repetitive sequences, and tends not to contain genes that are actively being transcribed. Within these regions, nucleosomes might be close together, restricting protein access to DNA. Both the centromeres and telomeres are usually heterochromatin, whereas other regions of chromosomes can be switched from heterochromatin to euchromatin or vice versa, often by proteins that modify histones and nucleosomes.

Keeping chromosomes organized in nuclei

During interphase, the decondensed chromosomes often have specific locations within the nucleus and relative to one another, which has been studied using a technique called FISH, fluorescent in situ hybridization. In FISH, fluorescently-tagged probes (pieces of single-stranded DNA) recognize complementary sequences specific to each chromosome and allow visualization of

specific chromosome locations within a nucleus full of DNA.

Breaking down terms to understand FISH:

F = fluorescent because a small molecule that is excited by a certain wavelength of light is attached to the probe

IS = in situ is a Latin term that means "in the original place" because the experiment examines something within a cell

H = hybridization because the probe and the cellular DNA are complementary and therefore bind each other (or hybridize)

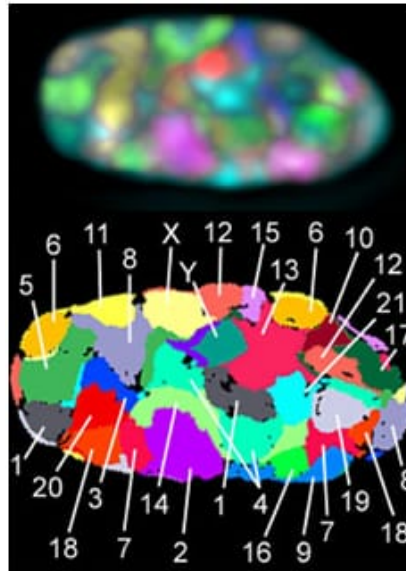


Figure 1.2.2 Chromosomes in a Human Nucleus Identified with Fluorescent Probes (CC-BY Bolzer et al., PLoS Biology 2005 (<https://doi.org/10.1371/journal.pbio.0030157>))

Exercise 1.2.1

How do DNA probes recognize sequences of DNA in cells in this experiment?

Answer available at the end of this section.

Karyotypes

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 when stained. A karyotype is a representation of a complete set of chromosomes. Karyotypes are usually determined by isolating mitotic chromosomes to view them as a karyogram. Chemicals that arrest the cells in metaphase of mitosis are used and then the chromosomes are released from nuclei, usually onto a slide. Chromosomes at this stage are at the most compact state. The images of these chromosomes can be captured digitally and arranged into pairs based on size, centromere position, and banding pattern to examine the complete set of chromosomes in the cell.

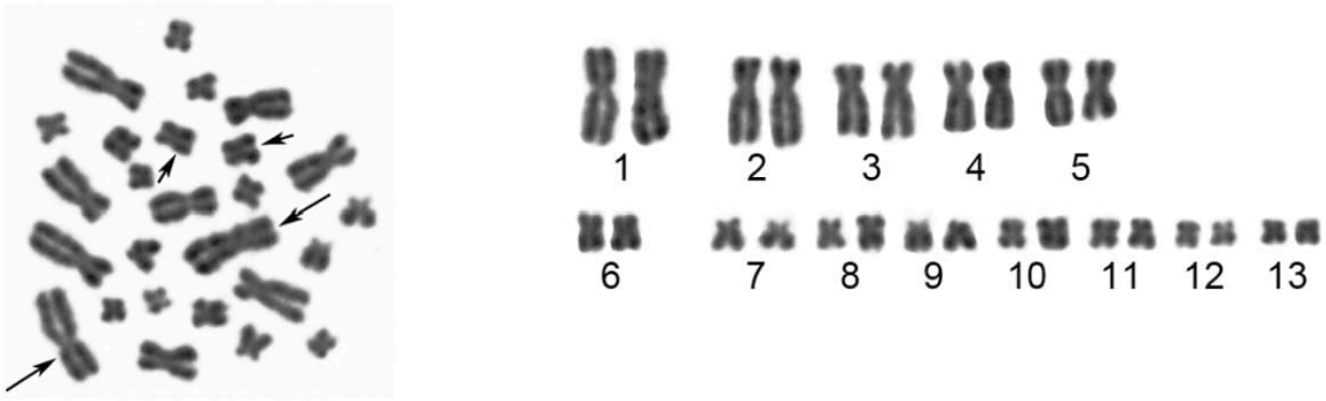


Figure 1.2.4: Karyotypes from a metaphase chromosome preparation from a spiny frog *Quasipaa boulengeri* (CC-BY Qing et al., 2012 *PLoS Biology* <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0046163>). Chromosomes are released from nuclei and randomly dispersed (left). Images of each chromosome are digitally arranged into homologous pairs (right).

Exercise 1.2.1

Based on the above karyotype from the spiny frog, what is the diploid ($2n$) number of this species?

Answer available at the end of this section.

Thinking ahead 1.2.2

A karyotype is a representation of a complete set of chromosomes. These chromosomes are typically extracted from a cell arrested in metaphase of mitosis. Why do you think the chromosomes are extracted during the mitotic phase instead of interphase?

Answer available at the end of this section.

Video: Performing Cytogenetic Test for Chromosomal Study

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Additional resources

For additional information about karyotyping: <https://www.nature.com/scitable/topicpage/karyotyping-for-chromosomal-abnormalities-298/>

References

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, et al. (2005) Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. PLoS Biol 3(5): e157. <https://doi.org/10.1371/journal.pbio.0030157>

Duan J, Jiang W, Cheng Z, Heikkila JJ, Glick BR (2013) The Complete Genome Sequence of the Plant Growth-Promoting Bacterium *Pseudomonas* sp. UW4. PLoS ONE 8(3): e58640. <https://doi.org/10.1371/journal.pone.0058640>

Qing L, Xia Y, Zheng Y, Zeng X (2012) A De Novo Case of Floating Chromosomal Polymorphisms by Translocation in *Quasipaa boulengeri* (Anura, Dicroglossidae). PLoS ONE 7(10): e46163. <https://doi.org/10.1371/journal.pone.0046163>

Contributors and Attributions

- Dr. Todd Nickle and Isabelle Barrette-Ng (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.
- Chromosome-DNA-gene image By Thomas Splettstoesser - This file has been altered from the original under a creative commons license. Original can be found at <https://commons.wikimedia.org/wiki/File:Chromosome-DNA-gene.png>, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=76810117>

Answers

Exercise 1.2.1

The rules of base pairing A-T and G-C and anti-parallel orientation! The probes are complementary to sequences of DNA specific to each chromosome.

Exercise 1.2.1

$$2n = 26$$

Thinking ahead 1.2.2

During interphase, chromosomes are in a decondensed configuration. However, during mitosis, chromosomes become highly condensed (17,000 times smaller than during interphase). It is much easier to extract chromosomes from a cell when they are tightly packaged as opposed to being in a loose configuration.

This page titled **1.2: Chromosomes and chromatin** is shared under a **CC BY-SA 4.0** license and was authored, remixed, and/or curated by **Stefanie West Leacock**.

- **1.2: Chromosomes and chromatin** by **Stefanie West Leacock** is licensed **CC BY-SA 4.0**.

1.3: Mitosis

Learning Objectives

- List and describe the key events that occur during the cell cycle and mitosis.
- Describe the movement of chromosomes and chromatids during mitosis and the role of microtubules.
- Determine the chromosome content of cells during various stages given information about the n or $2n$ number of the cell.

Copying and Dividing Chromosomes is a Cellular Challenge

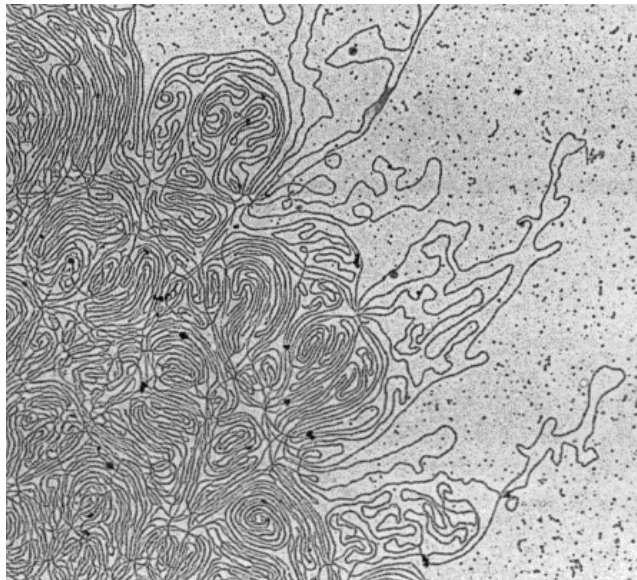


Figure 1.3.1: DNA released from human chromosome. It shows a bit (no more than 3%) of the single molecule of DNA released from a single human chromosome. (The chromosome was treated to remove its histones). Remembering that this is 3% of the DNA of only one of the 46 chromosomes in the human diploid cell, you can appreciate the problem faced by the cell of how to separate without error these great lengths of DNA without creating horrible tangles. (courtesy J. R. Paulson and U. C. Laemmli)

The solution to this challenge is:

1. Duplicate each chromosome during the S phase of the cell cycle.
2. This produces dyads, each made up of 2 identical sister chromatids. These are held together by a ring of proteins called cohesin.
3. Condense the chromosomes into a compact form. This requires ATP and protein complexes called condensins.
4. Separate the sister chromatids and
5. Distribute these equally between the two daughter cells

Cell Cycle

The cell cycle is an ordered series of events involving cell growth and cell division that produces two new daughter cells. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages of growth, DNA replication, and division that produces two identical (clone) cells. The cell cycle has two major phases: interphase and the mitotic phase. During interphase, the cell grows and DNA is replicated. During the mitotic phase, the replicated DNA and cytoplasmic contents are separated, and the cell divides.

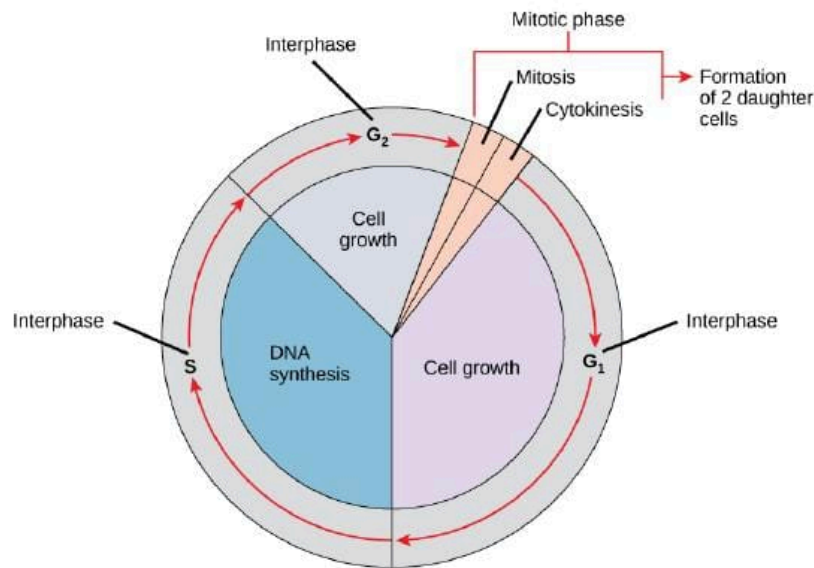


Figure 1.3.2: The cell cycle consists of interphase and the mitotic phase. During interphase, the cell grows and the nuclear DNA is duplicated. Interphase is followed by the mitotic phase. During the mitotic phase, the duplicated chromosomes are segregated and distributed into daughter nuclei. The cytoplasm is usually divided as well, resulting in two daughter cells. (OpenStaxCNX)

The cell cycle is an ordered series of events involving cell growth and cell division that produces two new daughter cells. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages of growth, DNA replication, and division that produces two identical (clone) cells. The cell cycle has two major phases: interphase and the mitotic phase. During interphase, the cell grows and DNA is replicated. During the mitotic phase, the replicated DNA and cytoplasmic contents are separated, and the cell divides.

Interphase

During interphase, the cell undergoes normal growth processes while also preparing for cell division. In order for a cell to move from interphase into the mitotic phase, many internal and external conditions must be met. The three stages of interphase are called G_1 , S, and G_2 .

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.3.1: Interphase | Cells | MCAT | Khan Academy

G₁ Phase (First Gap)

The first stage of interphase is called the G₁ phase (first gap) because, from a microscopic aspect, little change is visible. However, during the G₁ stage, the cell is quite active at the biochemical level. The cell is accumulating the building blocks of chromosomal DNA and the associated proteins as well as accumulating sufficient energy reserves to complete the task of replicating each chromosome in the nucleus.

S Phase (Synthesis of DNA)

Throughout interphase, nuclear DNA remains in a semi-condensed chromatin configuration. In the S phase, DNA replication can proceed through the mechanisms that result in the formation of identical pairs of DNA molecules—sister chromatids—that are firmly attached to the centromeric region via proteins called cohesins.

Counting chromosomes and chromatids

Note that the two sister chromatids, which are held together by cohesin proteins, are still counted as a single chromosome until they are split apart during mitosis.

The centrosome is also duplicated during the S phase. The two centrosomes will give rise to the mitotic spindle, the apparatus that orchestrates the movement of chromosomes during mitosis. At the center of each animal cell, the centrosomes of animal cells are associated with a pair of rod-like objects, the centrioles, which are at right angles to each other. Centrioles help organize cell division. Centrioles are not present in the centrosomes of other eukaryotic species, such as plants and most fungi.

G₂ Phase (Second Gap)

In the G₂ phase, the cell replenishes its energy stores and synthesizes proteins necessary for chromosome manipulation. Some cell organelles are duplicated, and the cytoskeleton is dismantled to provide resources for the mitotic phase. There may be additional cell growth during G₂. The final preparations for the mitotic phase must be completed before the cell is able to enter the first stage of mitosis.

What's in the gaps?

What specific types of proteins does the cell need to make during G₁? During G₂?

Mitosis

The mitotic phase is a multistep process during which the duplicated chromosomes are aligned, separated, and move into two new, identical daughter cells. The first portion of the mitotic phase is called karyokinesis, or nuclear division. The second portion of the mitotic phase, called cytokinesis, is the physical separation of the cytoplasmic components into the two daughter cells.

Mitosis is divided into a series of phases—prophase, prometaphase, metaphase, anaphase, and telophase—that result in the division of the cell nucleus. Karyokinesis is also called mitosis.

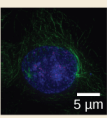
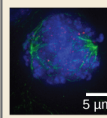
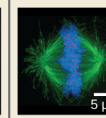
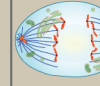
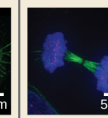
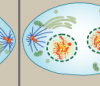
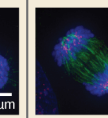
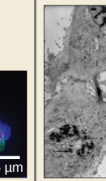
Prophase	Prometaphase	Metaphase	Anaphase	Telophase	Cytokinesis
 - Chromosomes condense and become visible - Spindle fibers emerge from the centrosomes - Nuclear envelope breaks down - Nucleolus disappears 	 - Chromosomes continue to condense - Kinetochores appear at the centromeres - Mitotic spindle microtubules attach to kinetochores - Centrosomes move toward opposite poles 	 - Mitotic spindle is fully developed, centrosomes are at opposite poles of the cell - Chromosomes are lined up at the metaphase plate - Each sister chromatid is attached to a spindle fiber originating from opposite poles 	 - Cohesin proteins binding the sister chromatids together break down - Sister chromatids (now called chromosomes) are pulled toward opposite poles - Non-kinetochore spindle fibers lengthen, elongating the cell 	 - Chromosomes arrive at opposite poles and begin to decondense - Nuclear envelope material surrounds each set of chromosomes - The mitotic spindle breaks down 	 - Animal cells: a cleavage furrow separates the daughter cells - Plant cells: a cell plate separates the daughter cells 
MITOSIS					

Figure 1.3.3: Karyokinesis (or mitosis) is divided into five stages—prophase, prometaphase, metaphase, anaphase, and telophase. The pictures at the bottom were taken by fluorescence microscopy (hence, the black background) of cells artificially stained by fluorescent dyes: blue fluorescence indicates DNA (chromosomes) and green fluorescence indicates microtubules (spindle apparatus). (credit “mitosis drawings”: modification of work by Mariana Ruiz Villareal; credit “micrographs”: modification of work by Roy van Heesbeen; credit “cytokinesis micrograph”: Wadsworth Center/New York State Department of Health; scale-bar data from Matt Russell)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.3.2: Real Microscopic Mitosis (Medical Research Community via www.youtube.com/watch?v=L61Gp_d7evo)

Prophase

During prophase, the nuclear envelope starts to dissociate into small vesicles, and the membranous organelles (such as the Golgi complex or Golgi apparatus, and endoplasmic reticulum), fragment and disperse toward the periphery of the cell. The nucleolus disappears (disperses). The centrosomes begin to move to opposite poles of the cell. Microtubules that will form the mitotic spindle extend between the centrosomes, pushing them farther apart as the microtubule fibers lengthen. The sister chromatids begin to coil more tightly with the aid of condensin proteins and become visible under a light microscope.

Prometaphase and Metaphase

During prometaphase, many processes that were begun in prophase continue to advance. The remnants of the nuclear envelope fragment. The mitotic spindle continues to develop as more microtubules assemble and stretch across the length of the former nuclear area. Chromosomes become more condensed and discrete. Each sister chromatid develops a protein structure called a kinetochore in the centromeric region. The proteins of the kinetochore attract and bind mitotic spindle microtubules. As the spindle microtubules extend from the centrosomes, some of these microtubules come into contact with and firmly bind to the kinetochores. Once a mitotic fiber attaches to a chromosome, the chromosome will be oriented until the kinetochores of sister chromatids face the opposite poles. Eventually, all the sister chromatids will be attached via their kinetochores to

microtubules from opposing poles. Spindle microtubules that do not engage the chromosomes are called polar microtubules. These microtubules overlap each other midway between the two poles and contribute to cell elongation. Astral microtubules are located near the poles, aid in spindle orientation, and are required for the regulation of mitosis.

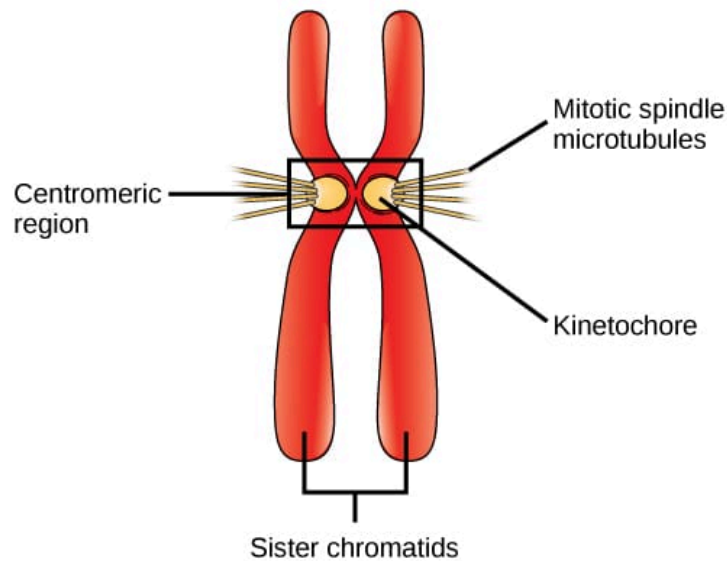


Figure 1.3.4: During prometaphase, mitotic spindle microtubules from opposite poles attach to each sister chromatid at the kinetochore. In anaphase, the connection between the sister chromatids breaks down, and the microtubules pull the chromosomes toward opposite poles.

During metaphase, all the chromosomes are aligned in a plane called the metaphase plate, or the equatorial plane, midway between the two poles of the cell. The sister chromatids are still tightly attached to each other by cohesin proteins. At this time, the chromosomes are maximally condensed.

Anaphase

During anaphase, the cohesin proteins degrade, and the sister chromatids separate at the centromere. Each chromatid, now called a chromosome, is pulled rapidly toward the centrosome to which its microtubule is attached. The cell becomes visibly elongated (oval shaped) as the polar microtubules slide against each other at the metaphase plate where they overlap.

Telophase

During telophase, the chromosomes reach the opposite poles and begin to decondense (unravel), relaxing into a chromatin configuration. The mitotic spindles are depolymerized into tubulin monomers that will be used to assemble cytoskeletal components for each daughter cell. Nuclear envelopes form around the chromosomes, and nucleosomes appear within the nuclear area.

Exercise 1.3.1

Which of the following is the correct order of events in mitosis?

1. Sister chromatids line up at the metaphase plate. The kinetochore becomes attached to the mitotic spindle. The nucleus reforms and the cell divides. Cohesin proteins break down and the sister chromatids separate.
2. The kinetochore becomes attached to the mitotic spindle. Cohesin proteins break down and the sister chromatids separate. Sister chromatids line up at the metaphase plate. The nucleus reforms and the cell divides.
3. The kinetochore becomes attached to the cohesin proteins. Sister chromatids line up at the metaphase plate. The kinetochore breaks down and the sister chromatids separate. The nucleus reforms and the cell divides.
4. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. Cohesin proteins break down and the sister chromatids separate. The nucleus reforms and the cell divides.

Answer available at the end of this section.

Exercise 1.3.2

Chemotherapy drugs such as vincristine and colchicine disrupt mitosis by binding to the protein tubulin (the subunit of microtubules) and interfering with microtubule assembly and disassembly. Exactly what mitotic structure is targeted by these drugs and what effect would that have on cell division?

Answer available at the end of this section.

Cytokinesis

Cytokinesis, or “cell motion,” is the second main stage of the mitotic phase during which cell division is completed via the physical separation of the cytoplasmic components into two daughter cells. Division is not complete until the cell components have been apportioned and completely separated into the two daughter cells. Although the stages of mitosis are similar for most eukaryotes, the process of cytokinesis is quite different for eukaryotes that have cell walls, such as plant cells.

In cells such as animal cells that lack cell walls, cytokinesis follows the onset of anaphase. A contractile ring composed of actin filaments forms just inside the plasma membrane at the former metaphase plate. The actin filaments pull the equator of the cell inward, forming a fissure. This fissure, or “crack,” is called the cleavage furrow. The furrow deepens as the actin ring contracts, and eventually the membrane is cleaved to separate the two new cells.

In plant cells, a new cell wall must form between the daughter cells. During interphase, the Golgi apparatus accumulates enzymes, structural proteins, and glucose molecules prior to breaking into vesicles and dispersing throughout the dividing cell. During telophase, these Golgi vesicles are transported on microtubules to form a phragmoplast (a vesicular structure) at the metaphase plate. There, the vesicles fuse and coalesce from the center toward the cell walls; this structure is called a cell plate. As more vesicles fuse, the cell plate enlarges until it merges with the cell walls at the periphery of the cell. Enzymes use the glucose that has accumulated between the membrane layers to build a new cell wall. The Golgi membranes become parts of the plasma membrane on either side of the new cell wall.

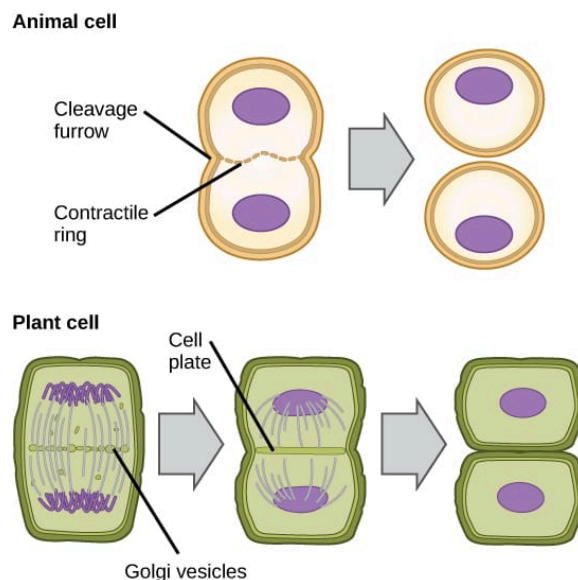


Figure 1.3.5: During cytokinesis in animal cells, a ring of actin filaments forms at the metaphase plate. The ring contracts, forming a cleavage furrow, which divides the cell in two. In plant cells, Golgi vesicles coalesce at the former metaphase plate, forming a phragmoplast. A cell plate formed by the fusion of the vesicles of the phragmoplast grows from the center toward the cell walls, and the membranes of the vesicles fuse to form a plasma membrane that divides the cell in two.

Exercise 1.3.3

Consider the events of interphase and mitosis. What cell cycle events will be affected in a cell that produces mutated (non-functional) condensin proteins?

Answer available at the end of this section.

Contributors and Attributions

- Connie Rye (East Mississippi Community College), Robert Wise (University of Wisconsin, Oshkosh), Vladimir Jurukovski (Suffolk County Community College), Jean DeSaix (University of North Carolina at Chapel Hill), Jung Choi (Georgia Institute of Technology), Yael Avissar (Rhode Island College) among other contributing authors. Original content by OpenStax (CC BY 4.0; Download for free at <http://cnx.org/contents/185cbf87-c72...f21b5eabd@9.87>).
- Figure 1.3.1:
- [John W. Kimball](#). This content is distributed under a Creative Commons Attribution 3.0 Unported (CC BY 3.0) license and made possible by funding from [The Saylor Foundation](#).

How to "count" the chromosomes

After DNA replication copies the chromosomes during S phase, sister chromatids remain held together by cohesin proteins. These sisters are still "counted" as one chromosome, until they separate during anaphase of mitosis, which ensures that each of the two new cells inherits one copy (sister chromatid) of each chromosome.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.3.3 : Following chromosome movements in mitosis

Example 1.3.1

If a cell has 10 chromosomes (total), how many chromosomes will be present during metaphase of mitosis? During anaphase? In the two daughter cells?

Solution

10 chromosomes (each made of two sister chromatids because DNA replication occurred in S phase before mitosis) are present during metaphase.

20 chromosomes are present during anaphase because the sister chromatids have been pulled apart.

After telophase and cytokinesis, the new daughter cells will each have 10 chromosomes, which is identical to the parental cell.

Exercise 1.3.4

Brown bears (*Ursus arctos*) have 74 ($2n=74$) chromosomes per diploid cell. How many chromosomes and sister chromatids are present at each of the following stages: G₁, G₂, anaphase of mitosis, and after cytokinesis?

Answer available at the end of this section.

Exercise 1.3.5

Human cells have 23 pairs of chromosomes. How many chromosomes are present during metaphase? During anaphase? In the cells that have completed mitosis?

Answer available at the end of this section.

Answers

Exercise 1.3.1

4

Exercise 1.3.2

The mitotic spindle is formed of microtubules. Microtubules are polymers of the protein tubulin; therefore, it is the mitotic spindle that is disrupted by these drugs. Without a functional mitotic spindle, the chromosomes will not be sorted or separated during mitosis. The cell will arrest in mitosis and die.

Exercise 1.3.3

If condensin is not functional, chromosomes are not packaged after DNA replication in the S phase of interphase. It is likely that the proteins of the centromeric region, such as the kinetochore, would not form. Even if the mitotic spindle fibers could attach to the chromatids without packing, the chromosomes would not be sorted or separated during mitosis.

Exercise 1.3.4

	# Chromosomes	# Chromatids
G₁	74	74
G₂	74	148
Anaphase of Mitosis	148	148
Cytokinesis	74	74

Exercise 1.3.5

Metaphase: 46 chromosomes

Anaphase: 92 chromosomes

After mitosis: Each cell has 46 chromosomes.

This page titled [1.3: Mitosis](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [1.4: Mitosis](#) by [Stefanie West Leacock](#) has no license indicated.

1.3.1: DNA replication

Learning Objectives

- Summarize the process of DNA replication, including bidirectional synthesis and explain the Meselson-Stahl experiment that demonstrated semiconservative replication.
- Recall that sister chromatids are identical copies of each other produced during S phase.
- Understand how DNA polymerase proofreads using exonuclease activity.
- Explain how telomerase replicates ends of chromosomes and the cellular significance of chromosome shortening, addressing the specific function of telomerase RNA.
- Compare prokaryotic and eukaryotic DNA replication.

DNA Replication

When the cell enters S (synthesis) phase in the cell cycle, all the chromosomal DNA must be replicated. **DNA polymerases** synthesize new strands by adding nucleotides to the 3'-OH group present on the previous nucleotide using the separated single strands of DNA as templates. This process generates two new double-stranded molecules, called sister chromatids, from one double helix. But how are the new and old strands distributed? The answer to this question was elucidated by classic experiments by Meselson and Stahl.

Videos about the experiment that demonstrated semiconservative DNA replication

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.3.1.1: THE MOST BEAUTIFUL EXPERIMENT IN BIOLOGY: Meselson & Stahl, The Semi-Conservative Replication of DNA. (Copyright YourekaScience via <https://youtu.be/4gdWOWjioBE>)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Figure 1.3.1.2: Matthew Meselson (Harvard): The Semi-Conservative Replication of DNA. (iBiology Science Stories via <https://youtu.be/V2evjmkur7k>)

The Mechanisms of Replication

Like many molecular events we will study, replication can be divided into three stages: initiation, elongation, and termination.

Initiation

DNA polymerase does not begin at random locations on chromosomes. DNA replication in both prokaryotes and eukaryotes begins at an **origin of replication** (Ori), which are specific sequences at specific positions on chromosomes. In *E. coli*, the *OriC* origin is ~245 bp in size. Chromosome replication begins with the binding of an initiator protein (DnaA) to an AT-rich sequence (for example TTATCCACA in *E. coli*) in the *OriC* and melts (disrupts the hydrogen bonding between) the two strands. **Helicase** enzymes will be added to unwind additional DNA from this point.

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. At least five prokaryotic DNA polymerases have been discovered to date with DNA Polymerase III (Pol III) being the primary polymerase for replication. Pol I is used primarily to fill in gaps created during lagging strand synthesis (described below) or through error-correcting mechanisms. DNA polymerase II, IV and V are used to synthesize DNA when certain types of repair is needed at other times in the cellular life cycle.

Thinking about sequences

Why would an initiator protein binding site in the *OriC* be an AT-rich region?

Hint: Think about the number of hydrogen bonds between AT and GC base pairs.

In 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 (6×10^9 base pairs). Even 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 than in prokaryotes, 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 S-phase.

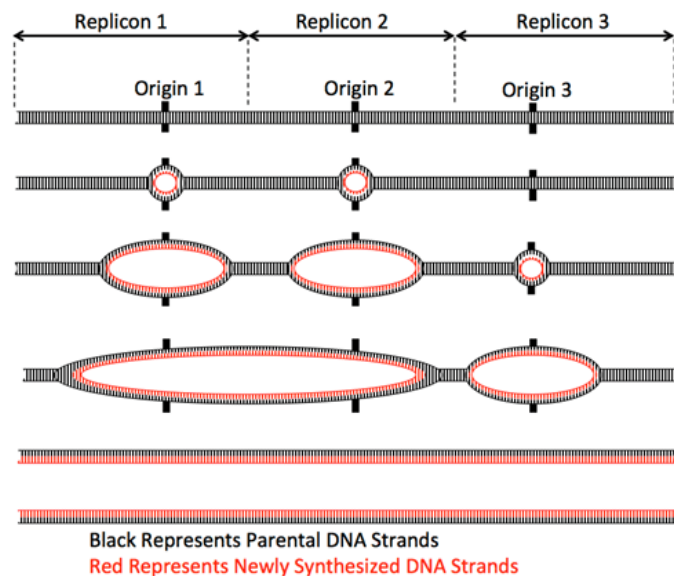


Figure 1.3.1.1: Schematic of a eukaryote chromosome showing multiple origins (1, 2, 3) of replication, each defining a replicon (1, 2, 3). Replication may start at different times in S-phase. Here #1 and #2 begin first then #3. As the replication forks proceed bidirectionally, they create "replication bubbles" that meet and form larger bubbles. The end result is two semi-conservatively replicated duplex DNA strands, with the parental strands in black and the newly-synthesized strands in red. (Original-Locke-CC:AN)

Elongation

As DNA polymerase proceeds along the template, the nucleotide that base pairs with each base on the template is covalently bonded to the 3' end of the growing strand. Note that the energy is provided by the nucleotide triphosphate itself; two phosphates

are released and one phosphate remains as a part of the phosphodiester bond.

Termination

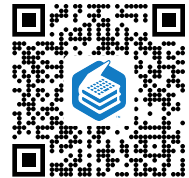
In prokaryotes, elongation proceeds bidirectionally until the replication forks meet. RNA primers are removed by a specialized DNA polymerase and then DNA is synthesized in their place. The resulting DNA fragments are then "sealed" together with DNA ligase, an enzyme that forms covalent phosphodiester bonds between two nucleotides.

In eukaryotes, replication also proceeds bidirectionally until adjacent forks meet, or the fork encounters the end of the chromosome.

Query 1.3.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Implications and limitations of the 5'-to-3' activity of DNA polymerase

Recall that enzymes are specific to their substrates. In the case of DNA polymerase, the structure only allows it to add nucleotides to the hydroxyl (-OH) group on the 3' carbon of an existing DNA strand, which presents some questions:

1. If the enzyme can only add nucleotides to an existing DNA strand (meaning that it cannot take two free nucleotides and attach them), how will it get started?
2. Because DNA is double stranded, each strand needs to be used as a template, but these strands are antiparallel and DNA polymerase replicates both strands via one complex of proteins. How can one complex make new DNA in opposite directions?
3. How will the 3' end be replicated when there is no longer a place for a primer on the complementary strand?

Obstacle #1: How to begin

Once *oriC* has been opened and the helicases have attached to the two sides of the replication fork, the replication machine, aka the replisome can begin to form. However, before the DNA polymerases take positions, they need to be primed. DNA polymerases are unable to join two individual free nucleotides together to begin forming a nucleic acid; they can only add onto a pre-existing strand of at least two nucleotides. Therefore, a specialized RNA polymerase (RNAP's do not have this limitation) known as **primase** is a part of the replisome, and reads creates a short RNA strand termed the *primer* for the DNA polymerase to add onto. Although only a few nucleotides are needed, the prokaryotic primers may be as long as 60 nt depending on the species. The helicase will continue to travel in front of the fork to unwind new DNA and allow primase to add new primers as needed.

Obstacle #2: Make two strands in opposite directions at the same time

Because DNA is being unwound in the direction of fork movement, both strands need to be synthesized in the unwound region at the same time. The two subunits of DNA polymerase that are adding nucleotides are actually tethered together, so they cannot travel in opposite directions.

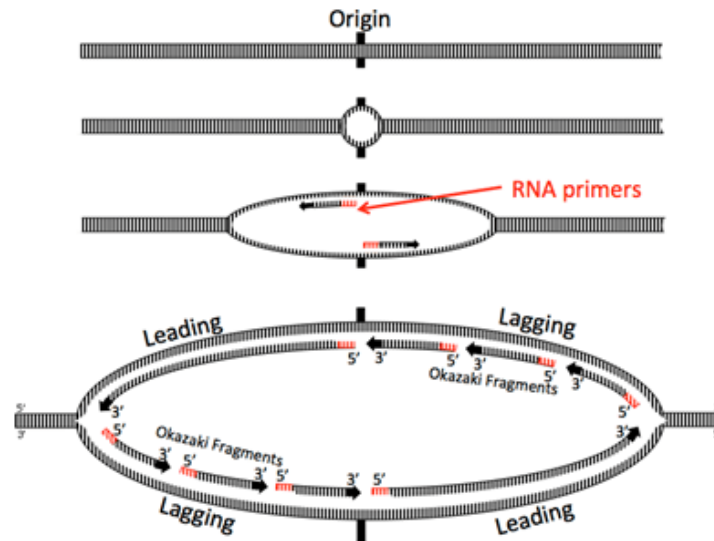


Figure 1.3.1.2: An origin of replication. The sequence specific DNA duplex is melted then the primase synthesizes RNA primers from which bidirectional DNA replication begins as the two replication forks head off in opposite directions. The leading and lagging strands are shown along with Okazaki fragments. Note the 5' and 3' orientation of all strands. (Original-Locke-CC:AN)

Helicase opens up the double stranded DNA and leads the rest of the replication machine along. Because nucleic acids are polymerized by adding the 5' phosphate of a new nucleotide to the 3' hydroxyl of the previous nucleotide, one of the strands, called the *leading strand*, is being synthesized in the same direction that the replication machine moves. No problem there.

The other strand is problematic: looked at linearly, the newly synthesized strand would be going 3' to 5' in the direction of helicase movement, but DNA polymerases cannot add nucleotides to the 5' end (remember that it has the phosphate, not a hydroxyl). How do cells resolve this problem? In the current model, the replication machine consists of the helicase, primases, and two DNA polymerase III holoenzymes moving in the same physical direction (following the helicase). In fact, the Pol III complexes are physically linked.

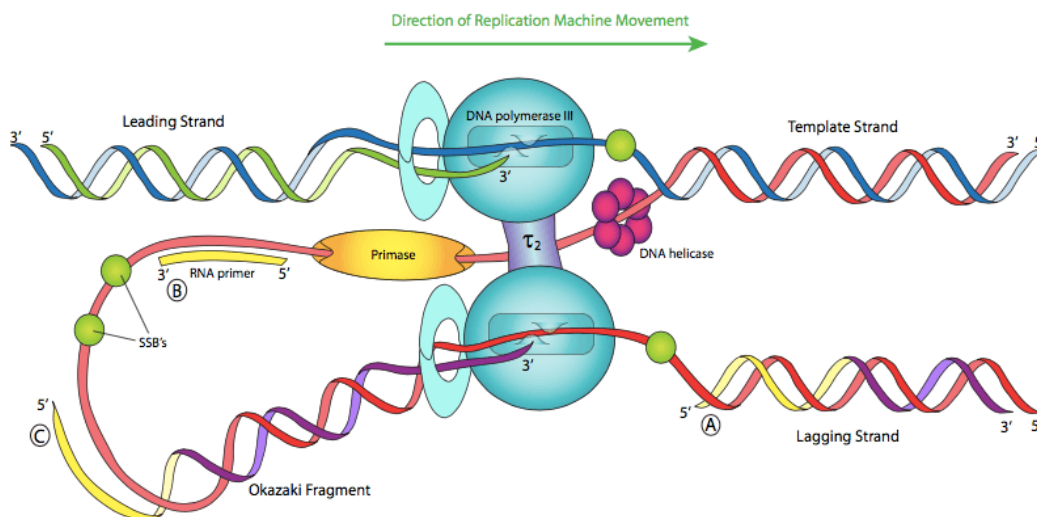


Figure 1.3.1.3: DNA Replication in prokaryotes. DNA polymerase III is a multi-subunit holoenzyme. The blue ring represents the β subunit (also known as the β clamp), a dimer of semicircular subunits that has a central hole through which the DNA is threaded. The core polymerase, via an α - β interaction, is attached to this β clamp so that it stays on the DNA longer, increasing the processivity of Pol III to over 5000nt.

For the other strand to be replicated, the strand must be fed into the polymerase backwards, by looping the DNA around. Moving behind DNA helicase, primase quickly synthesizes a short primer before having to move forward with the replisome and

synthesizing again, leaving intermittent primers in its wake. Thus, Pol III is forced to synthesize only short fragments of the chromosome at a time, called Okazaki fragments after their discoverers, Reiji and Tsuneko Okazaki. Pol III begins synthesizing by adding nucleotides onto the 3' end of a primer and continues until it hits the 5' end of the next primer. Pol III does not (and can not) connect the strand it is synthesizing with the 5' end of the RNA primer.

DNA replication is called a semi-discontinuous process because while the leading strand is being synthesized continuously, the lagging strand is synthesized in fragments. This leads to two major problems: first, there are little bits of RNA left behind in the newly made strands (just at the 5' end for the leading strand, in many places for the lagging); and second, Pol III can only add free nucleotides to a fragment of single stranded DNA; it cannot connect another fragment. Therefore, the new "strand" is not whole, but riddled with missing phosphodiester bonds.

The first problem is resolved by DNA polymerase I. Unlike Pol III, Pol I is a monomeric protein and acts alone, without additional proteins. There are also 10-20 times as many Pol I molecules as there are Pol III molecules, because they are needed for so many Okazaki fragments. DNA Polymerase I has three activities.

1. Like Pol III, it can synthesize a DNA strand based on a DNA template,
2. Also like Pol III, it is a 3'-5' proofreading exonuclease.
3. Unlike Pol III, it is also a 5'-3' exonuclease. The 5'-3' exonuclease activity is crucial in removing the RNA primer. The 5'-3' exonuclease binds to double-stranded DNA that has a single-stranded break in the phosphodiester backbone such as what happens after Okazaki fragments have been synthesized from one primer to the next, but cannot be connected. This 5'-3' exonuclease then removes the RNA primer. The polymerase activity then adds new DNA nucleotides to the upstream Okazaki fragment, filling in the gap created by the removal of the RNA primer. The proofreading exonuclease acts just like it does for Pol III, immediately removing a newly incorporated incorrect nucleotide. After proofreading, the overall error rate of nucleotide incorporation is approximately 1 in 10^7 .

Even though the RNA has been replaced with DNA, this still leaves a fragmented strand. The last major player in the DNA replication story finally appears: DNA ligase. This enzyme has one simple but crucial task: it catalyzes the attack of the 3'-OH from one fragment on the 5' phosphate of the next fragment, generating a phosphodiester bond.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](https://youtu.be/I9ArIJWYZHI)



Video 1.3.1.3: Mechanism of DNA replication. (DNA Learning Center via <https://youtu.be/I9ArIJWYZHI>)

Obstacle #3: How to copy the end?

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 where it would begin to lose essential gene sequences. Thus, this DNA must be replicated. Most eukaryotes solve this problem with a specialized DNA polymerase called **telomerase**, in combination with a regular polymerase. Telomerases are RNA-directed DNA polymerases that are **riboproteins**, as they are composed of both protein and RNA. 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 human telomeric DNA has the complementary sequence 5'-TTAGGG-3' repeated thousands of times. After telomerase has made the first strand, a primase synthesizes an RNA primer and a DNA polymerase then makes a complement of the extended sequence.

Telomerase = non-coding RNA + reverse transcriptase protein

In humans, the gene for the telomerase RNA is named TERC (telomerase RNA component) and is found on chromosome 3. This gene therefore encodes a product (an RNA) but not a protein!

The gene for the protein is named TERT (telomerase reverse transcriptase) and is found on chromosome 5.

These independently-encoded and independently-produced gene products must assemble to perform their function.

Note that the number of repeats, and thus the size of the telomere, is not set, but may fluctuate 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.

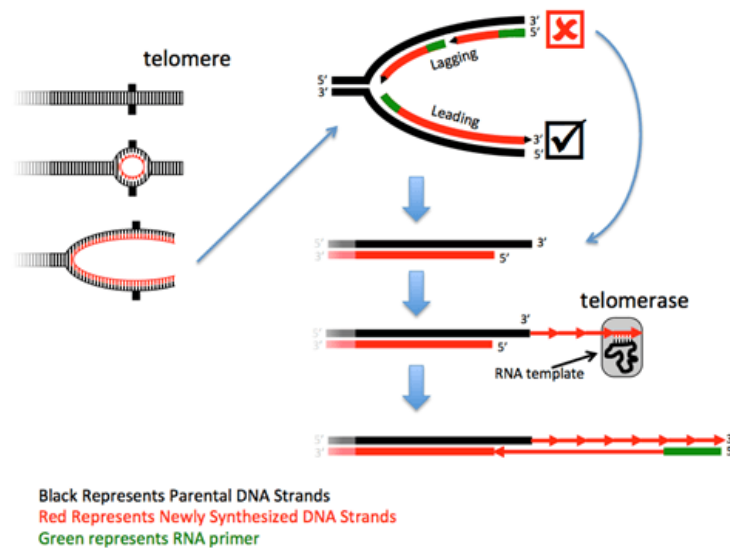


Figure 1.3.1.4: Telomere replication showing the completion of the leading strand and incomplete replication of the lagging strand. The gap is replicated by the extension of the 3' end by telomerase and then filled in by extension from an RNA primer. (Original-Locke-CC:AN)

Query 1.3.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



In the absence of telomerase, as is the case in human somatic cells, repeated cell division leads to telomere shortening. If a critical limit is reached, the cells enter a senescence phase of non-growth. The activation of telomerase expression, which often occurs in cancer cells, permits a cell and its descendants to become immortal because their telomeres will not reach the critical limit. Cells that can be grown in culture, such as **HeLa cells** that overexpress telomerase can be propagated essentially indefinitely. HeLa cells, which were originally isolated from an African-American cancer patient, Henrietta Lacks, without her consent, have been kept in culture since 1951.

DNA polymerase proofreading

DNA polymerase incorporates the correct nucleotides into each new strand based on the favorable hydrogen bonding that occurs between bases that bind using the rules of base pairing (A-T and G-C). If an incorrect base is incorporated, the structure

of the double helix will be distorted and will often be detected by the enzyme if it has not proceeded too far down the strand. DNA polymerase has 3' to 5' exonuclease activity to remove nucleotides through the mismatch and then repeat synthesis of that region of the strand with the correct nucleotides. As a result of this proofreading ability, DNA polymerases have very low error rates, allowing organisms to successfully pass genetic information from cell to cell and from generation to generation.

Thinking ahead 1.3.1.1

If a nucleotide that was missing the OH group at the 3' end was added to the growing strand by DNA polymerase, what do you think would happen to replication?

- (a) Replication would continue as normal
- (b) Replication would stop as another nucleotide cannot be added to this nucleotide

Answer available at the end of this section.

Contributors and Attributions

Todd Nickle and Isabelle Barrette-Ng Online Open Genetics

E.V. Wong, Cells - Molecules and Mechanisms ([https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book%3A_Cells_-_Molecules_and_Mechanisms_\(Wong\)/7%3A_DNA/7.3%3A_Prokaryotic_Replication](https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book%3A_Cells_-_Molecules_and_Mechanisms_(Wong)/7%3A_DNA/7.3%3A_Prokaryotic_Replication))

Answers

Thinking ahead 1.3.1.1

(b) Replication would stop as another nucleotide cannot be added to this nucleotide. Deoxyribose nucleotides that are missing the 3' OH are called dideoxyribose nucleotides and are utilized in Sanger DNA sequencing. The lack of a free OH prevents DNA polymerase from adding on another nucleotide, causing DNA synthesis to terminate.

This page titled [1.3.1: DNA replication](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [1.5: DNA replication](#) by [Stefanie West Leacock](#) has no license indicated.

1.4: In vitro applications of DNA replication

Learning Objectives

- Compare cellular DNA replication with the human applications of PCR and sequencing methods.
- Explain the role of primers, *Taq* polymerase, and the temperature cycles in PCR.
- Explain the relationship between the structure of dideoxynucleotides and their function in sequencing reactions.
- Describe the process of traditional (Sanger) sequencing.
- Give an example of a NextGen sequencing method and briefly explain how it differs from traditional sequencing.

Understanding and adapting cellular DNA replication has led to innovations that have had an immeasurable impact on research, biotechnology, and medicine. Both polymerase chain reaction (PCR) and DNA sequencing are in vitro applications of cellular DNA replication.

PCR

Molecular cloning was the first method available to isolate a gene of interest and make many copies of it to obtain sufficient amounts of the DNA to study. Today, there is a faster and easier way to obtain large amounts of a DNA sequence of interest: the polymerase chain reaction (PCR). PCR allows one to use the power of DNA replication to amplify DNA enormously in a short period of time. As you know, cells replicate their DNA before they divide, and in doing so, double the amount of the cell's DNA. PCR essentially mimics cellular DNA replication in the test tube, repeatedly copying the target DNA over and over, to produce large quantities of the desired DNA.

Selective Replication

In contrast to cellular DNA replication, which amplifies all of a cell's DNA during a replication cycle, PCR does targeted amplification to replicate only a segment of DNA bounded by the two primers that determine where DNA polymerase begins replication. Each cycle of PCR involves three steps, denaturing, annealing and extension, each of which occurs at a different temperature.

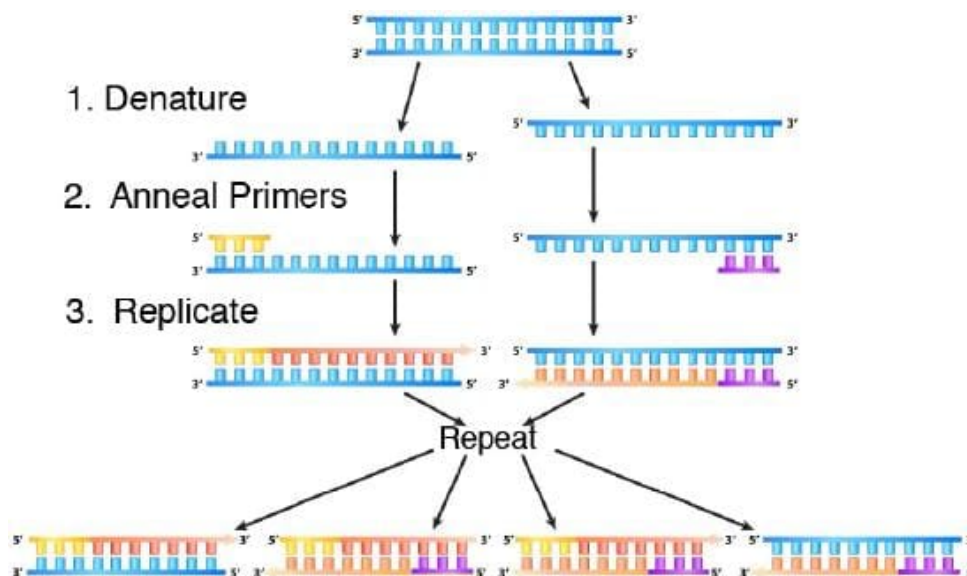


Figure 1.4.1: Steps in the polymerase chain reaction. Image by Aleia Kim

The Starting Materials

Since PCR is, basically, replication of DNA in a test-tube, many of the usual ingredients needed for DNA replication are required:

- A template (the DNA containing the target sequence that is being copied)
- Primers (to initiate the synthesis of the new DNA strands)
- Thermostable DNA polymerase (to carry out the synthesis). The polymerase needs to be heat stable, because heat is used to separate the template DNA strands in each cycle.
- dNTPs (DNA nucleotides to build the new DNA strands).
- The template is the DNA that contains the target you want to amplify (the "target" is the specific region of the DNA you want to amplify).

The primers are short synthetic single-stranded DNA molecules whose sequence is complementary to a region flanking the target sequence. These DNA molecules can be chemically-synthesized to produce any given base sequence to use as primers. Primers are present in millions of fold excess over the template, because each new DNA strand will begin from a primer. Therefore, it is necessary to know a little bit of the template sequence on either side of the region of DNA to be amplified. DNA polymerases and dNTPs are commercially available from biotechnology supply companies.

1. The first step of the process involves separating the strands of the target DNA by heating to near boiling.
2. Next, the solution is cooled to a temperature that favors complementary DNA sequences finding each other and making base pairs, a process called annealing. Since the primers are present in great excess, the complementary sequences they target are readily found and base-paired to the primers. These primers direct the synthesis of DNA. Only where a primer anneals to a DNA strand will replication occur, since DNA polymerases require a primer to begin synthesis of a new strand.
3. In the third step in the process, the DNA polymerase replicates DNA by extension from the 3' end of the primer, making a new DNA strand. At the end of the first cycle, there are twice as many DNA molecules, just as in cellular replication. But in PCR, steps 1-3 are repeated, usually for between 25 and 30 cycles. At the end of the process, there is a theoretical yield of 230 (over 1 billion times) more DNA than there was to start. (This enormous amplification power is the reason that PCR is so useful for forensic investigations, where very tiny amounts of DNA may be available at a crime scene.)

The temperature cycles are controlled in a thermocycler, which repeatedly raises and lowers temperatures according to the set program. Since each cycle can be completed in a couple of minutes, the entire amplification can be completed very rapidly. The resulting DNA is analyzed on a gel to ensure that it is of the expected size, and depending on what it is to be used for, may also be sequenced, to be certain that it is the desired fragment.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.4.1: PCR amplification uses repeated temperature cycling to denature double-stranded DNA, anneal primers, and synthesize copies of double-stranded DNA. (Copyright DNA Learning Center via youtube.com)

Example 1.4.1

PCR is used for one type of test for SARS-CoV-2, the virus associated with COVID-19. Why would this tool be helpful to detect the virus? What materials would be needed to perform this test?

Solution

Because PCR exponentially amplifies the number of copies of genetic material, even when only a small amount of genetic material is present, many copies can be produced and detected. In this case, a small amount of virus would be extremely hard to detect without PCR. However, this virus has an RNA (not DNA) genome, so an extra step must convert RNA to DNA, then copies can be made!

You would need: a sample of cells from an individual to isolate genetic material from, Taq polymerase and appropriate buffer, dNTPs, primers specific to the virus genetic sequence

Query 1.4.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



DNA sequencing

DNA sequencing determines the order of bases in a piece of DNA, a whole chromosome, or even a whole genome. Remember that DNA obeys the rules of base pairing, so determining the sequence of one strand means that we know the other strand as well. DNA sequencing is an application of DNA replication, because DNA strands are built using nucleotides or methods that allow the scientist to visualize which base was incorporated at each position.

Dideoxy sequencing (also known as Sanger sequencing)

The original method developed by Sanger and others relied on variant nucleotides (dNTPs), dideoxynucleotides (ddNTPs), which are also known as chain terminators. Remember, that DNA synthesis always proceeds from 5' to 3' by attaching the phosphate of the incoming nucleotide to the 3' OH of the previous nucleotide. When that hydroxyl group is absent as in a dideoxynucleotide, synthesis ends.

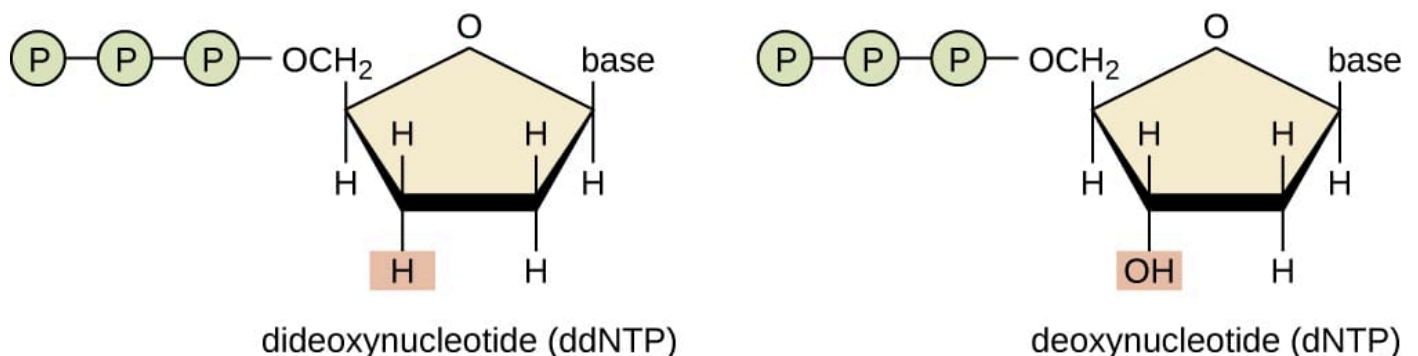


Figure 1.4.4: A dideoxynucleotide has a hydrogen on the 3' carbon, whereas a deoxynucleotide has a hydroxyl group. (<https://cnx.org/contents/5CvTdmJL@4.4>, CC BY 4.0, <https://commons.wikimedia.org/w/index.php?curid=53713264>)

When a combination of dNTPs and ddNTPs are included, a mixture of DNA sequences of different lengths is produced. Each fragment ends with a labeled dideoxynucleotide. Note that this process first utilized radioactively-labeled nucleotides; only one type of radioactively-labeled dideoxynucleotide (ddATP, ddGTP, ddTTP, or ddCTP) could be included in each reaction, meaning four individual reactions had to be performed. The use of fluorescent labels for the dideoxynucleotides simplified this process, allowing all four bases to be sequenced in a single reaction. Ordering the mixture of fragments by size using electrophoresis and

reading the wavelength of each label using a laser will reveal the order of bases.

The reagents needed for a sequencing reaction include:

- Template DNA (the DNA you want to know the sequence of)
- A mixture of all four deoxynucleotides (dATP, dCTP, dGTP, AND dTTP)
- Fluorescently-labeled dideoxynucleotide (ddATP, ddCTP, ddGTP, OR ddTTP)
- DNA polymerase (and buffer for efficient enzyme function)
- DNA oligonucleotide primer (complementary to the sequence 5' of the unknown sequence)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)

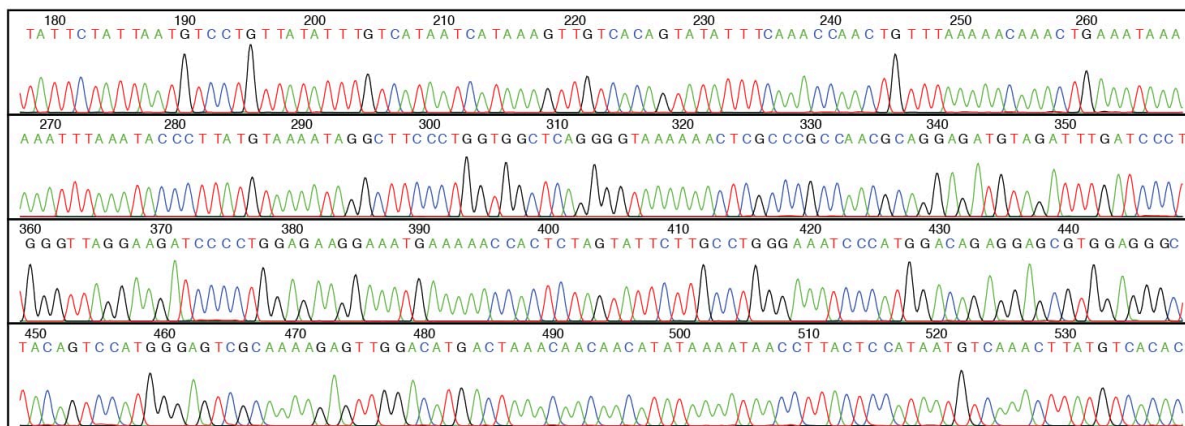


Video 1.4.2: See how the mixture of fragments produced by this reaction can be used to determine a sequence of DNA in this video. Note that whether a dNTP or ddNTP is incorporated at each position occurs at random. By chance, all possible fragments lengths will be generated in a successful sequencing reaction. (Copyright SWLeacock own work)

Visualizing the sequence

After the sequencing reaction is completed, the randomly-terminated fragments are separated by size using electrophoresis. If the reaction proceeded optimally, the products are a mixture of fragments with each base represented at the end of a fragment. Each fragment ends with a fluorescently-labeled dideoxyribonucleotide. The smallest fragments will migrate the fastest and pass by a laser that will detect the intensity of the fluorescence and record the corresponding base. The results can be

visualized in a plot that represents the peaks of fluorescence intensity at each size.



DNA sequence data from an automated sequencing machine

Figure 1.4.5: DNA sequence data from an automated sequencing machine. (Public domain via [genome.gov](https://www.genome.gov))

Exercise 1.4.1

1. Starting at the A at nucleotide 450 in the figure above, give 20 nucleotides of sequence that recorded? (Show the 5' and 3' ends.)
2. Based on your answer above, what was the sequence of the template DNA strand? (Show the 5' and 3' ends.)

Answer available at the end of this section.

Next-generation sequencing

Traditional Sanger sequencing is still used for some sequencing applications, particularly for small pieces of DNA. However, newer technologies have improved the speed and lowered the cost of DNA sequencing.

Three of these next-generation technologies (named for the companies that developed them) are highlighted here, but others exist and new ones are likely.

Illumina Sequencing by Synthesis

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.4.3: The Illumina method uses sequencing-by-synthesis by creating clusters of DNA fragment copies. Incorporation of different nucleotides produces fluorescent signals as the DNA polymerase synthesizes DNA complementary to the fragments. (Illumina via youtube.com)

Pacific Biosciences

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 1.4.4: In PacBio sequencing, adaptor sequences are added to the ends of double-stranded DNA to create circular molecules. These molecules are then used as templates for DNA polymerase to synthesize new DNA strands; the sequence is read as light is emitted when each new nucleotide is added. (Pacific Biosciences https://www.youtube.com/embed/_ID8JyAbwEo)

Nanopore

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Videos 1.4.4: Nanopore sequencing uses a current to detect which bases are passing through a small pore. (National DNA Database <https://youtu.be/E9-Rm5AoZGw> and Oxford Nanopore Technologies <https://youtu.be/RcP85JHLmnI>)

Compare and contrast

- Compare and contrast cellular DNA replication with DNA sequencing, addressing how DNA sequencing depends on the rules of base pairing, DNA structure, and DNA synthesis.
- Compare and contrast traditional sequencing and next-generation methods.

Exercise 1.4.1

Which platform of next generation sequencing does not involve sequencing the DNA as it is being synthesized by a DNA polymerase?

- A. Sanger sequencing
- B. Illumina sequencing
- C. Pacific Biosciences sequencing
- D. Nanopore sequencing

Answer available at the end of this section.

Answers

Exercise 1.4.1

1. 5' ACAGTCCATGGGAGTCGCAA 3'
2. 3' TGTCAGGTACCCTCAGCGTT 5'

Exercise 1.4.1

D. In Oxford Nanopore sequencing a double-stranded DNA is separated and one strand is processed through the "pore" and changes in current are measured as nucleotides pass through the channel. DNA polymerase is not synthesizing the strand while it is being sequenced.

This page titled [1.4: In vitro applications of DNA replication](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [1.6: In vitro applications of DNA replication](#) by [Stefanie West Leacock](#) has no license indicated.

1.5: Genomes and variation

Learning Objectives

- Distinguish between genes, chromosomes, and genomes.
- Describe the range in the size of genomes and numbers of genes between species.
- Distinguish between SNPs and STRs as variations in DNA sequence.
- List applications of the field of genomics.

What is a genome?

The complete set of an organism's genetic information is the **genome**.

Variation exists when sequences of bases differ between chromosomes. Some examples of types of variation include:

- differences between the base sequence of homologous chromosomes in an individual
- differences between the base sequence of chromosomes in two individuals
- differences in base sequences between two different species
- differences in numbers of chromosomes between species
- differences in arrangement of similar sequences between species (synteny)
- differences in base sequence individuals with and without a disease or trait

....and more

The study of sequences at the whole genome level is known as **genomics**.

Why study genomes?

Understanding genomes has a variety of applications including:

- comparing species to understand gene function and evolution
- understanding health differences between individuals
- performing research into gene function using model organisms
- identifying individuals or species

Genome sequencing projects

The first organismal genome to be sequenced was the prokaryote *H. influenzae* in 1995 (Fleischmann et al, 1995).

Determining the sequence of the entire human genome took more than two decades from its initial official start in 1990 to the draft sequence published in 2001 and the refinement of that draft sequence over the next several years (review key dates in the timeline here: <https://www.genome.gov/about-nhgri/Brief-History-Timeline#one>). Sequences produced by genome projects are referred to as **reference genomes**.

Competing genome projects?

The results of human genome sequencing efforts were published at the same time in *Nature* (results of the International Human Genome Sequencing Consortium) and *Science* (by Craig Venter and colleagues with Celera Genomics) in February 2001. These two groups had taken somewhat different approaches to sequencing efforts.

However, there were still some "gaps" in the sequence. For example, regions that are highly repetitive or similar to other chromosomes are difficult to sequence precisely. The next-generation sequencing techniques described previously have facilitated obtaining complete chromosome sequences for some of these missing regions. Recently, the complete sequence of an X chromosome from "telomere-to-telomere" was published (Miga et al., 2020), followed by a complete assembly of chromosome 8 (Logsdon et al., 2021).

Exercise 1.5.1

Whose genome was the first sequenced?

For the Celera project, refer to the Sources of DNA and Sequencing Methods section of this paper <https://science.sciencemag.org/content/291/5507/1304.long>.

For the International consortium project, the sample collection is reported in this paper <https://www.nature.com/articles/35057062>.

Answer available at the end of this section.

How big are genomes? And does it matter?

Genomes vary in size and complexity. Bacterial genomes are the smallest in numbers of base pairs, followed by fungi and algae. Beyond these organisms, genome size does not necessarily increase with complexity. Plants have some of the largest genomes (often because they have more than two sets of chromosomes). However, regardless of the total genome size, most organisms does not have much more than 25,000 genes, so total size does not necessarily increase gene number.

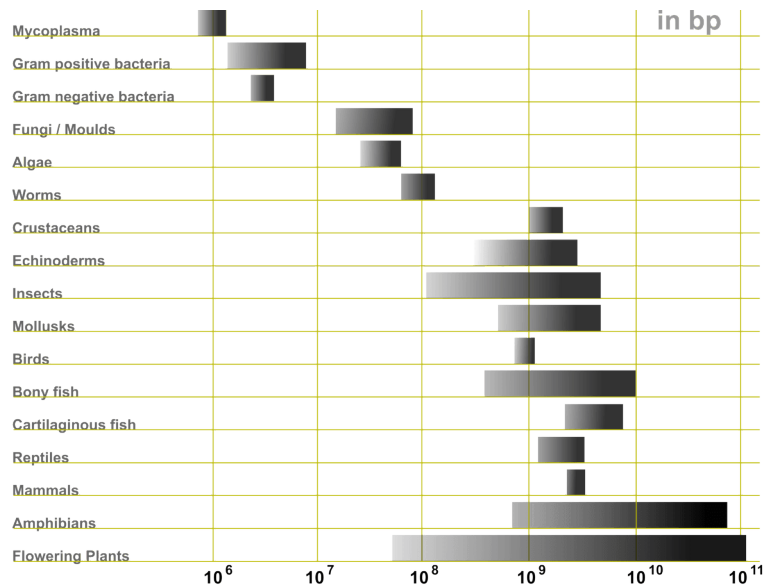


Figure 1.5.1: Genome sizes (in base pairs) vary by organism. (CC:BY-SA Abizar via Wikipedia Commons)

Explore genome data

Remember that much genome research is funded by grants from governments and is made publicly accessible. Some databases that maintain genome information are:

- National Center for Biotechnology Information (NCBI) Genome Data Viewer <https://www.ncbi.nlm.nih.gov/gdv/>
- European Molecular Biology Laboratory's European Bioinformatics Institute Ensembl <https://www.ensembl.org/index.html>
- University of California Santa Cruz Genomics Institute <https://genome.ucsc.edu/>

Variation: SNPs and STRs

Genome sequencing projects reveal the reference genome that is largely shared between species or individuals, but they also reveal variation. Although other types of variation exist, two major classes include **single nucleotide polymorphisms (SNPs)** and **short tandem repeats (STRs)**. SNPs are positions along chromosomes that can vary in the nucleotide at that position. STRs are repeats of a few nucleotides that can vary in the number of repeats.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



References

Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. 1995;269(5223):496-512. doi:10.1126/science.7542800

International Human Genome Sequencing Consortium., Whitehead Institute for Biomedical Research, Center for Genome Research., Lander, E. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001). <https://doi.org/10.1038/35057062>

Logsdon, G.A., Vollger, M.R., Hsieh, P. et al. The structure, function and evolution of a complete human chromosome 8. *Nature* **593**, 101–107 (2021). <https://doi.org/10.1038/s41586-021-03420-7>

Miga, K.H., Koren, S., Rhie, A. et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **585**, 79–84 (2020). <https://doi.org/10.1038/s41586-020-2547-7>

Milo, Ron and Philips, Rob. *Cell Biology by the Numbers*. Garland Science, 2015.

Answers

Exercise 1.5.1

Celera enrolled 21 donors initially, but only 5 of these were used for the complete sequencing. What factors determined which samples were selected?

For the International consortium, volunteer donors were solicited from areas near research labs. After sample collection and processing all identifying information was removed. Many more samples were collected than were actually used. So truly, we don't know whose genome is included!

This page titled [1.5: Genomes and variation](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [1.3: Genomes and variation](#) by [Stefanie West Leacock](#) has no license indicated.

1.6: Chapter 1 Review Questions

1.1: The Structure of DNA

Query 1.6.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.6.1.5

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



1.2: Chromosomes and Chromatin

Query 1.6

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.7

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.8

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.9

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.10

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.11

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



1.3: Mitosis

Query 1.12

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.13

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.14

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.15

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



1.3.1: DNA Replication

Query 1.16

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.17

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.18

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.19

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.20

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.21

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



1.4: In Vitro Applications of DNA Replication

Query 1.22

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 1.23

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



1.5: Genomes and Variation

Query 1.24

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



2: CENTRAL DOGMA

- 2.1: Overview of Transcription
- 2.2: RNA processing
- 2.3: Genetic Code and Translation
- 2.4: Chapter 2 Review Questions

Template:HideTOC

This page titled [2: Central Dogma](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [2: Central Dogma](#) by [Stefanie West Leacock](#) has no license indicated.

2.1: Overview of Transcription

Learning Objectives

- Identify the key steps of transcription, the function of the promoter and the function of RNA polymerase.
- Distinguish between coding (RNA-like) and non-coding (template) strands of DNA. Understand that within a single piece of DNA, either strand can be used as the template for different genes, but the RNA will still be produced from 5' → 3'.
- Draw a line diagram showing a segment of DNA from a gene and its RNA transcript, indicating which DNA strand is the template, the direction of transcription and the polarities of all DNA and RNA strands.
- Give examples of non-coding RNA molecules.

What is transcription?

Consider that all of the cells in a multicellular organism have arisen by division from a single fertilized egg and therefore, all have the same DNA. Division of that original fertilized egg produces, in the case of humans, over a trillion cells, by the time a baby is produced from that egg (that's a lot of DNA replication!). Yet, we also know that a baby is not a giant ball of a trillion identical cells, but has the many different kinds of cells that make up tissues like skin and muscle and bone and nerves. How did cells that have identical DNA turn out so different?

The answer lies in gene expression, which is the process by which the information in DNA is used. Although all the cells in a baby have the same DNA, each different cell type uses a different subset of the genes in that DNA to direct the synthesis of a distinctive set of RNAs and proteins. The first step in gene expression is transcription, the process of copying information from DNA sequences into RNA sequences. This process is also known as DNA-dependent RNA synthesis. When a sequence of DNA is transcribed, only one of the two DNA strands is copied into RNA, when this RNA encodes a protein is it known as messenger RNA (mRNA).

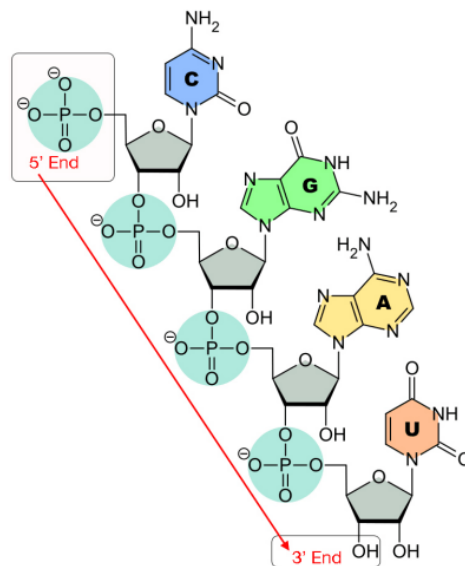


Figure 2.1.1: Structure of a single strand of RNA

Important features of transcription

- All RNA, mRNA as well as tRNA, rRNA, microRNA and more, is produced by transcription.
- Only one strand of DNA is used as a template by enzymes called RNA polymerases
- RNA is synthesized from 5' to 3'.

- RNA polymerases do not need primers to begin transcription.
- The four ribonucleotide triphosphates (rNTPs) are ATP, GTP, UTP, and CTP.
- RNA polymerases begin transcription at DNA sequences called promoters.
- RNA polymerases end transcription at sequences called terminators.

In transcription, an RNA polymerase uses only one strand of DNA, called the template strand, of a gene to catalyze synthesis of a complementary, antiparallel RNA strand. RNA polymerases use ribose nucleotide triphosphate (NTP) precursors, in contrast to DNA polymerases, which use deoxyribose nucleotide (dNTP) precursors (compared on page 1.1: [The Structure of DNA](#)). In addition, RNAs incorporate uracil (U) nucleotides into RNA strands instead of the thymine (T) nucleotides used in DNA. RNA polymerases differ from DNA polymerases in that they do not require primers. With the help of transcription initiation factors, RNA polymerase locates the transcription start site of a gene and begins synthesis of a new RNA strand from scratch by joining the two ribonucleotides that are complementary to the first two bases of the template strand.

Overview of the Stages of Transcription

The basic steps of transcription are initiation, elongation, and termination. Here we can identify several of the DNA sequences that characterize a gene. The promoter is the binding site for RNA polymerase. It usually lies 5' to, or upstream of the transcription start site, where it will start unwinding the double helix and begin synthesizing new RNA. The transcribed grey DNA region in each of the three panels are the **transcription unit** of the gene. Termination sites are typically 3' to, or downstream from the transcribed region of the gene. By convention, *upstream* refers to DNA 5' to a given reference point on the DNA (e.g., the transcription start-site of a gene). *Downstream* then, refers to DNA 3' to a given reference point on the DNA.

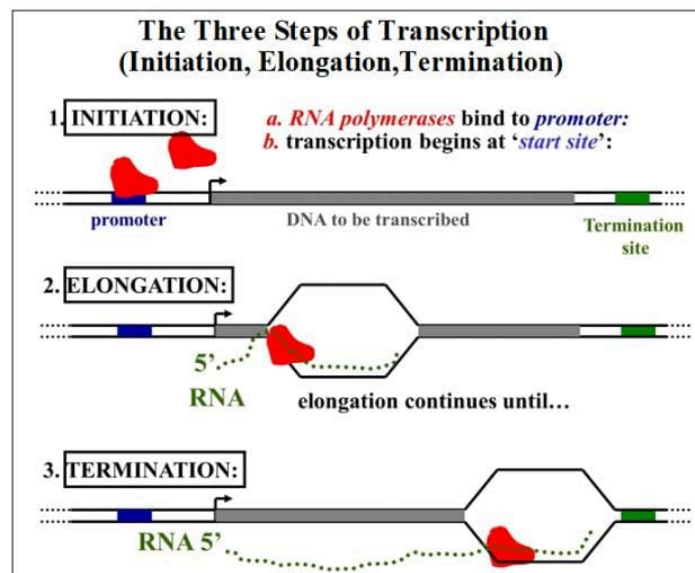


Figure 2.1.2: Three steps of transcription. (Copyright)

RNA polymerase

Building an RNA strand is very similar to building a DNA strand. This is not surprising, knowing that DNA and RNA are very similar molecules. What enzyme carries out transcription? Transcription is catalyzed by the enzyme RNA Polymerase. "RNA polymerase" is a general term for an enzyme that makes RNA. There are many different RNA polymerases.

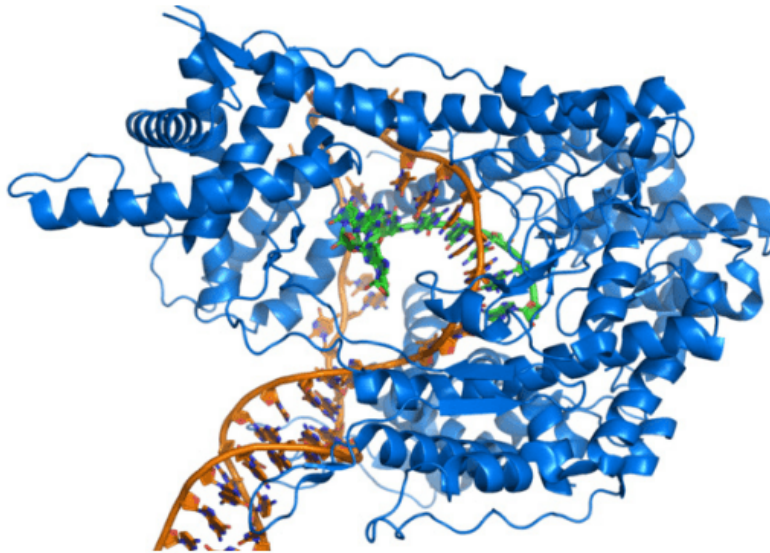


Figure 2.1.3: T7 RNA Polymerase (blue) making RNA (green) using DNA template (brown)

Like DNA polymerases, RNA polymerases synthesize new strands only in the 5' to 3' direction, but because they are making RNA, they use ribonucleotides (i.e., RNA nucleotides) rather than deoxyribonucleotides. Ribonucleotides are joined in exactly the same way as deoxyribonucleotides, which is to say that the 3'OH of the last nucleotide on the growing chain is joined to the 5' phosphate on the incoming nucleotide.

One important difference between DNA polymerases and RNA polymerases is that the latter do not require a primer to start making RNA. Once RNA polymerases are in the right place to start copying DNA, they just begin making RNA by stringing together RNA nucleotides complementary to the DNA template.

This, of course, brings us to an obvious question- how do RNA polymerases "know" where to start copying on the DNA. Unlike the situation in replication, where every nucleotide of the parental DNA must eventually be copied, transcription, as we have already noted, only copies selected genes into RNA at any given time. What indicates to an RNA polymerase where to start copying DNA to make a transcript? Signals in DNA indicate to RNA polymerase where it should start (and end) transcription. These signals are special sequences in DNA that are recognized by the RNA polymerase or by proteins that help RNA polymerase determine where it should bind the DNA to start transcription. A DNA sequence at which the RNA polymerase binds to start transcription is called a promoter.

A promoter is generally situated upstream of the gene that it controls. What this means is that on the DNA strand that the gene is on, the promoter sequence is "before" the gene. Remember that, by convention, DNA sequences are read from 5' to 3'. So the promoter lies 5' to the start point of transcription.

Also notice that the promoter is said to "control" the gene it is associated with. This is because expression of the gene is dependent on the binding of RNA polymerase to the promoter sequence to begin transcription. If the RNA polymerase and its helper proteins do not bind the promoter, the gene cannot be transcribed and it will therefore, not be expressed.

Gene	-35 sequence	-10 sequence	Transcription start site
<i>rrnE1</i>	CAATTTTCTATTGAGGAATGAGGAGAACTCCC	TATAATGCGCCTCC	ATC
<i>tRNA^{trp}</i>	CAACGTAACACTTTACAGCGGCGCGTCATTTGATATGATGCGCCCC	GCT	
<i>trp</i>	AAATGAGCTGTTTGACAATTAATCATCGAACTAGT	TAACTAGTACGCA	AG
<i>araBAD</i>	GGATCCTACCTGACGCTTTTATCGCAACTCTC	TACTGTTTCTCCAT	AC
<i>araC</i>	GCCGTGATTATAGACACTTTTGTTACGCGTTTT	TGTCATGGCTTTG	GTC
<i>bioA</i>	TTCCAAAACGTGTTTTTTTGTTGTTAATTCGGTG	TAGACTTGTA	ACCTA

Figure 2.1.4: Promoter Sequences

What is special about a promoter sequence? In an effort to answer this question, scientists looked at many genes and their surrounding sequences. It makes sense that because the same RNA polymerase has to bind to many different promoters, the promoters should have some similarities in their sequences. Sure enough, common sequence patterns were seen to be present in many promoters. We will first take a look at prokaryotic promoters. When prokaryotic genes were examined, the following features commonly emerged:

- A transcription start site (this the base in the DNA across from which the first RNA nucleotide is paired).
- A -10 sequence: this is a 6 bp region centered about 10 bp upstream of the start site. The consensus sequence at this position is TATAAT. In other words, if you count back from the transcription start site, which by convention, is called the +1, the sequence found at -10 in the majority of promoters studied is TATAAT).
- A -35 sequence: this is a sequence at about 35 basepairs upstream from the start of transcription. The consensus sequence at this position is TTGACA.

What is the significance of these sequences? It turns out that the sequences at -10 and -35 are recognized and bound by a subunit of prokaryotic RNA polymerase before transcription can begin.

The RNA polymerase of *E. coli*, for example, has a subunit called the **sigma (σ) subunit** (or sigma factor) in addition to the core polymerase, which is the part of the enzyme that actually makes RNA. Together, the sigma subunit and core polymerase make up what is termed the **RNA polymerase holoenzyme**. The sigma subunit of the polymerase can recognize and bind to the -10 and -35 sequences in the promoter, thus positioning the RNA polymerase at the right place to initiate transcription. Once transcription begins, the core polymerase and the sigma subunit separate, with the core polymerase continuing RNA synthesis and the sigma subunit wandering off to escort another core polymerase molecule to a promoter. The sigma subunit can be thought of as a sort of usher that leads the polymerase to its "seat" on the promoter.

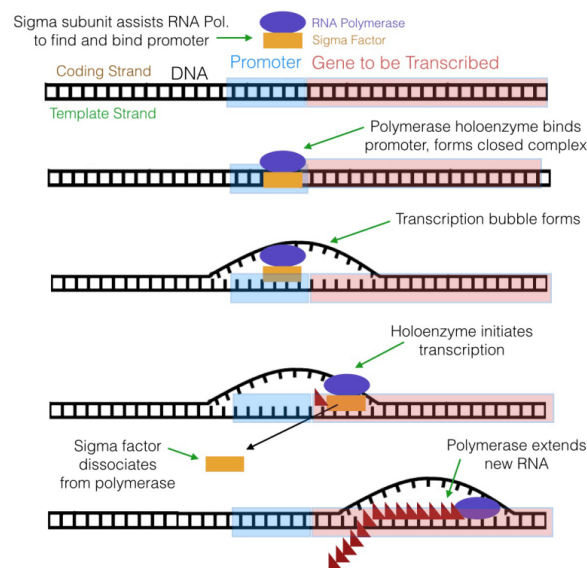


Figure 2.1.5: Transcription initiation in *E. coli*

As already mentioned, an RNA chain, complementary to the DNA template, is built by the RNA polymerase by the joining of the 5' phosphate of an incoming ribonucleotide to the 3'OH on the last nucleotide of the growing RNA strand. How does the polymerase know where to stop? A sequence of nucleotides called the terminator is the signal to the RNA polymerase to stop transcription and dissociate from the template.

Although the process of RNA synthesis is the same in eukaryotes as in prokaryotes, there are some additional issues to keep in mind in eukaryotes. One is that in eukaryotes, the DNA template exists as chromatin, where the DNA is tightly associated with histones and other proteins. The "packaging" of the DNA must therefore be opened up to allow the RNA polymerase access to the template in the region to be transcribed.

A second difference is that eukaryotes have multiple RNA polymerases, not one as in bacterial cells. The different polymerases transcribe different genes. For example, RNA polymerase I transcribes the ribosomal RNA genes, while RNA polymerase III copies tRNA genes. The RNA polymerase we will focus on most is RNA polymerase II, which transcribes protein-coding genes to make mRNAs.

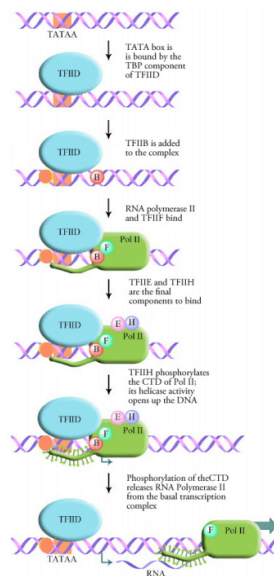


Figure 2.1.6: Assembly of basal transcription complex and initiation of transcription

All three eukaryotic RNA polymerases need additional proteins to help them get transcription started. In prokaryotes, RNA polymerase by itself can initiate transcription (remember that the sigma subunit is a subunit of the prokaryotic RNA polymerase). The additional proteins needed by eukaryotic RNA polymerases are referred to as transcription factors.

Finally, in eukaryotic cells, transcription is separated in space and time from translation. Transcription happens in the nucleus, and the mRNAs produced are processed further before they are sent into the cytoplasm. Protein synthesis (translation) happens in the cytoplasm. In prokaryotic cells, mRNAs can be translated as they are coming off the DNA template, and because there is no nucleus, transcription and protein synthesis occur in a single cellular compartment.

Like genes in prokaryotes, eukaryotic genes also have promoters. Eukaryotic promoters commonly have a TATA box, a sequence about 25 base pairs upstream of the start of transcription that is recognized and bound by proteins that help the RNA polymerase to position itself correctly to begin transcription. (Some eukaryotic promoters lack TATA boxes, and have, instead, other recognition sequences to help the RNA polymerase find the spot on the DNA where it binds and initiates transcription.)

We noted earlier that eukaryotic RNA polymerases need additional proteins to bind promoters and start transcription. What are these additional proteins that are needed to start transcription? General transcription factors are proteins that help eukaryotic RNA polymerases find transcription start sites and initiate RNA synthesis. We will focus on the transcription factors that assist RNA polymerase II. These transcription factors are named TFIIA, TFIIB and so on (TF= transcription factor, II=RNA polymerase II, and the letters distinguish individual transcription factors).

Transcription in eukaryotes requires the general transcription factors and the RNA polymerase to form a complex at the TATA box called the basal transcription complex or transcription initiation complex. This is the minimum requirement for any gene to be transcribed. The first step in the formation of this complex is the binding of the TATA box by a transcription factor called the TATA Binding Protein or TBP. Binding of the TBP causes the DNA to bend at this spot and take on a structure that is suitable for the binding of additional transcription factors and RNA polymerase. As shown in the figure at left, a number of different general transcription factors, together with RNA polymerase (Pol II) form a complex at the TATA box.

The final step in the assembly of the basal transcription complex is the binding of a general transcription factor called TFIIH. TFIIH is a multifunctional protein that has helicase activity (i.e., it is capable of opening up a DNA double helix) as well as kinase activity. The kinase activity of TFIIH adds a phosphate onto the C-terminal domain (CTD) of the RNA polymerase. This phosphorylation appears to be the signal that releases the RNA polymerase from the basal transcription complex and allows it to move forward and begin transcription.

Contributors and Attributions

- Dr. Kevin Ahern and Dr. Indira Rajagopal (Oregon State University)

Either DNA strand can be a template

The promoter is the sequence of DNA that encodes the information about where to begin transcription for each gene. Depending on the promoter, either strand of DNA can be used as the template strand.

Watch this video to see how either strand of DNA can be used as a template for different genes on the same chromosome.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



template vs. non-template strands summary

- The **template strand** is the one that RNA polymerase uses as the basis to build the RNA. This strand is also called the **non-coding strand** or the **antisense strand**.
- The **non-template strand** has the identical sequence of the RNA (except for the substitution of U for T). This strand is also called the **coding strand** or **sense strand**.

Major Types of Cellular RNA

Cells make several different kinds of RNA:

- mRNAs that code for proteins
- rRNAs that form part of ribosomes
- tRNAs that serve as adaptors between mRNA and amino acids during translation
- MicroRNAs that regulate gene expression
- Other small RNAs that have a variety of functions.

Contributors and Attributions

This page titled [2.1: Overview of Transcription](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [2.1: Overview of Transcription](#) by [Stefanie West Leacock](#) is licensed [CC BY 4.0](#).

2.2: RNA processing

Learning Objectives

- List three mRNA-processing events that occur after transcription in eukaryotes and identify the purpose of each process.
- Explain how alternative splicing produces multiple distinct mRNAs from a single gene.

mRNA processing in eukaryotes

Transcription is the mechanism by which the information in genes (DNA) is used to produce RNA. The newly-made RNA, also known as the primary transcript (the product of transcription is known as a transcript) or pre-mRNA, is further processed before it is functional. Transcripts that encode proteins will proceed to the ribosome for translation. In bacterial cells, the mRNA can be translated directly as it comes off the DNA template. In eukaryotic cells, RNA synthesis, which occurs in the nucleus, is separated from the protein synthesis machinery, which is in the cytoplasm. In addition, eukaryotic genes have introns, which are non-coding regions that interrupt the gene's coding sequence. The primary RNA copied from genes containing introns will also therefore have regions that interrupt the coding sequence of the gene. These regions must be removed before the mRNA is sent out of the nucleus to be used to direct protein synthesis. The process of removing the introns and rejoining the coding sections or exons, of the mRNA, is called splicing. Proteins that interact with the cap, mRNA, and poly(A) tail allow the mature mRNA to be exported through nuclear pores into the cytoplasm for translation.

What are the processing steps for messenger RNAs?

In eukaryotic cells, pre-mRNAs undergo three main processing steps:

- Capping at the 5' end
- Addition of a poly(A) tail at the 3' end
- Splicing to remove introns

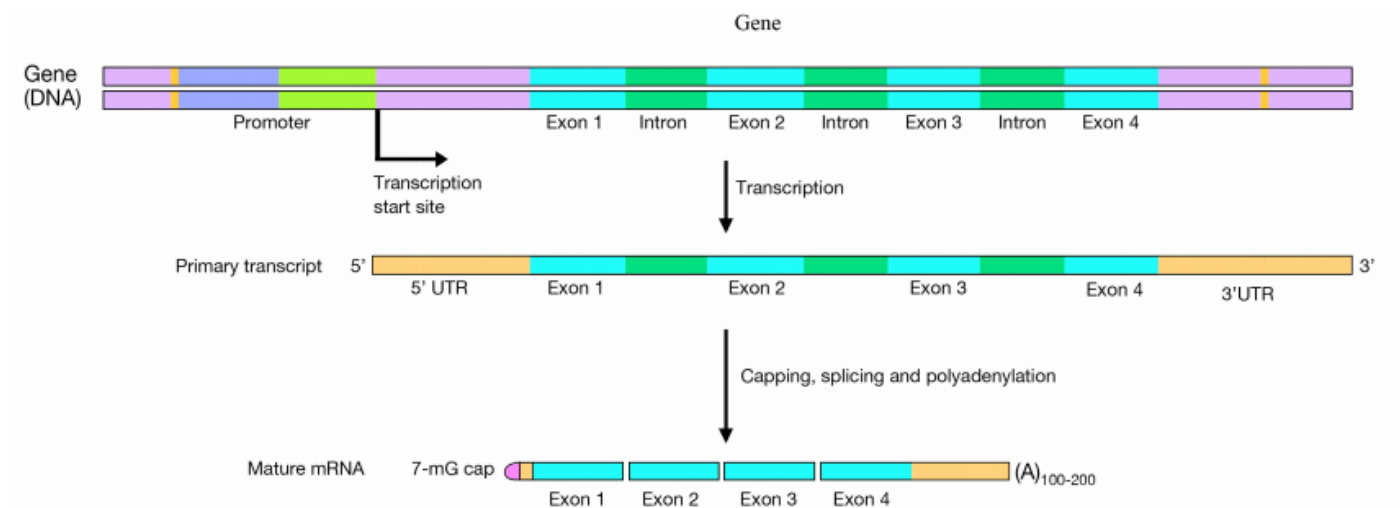


Figure 2.2.1: Summary of steps in processing eukaryotic messenger RNAs

Capping

In the capping step of mRNA processing, a methylated-guanosine (7-methyl-G) is linked to the phosphates at the 5' end of the mRNA. The cap protects the 5' end of the mRNA from degradation by nucleases and also helps to position the mRNA correctly on the ribosomes during protein synthesis.

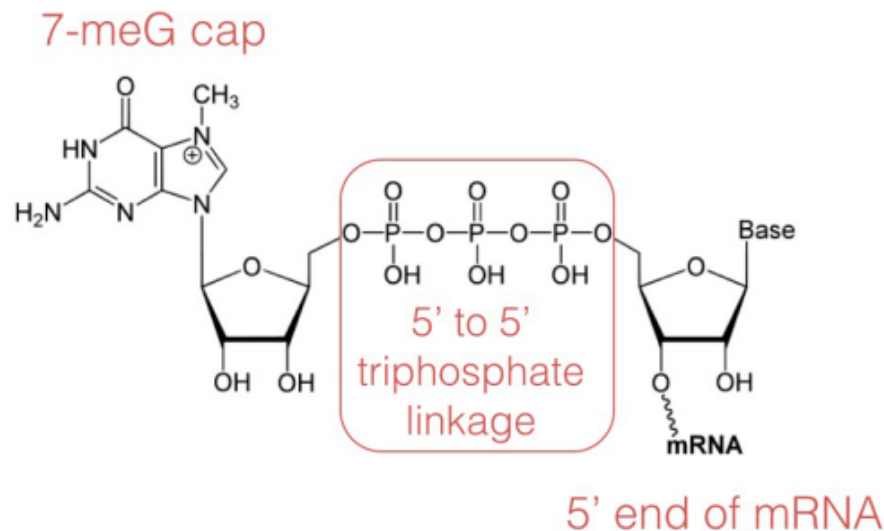


Figure 2.2.2: The mRNA capping structure is a methylated G nucleotide that is linked with triphosphates to the 5' nucleotide instead of via a phosphodiester bond.

Poly(A) tail addition

The 3' end of a eukaryotic mRNA is first trimmed, then an enzyme called Poly(A) Polymerase adds a "tail" of about 200 'A' nucleotides to the 3' end. There is evidence that the poly(A) tail plays a role in efficient translation of the mRNA, as well as in the stability of the mRNA. The cap and the poly(A) tail on an mRNA indicate that the mRNA is complete (i.e., not defective).

Splicing

In many cases, splicing now appears to occur as the transcript is being synthesized. Introns are removed from the pre-mRNA by the activity of a complex called the **spliceosome**. The spliceosome is made up of proteins and small nuclear RNAs (snRNAs) that associate to form protein-RNA enzymes called small nuclear ribonucleoproteins or snRNPs (pronounced SNURPS). The splicing machinery must be able to recognize sequences that are specific to splice junctions (i.e., the end of each exon and the start of the next) in order to correctly cut out the introns and join the exons together to make the mature, spliced mRNA.

Introns vs Exons: Which one is "in" ?

- The word intron comes from the word intervening; these are the intervening sequences that break up the coding sequence. You can remember that introns are "in the way" of the coding sequence.
- The word exon comes from the word expressed; these sequences are present in the mature mRNA.

What signals indicate where an intron starts and ends? The base sequence at the start (5' or left end, also called the splice donor or donor site) of an intron is GU while the sequence at the 3' or right end (also called the splice acceptor or acceptor site) is AG. There is also a third important sequence within the intron, called a branch point, that is important for splicing. These sequences interact with the snRNAs

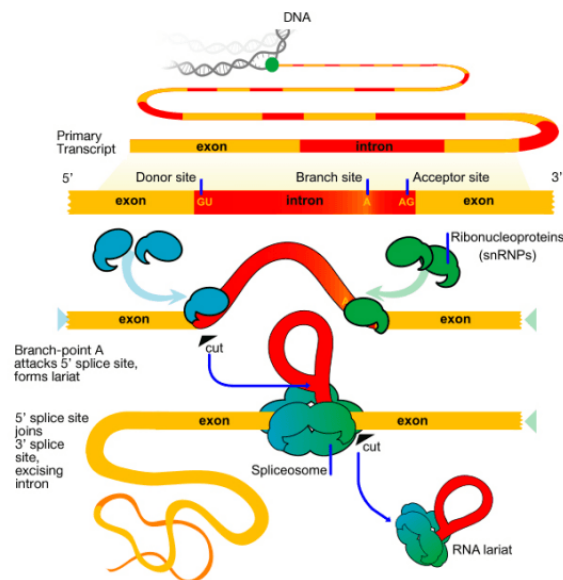


Figure 2.2.3: Summary of splicing. Ribonucleoprotein complexes (snRNPs) bind exon-intron boundaries at splice donor and splice acceptor sites. The branch point A nucleotide attacks the 5' splice site cutting the mRNA at that position and forming a loop. The spliceosome joins the end of one exon to the start of the next exon and the intron is excised.

There are two main steps in splicing:

- In the first step, the pre-mRNA is cut at the 5' splice site (the junction of the 5' exon and the intron). The 5' end of the intron then is joined to the branch point within the intron. This generates the lariat-shaped molecule characteristic of the splicing process
- In the second step, the 3' splice site is cut, and the two exons are joined together, and the intron is released.

Exercise 2.2.1

Draw a diagram of a section of DNA that has a gene with three exons and two introns. Draw the corresponding primary transcript and mature mRNA. Label the promoter, transcription start site, exons, introns, splice donors (SD) and splice acceptors (SA), transcription termination, 5' cap, and polyA tail.

Answer available at the end of this section.

Exercise 2.2.2

Where does eukaryotic mRNA processing occur?

- All three steps occur in the nucleus.
- All three steps occur in the cytoplasm.
- The cap and tail are added in the nucleus, but splicing occurs in the cytoplasm.
- Splicing occurs in the nucleus, but the cap and tail are added in the cytoplasm.

Answer available at the end of this section.

Alternative Splicing

Many pre-mRNAs have a large number of exons that can be spliced together in different combinations to generate different mature mRNAs. This is called **alternative splicing**, and allows the production of many different proteins using relatively few

genes, because a single RNA can, by combining different exons during splicing, create many different protein coding messages. Because of alternative splicing, each gene in our DNA gives rise, on average, to three different proteins.

Alternative splicing may also be used to change or downregulate protein production from a gene, by retaining an intron that includes a premature stop codon, which will end protein synthesis before the final protein can be completed. This role for alternative splicing may be useful for organisms that need to respond to a changing environment, such as temperature or light for plants. Profiling of mRNAs after cold treatment in tea plants revealed that 28.8% of alternatively spliced genes retained an intron (Li et al, 2020).

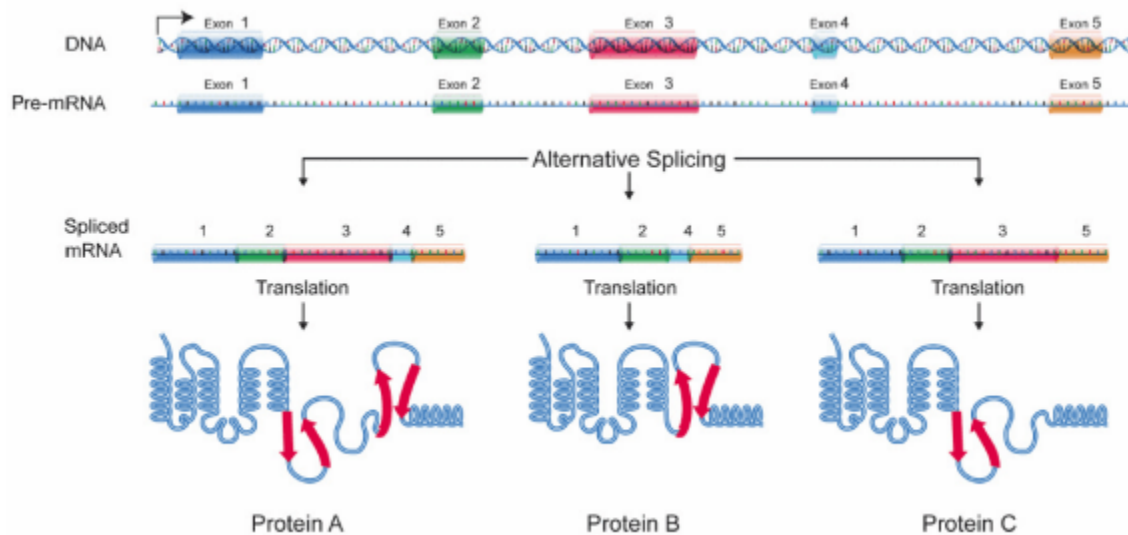


Figure 2.2.4: Splicing and protein diversity

tRNAs and rRNAs are Processed After Transcription

In addition to mRNA processing, non-coding eukaryotic RNAs are processed (trimmed, chemically modified) from large precursor RNAs to mature, functional RNAs. These precursor RNAs (pre-RNAs, or primary transcripts) contain in their sequences the information necessary for their function in the cell. Both prokaryotes and eukaryotes process their ribosomal and transfer RNAs.

- Pre-rRNA (ribosomal RNA) is cleaved and/or trimmed (not spliced because the pieces become separate rRNAs) to make multiple, shorter mature rRNAs.
- Pre-tRNAs (transfer RNA) are trimmed, some bases within the transcript are chemically modified and 3' bases (not encoded by the tRNA gene) are enzymatically added to the 3'-end.

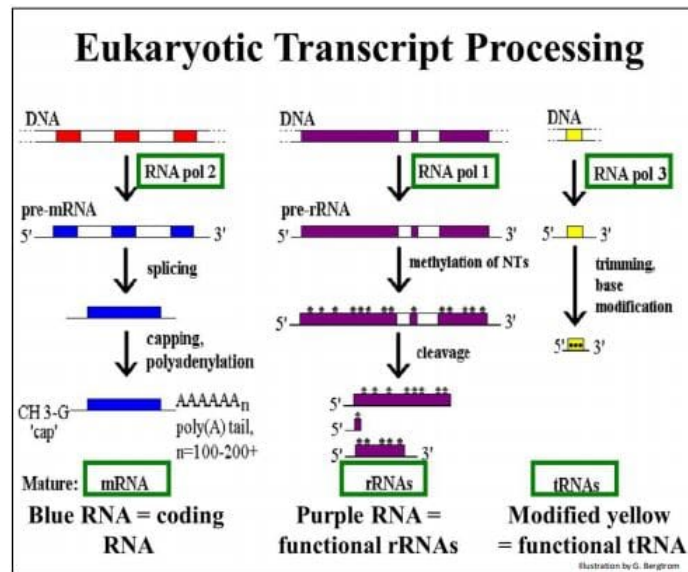


Figure 2.2.1: Left: Many eukaryotic genes are 'split' into coding regions (exons, blue) and non-coding intervening regions (introns, white). Transcription generates a primary mRNA transcript (pre-mRNA) that contains both exons and introns. Primary transcripts are spliced to remove the introns from the exons; exons are then ligated into a continuous mature mRNA. (Copyright Gerald Bergtrom via 10.2: Overview of Transcription)

References

Li, Y., Mi, X., Zhao, S. *et al.* Comprehensive profiling of alternative splicing landscape during cold acclimation in tea plant. *BMC Genomics* **21**, 65 (2020). <https://doi.org/10.1186/s12864-020-6491-6>

Zahler AM. Alternative splicing in *C. elegans*. WormBook. 2005 Sep 26:1-13. doi: 10.1895/wormbook.1.31.1. PMID: 18050427; PMCID: PMC4781185.

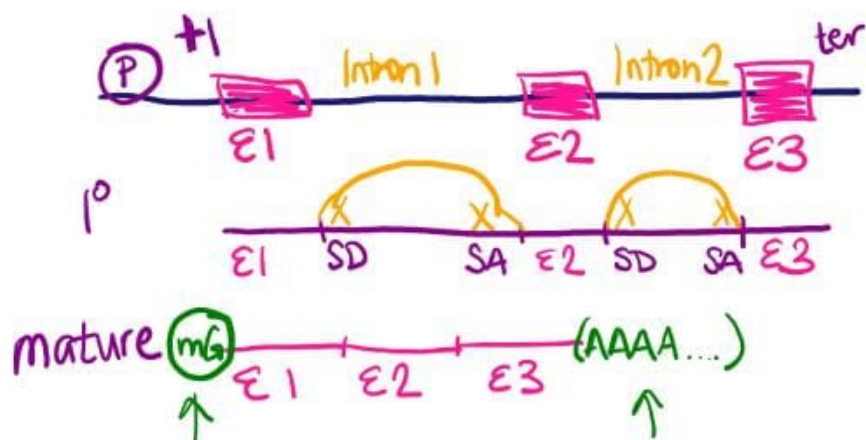
Contributors and Attributions

- Dr. Kevin Ahern and Dr. Indira Rajagopal (Oregon State University)
- Gerald Bergtrom

Answers

Exercise 2.2.1

Your gene diagram should look something like this:



For a walk-through of creating the diagram, you can view this video:

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Exercise 2.2.2

A.

- [2.2: RNA processing](#) has no license indicated.

2.3: Genetic Code and Translation

Learning Objectives

- Identify the key steps of translation and the role of tRNAs, aminoacyl tRNA synthetases, and ribosomal RNAs.
- Use the codon table to determine the sequence of amino acids that will be produced from a DNA or mRNA sequence.
- Use the codon table to determine which amino acid will be carried by a tRNA, given the anticodon sequence.
- Identify untranslated regions and include these regions in gene diagrams.
- Explain how some antibiotics affect ribosomes to kill bacteria.

Overview of translation

Genes contain the instructions a cell needs to make proteins. Making proteins from DNA requires a two-step process:

1. Transcription: the process of copying the gene's DNA into RNA
2. Translation: the process of using RNA to synthesize a sequence of amino acids to form a polypeptide (or protein)

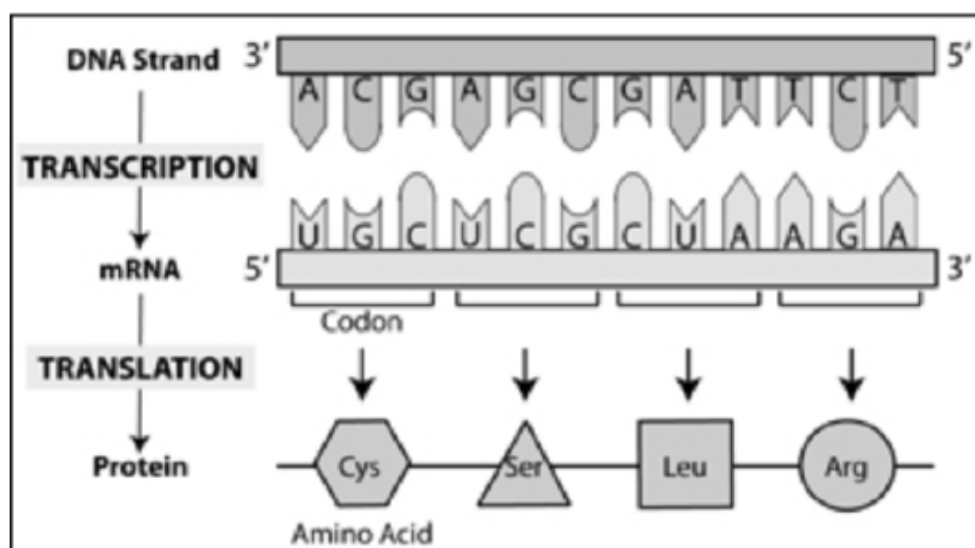


Figure 2.3.1. Summary of transcription and translation showing the relationship of DNA and RNA nucleotides with amino acids (CC BY-NC-SA)

Important Players in Translation

messenger RNA(mRNA): RNA copy of DNA that contains the instructions to make a protein

transfer RNA (tRNA): RNA molecule responsible for delivering amino acids to the ribosome

amino acids: The basic building block of a protein. There are 20 different amino acids, each has an amino group (NH_2) and a carboxyl group (COOH), but differs in its functional group (also known as side chain or R group). An infographic showing all the amino acids can be found at this link (<https://www.compoundchem.com/wp-content/uploads/2014/09/20-Common-Amino-Acids-Physiological-Structures.pdf>).

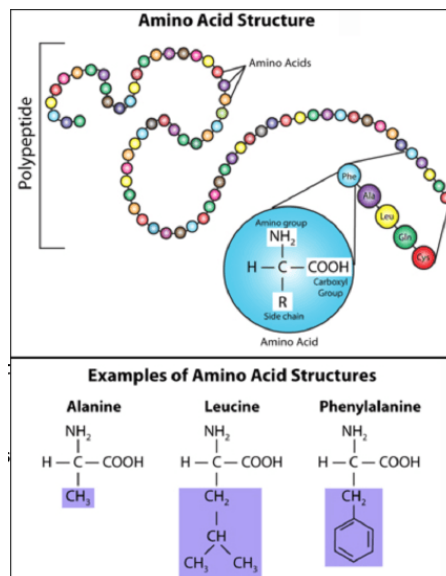


Figure 2.3.2. (CC BY-NC-SA)

protein: A chain of amino acids, also known as a polypeptide; some proteins require multiple polypeptides to function.

ribosome: The organelle on which mRNA is translated into protein. Consists of a large (60S) and small subunit (40S) which are made up of ribosomal RNA (rRNA) and proteins.

The Genetic Code: Translating RNA into Protein

How does the nucleotide sequence of RNA specify the specific order of amino acids in a protein? The answer lies in what is known as the genetic code.

- If each RNA base codes for just 1 amino acid, RNA could code for only 4 amino acids (not enough to include all 20 amino acids).
- If two RNA bases code for 1 amino acid, RNA could code for 16 amino acids (still not enough to include all 20 amino acids).
- If three RNA bases code for 1 amino acid, RNA could code for 64 amino acids (more than enough to include all 20 amino acids). Thus, the genetic code is a triplet code in which three nucleotides in RNA specify one amino acid in protein.

In the 1960s, the mRNA code for each amino acid was determined experimentally; today, tables are used to find the amino acid encoded by each codon. Remember that each codon must be read 5' to 3'.

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G
						Third letter

Figure 2.3.3: A codon table shows the amino acid specified by each codon. The codon AUG is the start signal for translation which places the amino acid, methionine (Met) at the beginning of each protein. Three codons, UAA, UAG, and UGA, called stop codons, act as signals to terminate translation. (Copyright: General Biology at [OpenStax](#))

The role of transfer RNA

Consider nucleic acids and amino acids as different languages. DNA and RNA speak the language of nucleic acids (A, U/T, G and C); whereas proteins speak the language of amino acids. The tRNA (transfer RNA) is the "interpreter" in the cell because it is the molecule that speaks both languages!

Sets of three nucleotides that code for a specific amino acid are known as codons. These codons are recognized by, and base pair with, the tRNA molecule with the complementary sequence called anticodon. tRNA molecules act as translators because they are able to read the nucleic acid words (as mRNA codons) and deliver the corresponding words (as amino acids). There is at least one tRNA for each of the 20 amino acids (some amino acids bind to 2 or 3 different tRNAs, so cells may contain as many as 32 different tRNAs). The secondary structure of the tRNA relies on complementary sequences within a single tRNA to fold into a function structure with the anticodon loop at the fold opposite the amino acid on the 3' end. Enzymes called aminoacyl-tRNA synthetases catalyze the attachment of the correct amino acid to each tRNA, this process is known as tRNA charging.

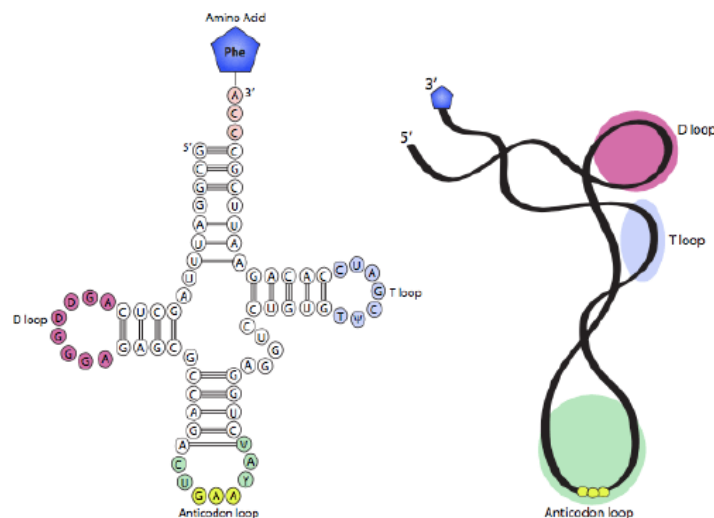


Figure 2.3.4: Two representations of tRNA structure. At left, the classic clover-leaf, splayed out for simplicity. At right, a more accurate representation of the tRNA in pseudo-3D. (E.V. Wong CC-BY-NC-SA [10.5: tRNA are Rather Odd Ducks](#))

Exercise 2.3.1

A protein is being translated from an mRNA with the sequence 5' augcaucauaugcuucugcuauguuaagc.. 3'. Starting with aug, what are the first five amino acids of the peptide?

Answer available at the end of this section.

Exercise 2.3.2

A DNA sequence is 5' ...ATGCCGTATAGG... 3'.

1. If this strand is the coding strand, determine the RNA and possible amino acid sequence. (Assume you begin translation at the 5' nucleotide.)
2. If this strand is the non-coding (template) strand, determine the RNA and possible amino acid sequence. (Assume you begin translation at the 5' nucleotide.)

Answer available at the end of this section.

Exercise 2.3.3

If a tRNA has the anticodon sequence 5' UAC 3', what amino acid will it be carrying?

Answer available at the end of this section.

The Steps of Translation

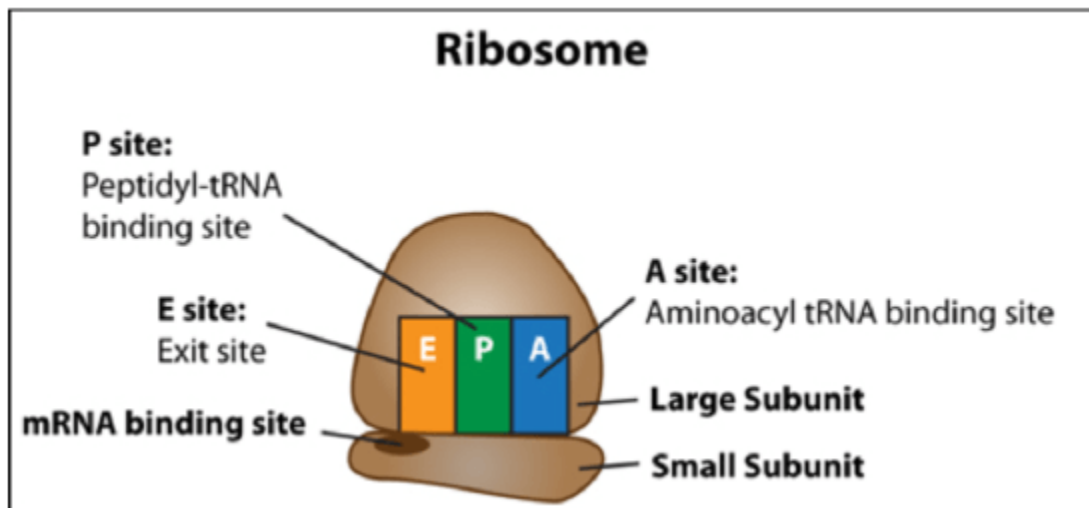


Figure 2.3.5: Summary of ribosome structure and positions of E, P, and A sites. (CC BY-NC-SA)

Translation Initiation: The small subunit binds to a site upstream (on the 5' side) of the start of the mRNA. It proceeds to scan the mRNA in the 5'-to-3' direction until it encounters the START codon (AUG). Nearby sequences help position the small subunit. In prokaryotes, this sequence, called the Shine-Dalgarno sequence, is upstream of the AUG. In eukaryotes, a consensus sequence (Kozak sequence) surrounds the AUG.

When the small subunit and mRNA are correctly positioned with a tRNA, the large subunit attaches and the initiator tRNA, which carries methionine (Met), binds to the P site on the ribosome.

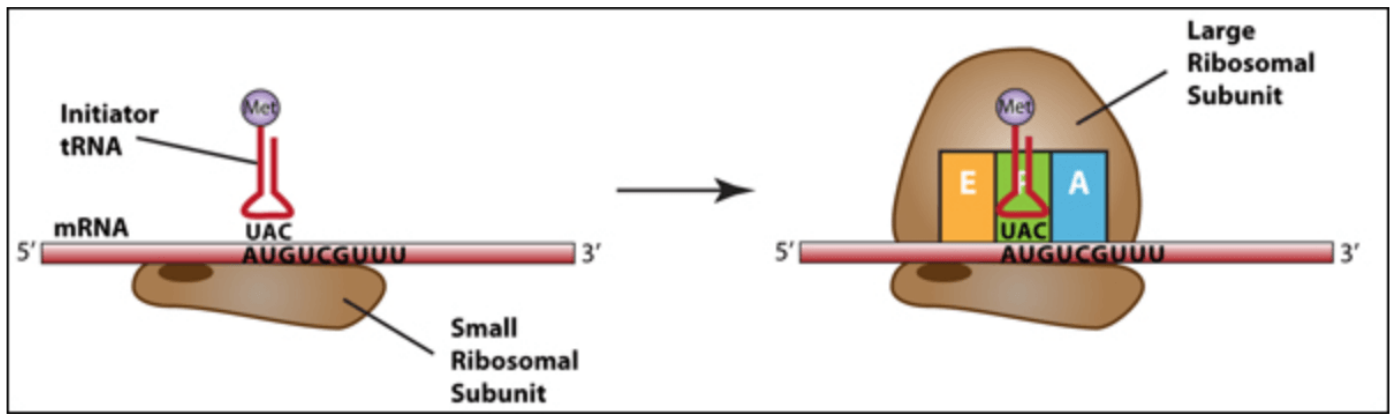


Figure 2.3.6. Translation Initiation: The assembly of the small ribosomal subunit, mRNA and initiator tRNA is followed by the addition of the large ribosomal subunit. The initiator tRNA has sequence UAC that is complementary to AUG in the mRNA. (CC BY-NC-SA)

Elongation: A tRNA bound to its amino acid (known as an aminoacyl-tRNA) that is able to base pair with the next codon on the mRNA arrives at the A site. The ribosome catalyzes the formation of a peptide bond between the preceding amino acid (Met at the start of translation) in the P site and the amino acid held by the tRNA in the A site. The initiator tRNA moves to the E site and the ribosome translocates (moves down) one codon downstream. This shifts the more most recent tRNA from the A site to the P site, opening up the A site for the arrival of a new aminoacyl-tRNA. Polypeptide synthesis repeats, the tRNA residing in the E site is released from the complex, the tRNAs in the P site and A site shift over and the next amino acid is added to the growing polypeptide chain. This cycle repeats until a stop codon is reached.

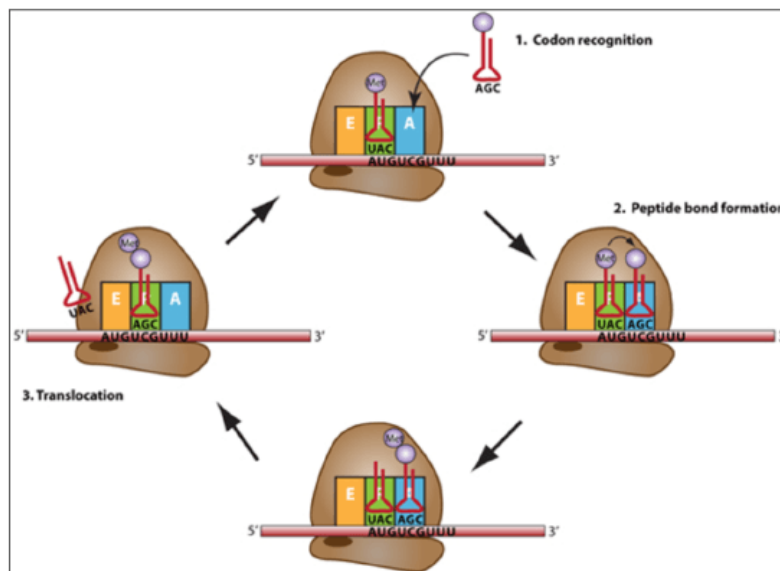


Figure 2.3.7: The elongation cycle of translation begins with codon recognition in the A site by a tRNA with a complementary anticodon, then peptide bond formation and transfer of the amino acid to the A site tRNA, then translocation of the ribosome toward the 3 end of the mRNA for the next codon to be in the A site. (CC BY-NC-SA)

Remembering what happens in the E, P, and A sites

- P = peptidyl because the P site holds the tRNA with the growing peptide
- A = aminoacyl because the A site holds the incoming amino acid that will accept the transfer of the peptide from the P site

- E = exit site because the tRNA from the P site will exit the ribosome through this site after peptide bond formation and translocation

Termination: Translation ends when the ribosome reaches a STOP codon (UAA, UAG or UGA). There are no tRNA molecules with anticodons complementary to stop codons, instead protein release factors (RF) recognize these codons when they arrive at the A site. Binding of a release factor causes the polypeptide (protein) to be released from the ribosome. The ribosome subunits dissociate (split) from each other and can be reassembled later for another round of protein synthesis.

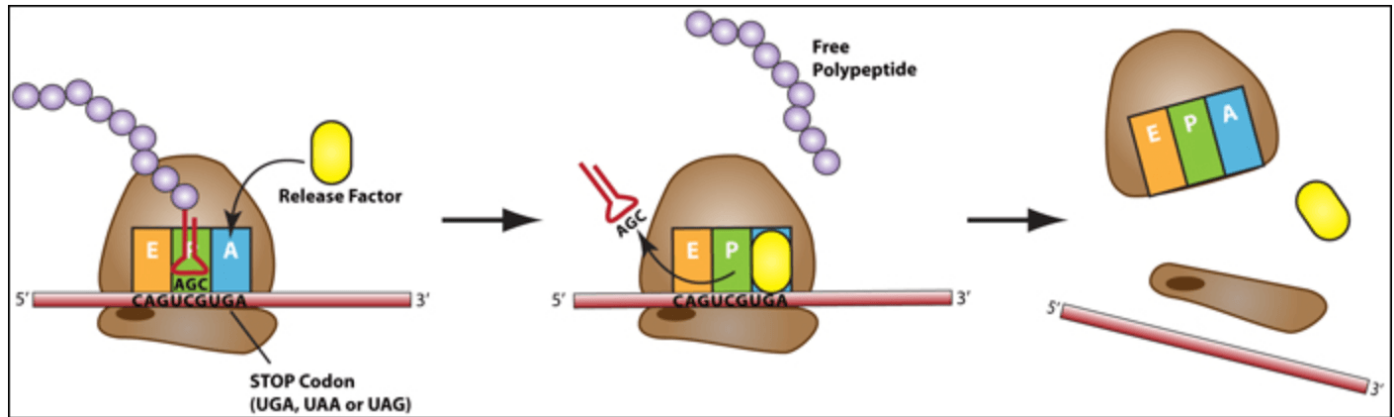


Figure 2.3.8: During termination release factor proteins enter the A site and the small and large ribosomal subunits dissociate from each other and the mRNA. The newly made polypeptide is released. (CC BY-NC-SA)

Translation tutorial modified from:



Translation Tutorial by [Dr. Katherine Harris](#) is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#).

Funded by the U.S. Department of Education, College Cost Reduction and Access (CCRAA) grant award # P031C080096.

Untranslated regions (UTRs)

In both prokaryotes and eukaryotes, the start codon for translation (AUG) is not at the +1 site where transcription begins. Similarly, the stop codon for translation is not at the 3' end of the mRNA. This genetic architecture leads to two regions of the mRNA that are called the **5' and 3' untranslated regions (UTRs)**. To understand what a UTR is, review these key steps of translation initiation and termination.

- For initiation, the small ribosomal subunit moves along the mRNA from 5' to 3' until it is positioned at the start codon. Then the large subunit load onto the complex with the initiator tRNA carrying methionine. Then the ribosome translates the RNA in the 5' to 3' direction, therefore all the nucleotides prior to the start codon were skipped, or untranslated.
- For termination, when the ribosome encounters a stop codon, there is not a tRNA with a complementary codon. Instead release factor proteins bind in that site and disrupt the complex, causing the mRNA and ribosome to disassociate. Therefore, any nucleotides after the stop codon cannot be translated, because the ribosome no longer moves along this part of the mRNA.

Example 2.3.1

Consider a relatively small gene, the SRY gene on the human Y chromosome. The SRY mRNA is 828 nucleotides. If each amino acid is coded by 3 nucleotides (nt) and the entire mRNA encoded amino acids, the translated protein would be 276 amino acids.

However, the SRY protein is 204 amino acids. How many nucleotides are in the untranslated regions?

Solution

If 1 amino acid is encoded by 3 nucleotides, then 204 amino acids x 3 nucleotides = 612 nucleotides to encode a 204 amino acid protein. Adding 3 nucleotides for the stop codon adds 3 more nucleotides (612 + 3) = 615 nucleotides of coding sequence.

The SRY mRNA is 828 nucleotides, so the remaining (828 - 615) 213 nucleotides must be in untranslated regions. The nucleotides upstream (5' of) the AUG start codon are the 5' UTR and the nucleotides downstream of the stop codon (in this gene, UAG) are the 3' UTR.

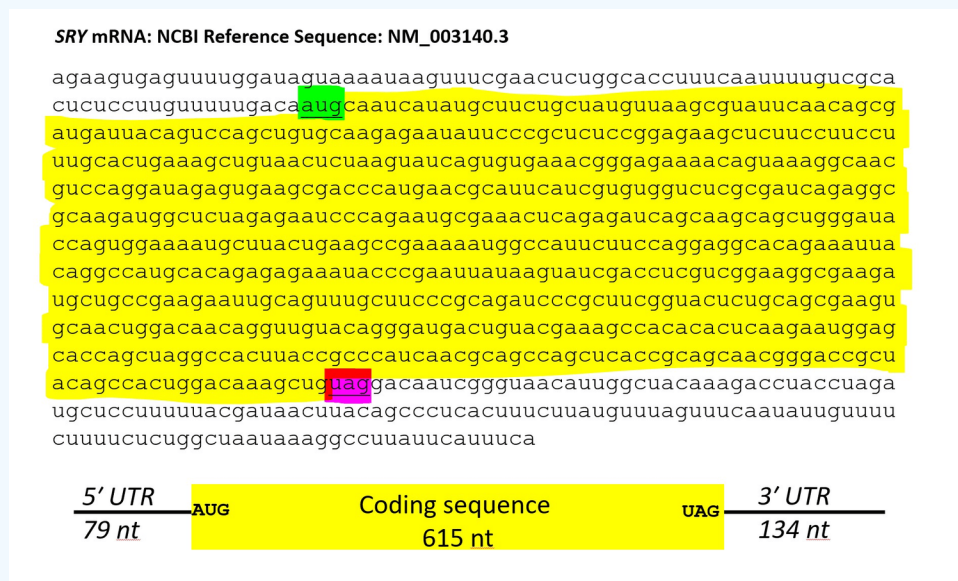


Figure 2.3.11: Homo sapiens SRY mRNA sequence and gene diagram (https://www.ncbi.nlm.nih.gov/nucore/NM_003140.3) with start codon AUG highlighted in green, stop codon UAG highlighted in pink, and remaining coding sequence highlighted in yellow. Gene diagram is labeled with 79 nucleotide of 5' UTR, 615 nucleotides of coding sequence, and 134 nucleotides of 3' UTR. (Image CC-BY S. Leacock)

Why have untranslated regions?

Untranslated regions often function in RNA stability, translation initiation, or other aspects of post-transcriptional regulation; however the complete function of most untranslated regions is not entirely known.

The importance of ribosomal RNA

The ribosome itself is made of both ribosomal proteins and rRNA (ribosomal RNA). In a Perspective from 2000 titled "The Ribosome is a Ribozyme," (<https://science.sciencemag.org/content/289/5481/878>) Thomas Cech describes how the crystal structure of the ribosome advanced our understanding of the roles of the rRNA and protein components of this critical complex. The crystal structure of the large ribosomal subunit from the archaean *Haloarcula marismortui* at high resolution was published in 2000 (<https://science.sciencemag.org/content/289/5481/905> Ban et al.). Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath later shared the 2009 Nobel Prize in Chemistry for their contributions to our understanding of this structure. This structure revealed that the active site, where the peptide bond forms between the amino acids in the P and A sites, is largely not

proteins, which most cellular catalysts are, but instead ribosomal RNA. A ribozyme is the term for an RNA that can catalyze an enzymatic reaction. Therefore, the ribosome is a ribozyme.

Structure of the 50S ribosomal subunit assembly (intermediate state 1) from *E. coli* K-12 (<https://www.rcsb.org/structure/6GC7>) with ribosomal RNA in orange and proteins in other colors.

Antibiotics and the ribosome

In general, protein synthesis inhibitors work at different stages of prokaryotic mRNA translation into proteins like initiation, elongation (including aminoacyl tRNA entry, proofreading, peptidyl transfer, and ribosomal translocation), and termination. The following is a list of common antibacterial drugs and the stages which they target.

- Linezolid acts at the initiation stage, probably by preventing the formation of the initiation complex, although the mechanism is not fully understood.
- Tetracyclines and Tigecycline (a glycylcycline related to tetracyclines) block the A site on the ribosome, preventing the binding of aminoacyl tRNAs.
- Aminoglycosides, among other potential mechanisms of action, interfere with the proofreading process, causing an increased rate of error in synthesis with premature termination.
- Chloramphenicol blocks the peptidyl transfer step of elongation on the 50S ribosomal subunit in both bacteria and mitochondria.
- Macrolides, clindamycin, and aminoglycosides have evidence of inhibition of ribosomal translocation.
- Streptogramins also cause premature release of the peptide chain.

By targeting different stages of the mRNA translation, antimicrobial drugs can be changed if resistance develops to one or many of the drugs.

Contributor:

- General Microbiology Boundless from: [https://bio.libretexts.org/Courses/Northwest_University/MKBN211%3A_Introductory_Microbiology_\(Bezuidenhout\)/07%3A_Antimicrobial_Drugs/7.02%3A_Functions_of_Antimicrobial_Drugs/7.2.04%3A_Inhibiting_Protein_Synthesis](https://bio.libretexts.org/Courses/Northwest_University/MKBN211%3A_Introductory_Microbiology_(Bezuidenhout)/07%3A_Antimicrobial_Drugs/7.02%3A_Functions_of_Antimicrobial_Drugs/7.2.04%3A_Inhibiting_Protein_Synthesis)

Thinking about antibiotics and specificity

Antibiotics are used to treat bacterial infections. Consider these questions:

- Why are antibiotics not prescribed for viral infections?
- If humans have ribosomes as well, why would the antibiotics not inhibit the human ribosomes and prevent our own protein production?
- What does antibiotic resistance mean and how does it arise?
- Does the same antibiotic treat all types of bacterial infections? Why or why not?
- How does studying the structure of the ribosome help us to understand the effectiveness of antibiotics?

Answers

Exercise 2.3.1

Met - Gln - Ser - Tyr - Ala

Exercise 2.3.2

1. The RNA has the same sequence as the coding strand, except for U instead of T. The mRNA will be 5' ...AUGCCGUAUAGG... 3' and the protein sequence will be Met - Pro - Tyr - Arg.
2. The RNA has a complementary sequence to the template strand. (Remember that transcription would occur from right to left off this template to make RNA from 5' to 3'.) The mRNA will be 5' ...CCUAUACGGCAU... 3' and the protein sequence will be Pro - Ile - Arg - His.

Exercise 2.3.3

Valine (val)

To see this how this problem is solved, watch this video:

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



This page titled [2.3: Genetic Code and Translation](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [2.3: Genetic Code and Translation](#) by [Stefanie West Leacock](#) has no license indicated.

2.4: Chapter 2 Review Questions

2.1: Overview of Transcription

Query 2.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.5

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.6

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.7

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



2.2: RNA processing

Query 2.4.2.8

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.9

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.10

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.11

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



2.3: Genetic Code and Translation

Query 2.4.2.12

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.13

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.14

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.15

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 2.4.2.16

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Try a crossword puzzle (covers the whole chapter)

Query 2.4.2.17

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



3: MUTATION

Mutations are any change in DNA nucleotide sequence. Because DNA stores information in the sequence of bases, a change in a base could (but does not always) have an effect on gene products. We already learned that genomes have variation; mutations are the source of that variation. In essence, without mutations there would be no diversity - within a species or between species!

[3.1: Origins of Mutations](#)

[3.2: Consequences of mutations](#)

[3.3: Example - Mutations and Cystic Fibrosis](#)

[3.4: Ames Test](#)

[3.5: DNA Repair](#)

[3.6: Chapter 3 Review Questions](#)

This page titled [3: Mutation](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [3: Mutation](#) by [Stefanie West Leacock](#) has no license indicated.

3.1: Origins of Mutations

Learning Objectives

- Identify some causes and types of mutations.
- Define mutagen and predict the effects of mutagens on DNA sequences.

Introduction to mutations

Mutations have a bad reputation, BUT they are ESSENTIAL to life! Watch the video for an introduction to why understanding mutations is so important for us.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 3.1.1: Introduction to mutations. (CC-BY Leacock <https://www.youtube.com/watch?v=-BVFRB2hsCI&t=3s>)

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 DNA polymerase proofreading activity 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 descendants.

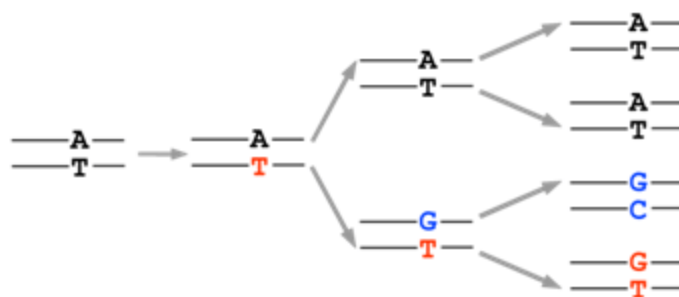


Figure 3.1.1: Mismatching of bases (e.g. G with T) can occur due to tautomerism, alkylating agents, or other effects. As a result, in this example the AT base pair in the original DNA strand will become permanently substituted by a GC based pair in some progeny. The mismatched GT pair will likely be repaired or eliminated before further rounds of replication. (Original-Deyholos-CC:AN)

The most common insult to the DNA of living organisms is **depurination**, in which the bond between an adenine or guanine and the deoxyribose is hydrolyzed. Depyrimidination of cytosine and thymine residues can also occur, but do so at a much slower rate than depurination. Despite the high rate of loss of these bases, they are generally identified and remediated by base excision repair (BER). Therefore, it is rare for depurination or depyrimidination to lead to mutation.

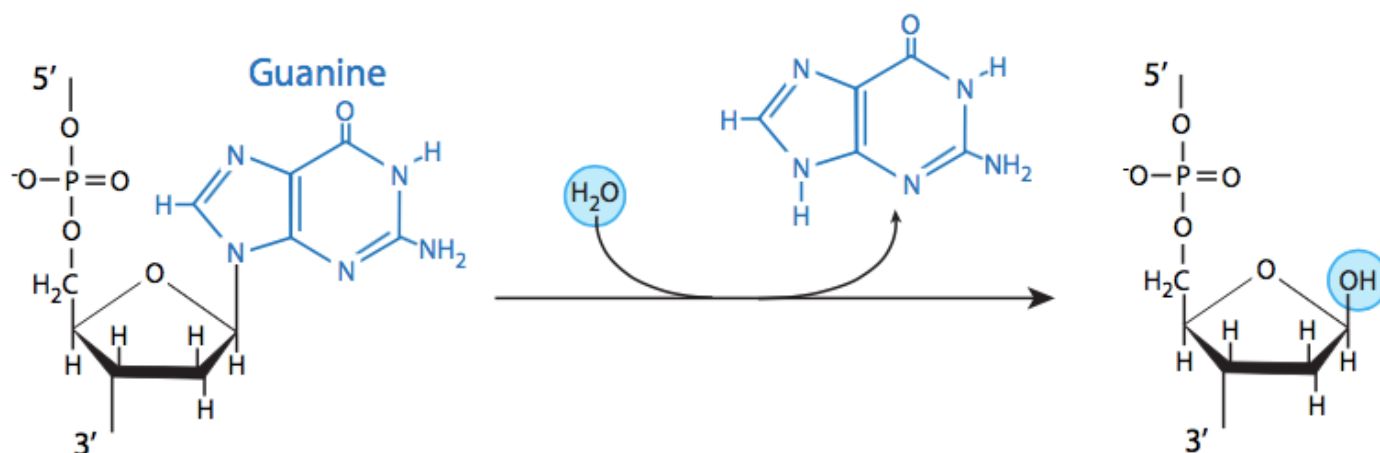


Figure 3.1.2: Depurination of guanines (or adenines) in which the base is cleaved from the sugar is a common DNA lesion.

Three of the four DNA bases, adenine, guanine, and cytosine, contain amine ($-NH_2$) groups that can be lost in a variety of pH and temperature-dependent reactions that convert the bases to hypoxanthine, xanthine, and uracil, respectively. This **deamination** can sometimes lead to permanent mutations during replication. Hypoxanthine base pairs with C (cytosine); xanthine base pairs with cytosine; uracil base pairs with adenine. If hypoxanthine is present when DNA polymerase is replicating DNA, a C would be incorporated instead of T; similarly if uracil is present an A would be incorporated instead of G.

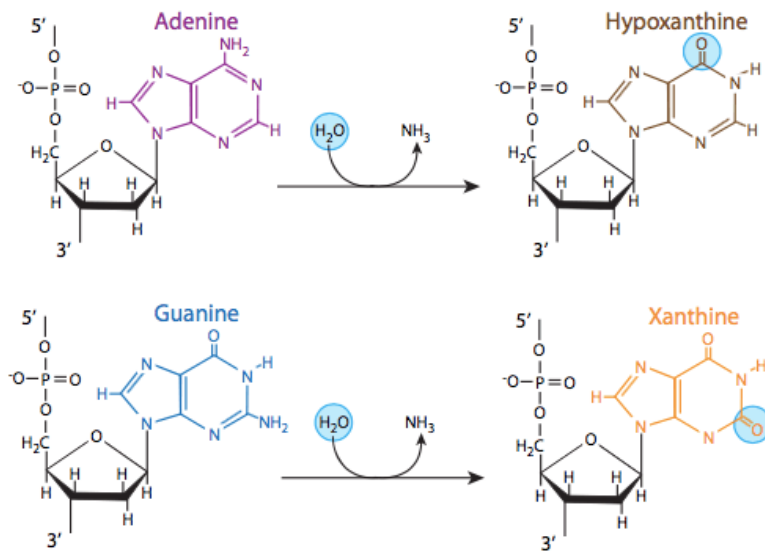


Figure 3.1.3: Deamination converts adenine to hypoxanthine and guanine to xanthine.

Query 3.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Another deamination, of the modified base methylcytosine, can also lead to a mutation upon replication. Some cytosines may be methylated as part of a regulatory process to inactivate certain genes in eukaryotes, or in prokaryotes as protection against restriction endonucleases. When the methylated cytosine is deaminated, it produces a thymine, which changes the complementary nucleotide (upon replication) from a guanine to an adenine. Deamination of cytosines occurs at nearly the same rate as depurination, but deamination of other bases are not as pervasive: deamination of adenines, for example, is 50 times less likely than deamination of cytosine.

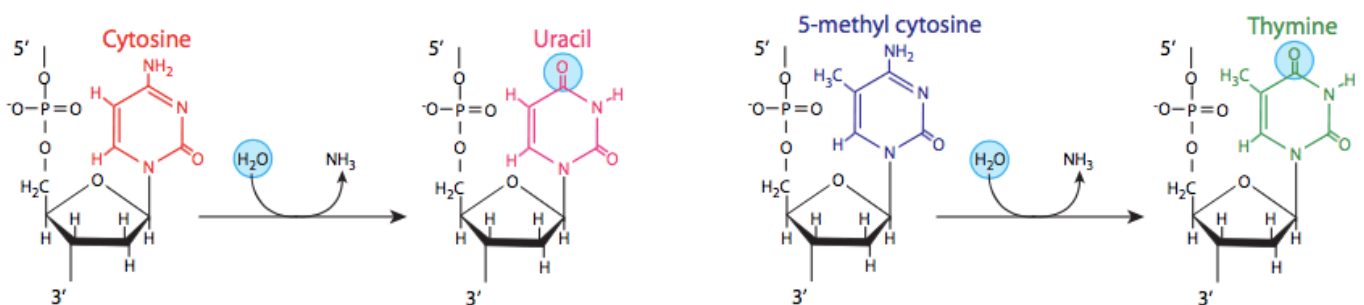


Figure 3.1.4: Deamination of cytosine, and methylcytosine can lead to mutations upon replication if unrepaired.

Query 3.1.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



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 3.1.5). 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.

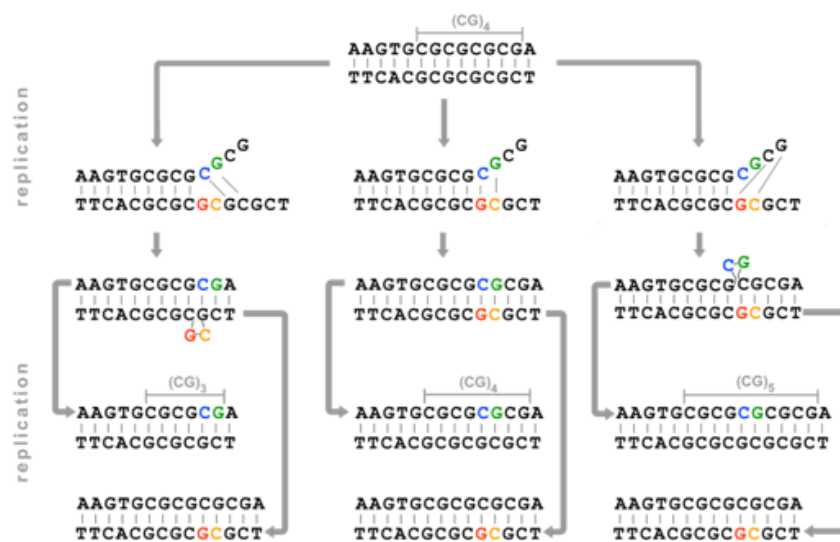


Figure 3.1.5 Strand-slippage can occur occasionally during replication, especially in regions with short, repeated sequences. This can lead to either deletion (left) or insertion (right) of sequences compared to the products of normal replication (center), depending on whether the template strand or daughter strand “loops-out” during replication. (Original-Deyholos-CC:AN)

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 tandem repeats (STRs)** or short sequence repeats (SSRs) are tend to be highly polymorphic, and are therefore particularly useful in genetics. They are also 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:

- **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.

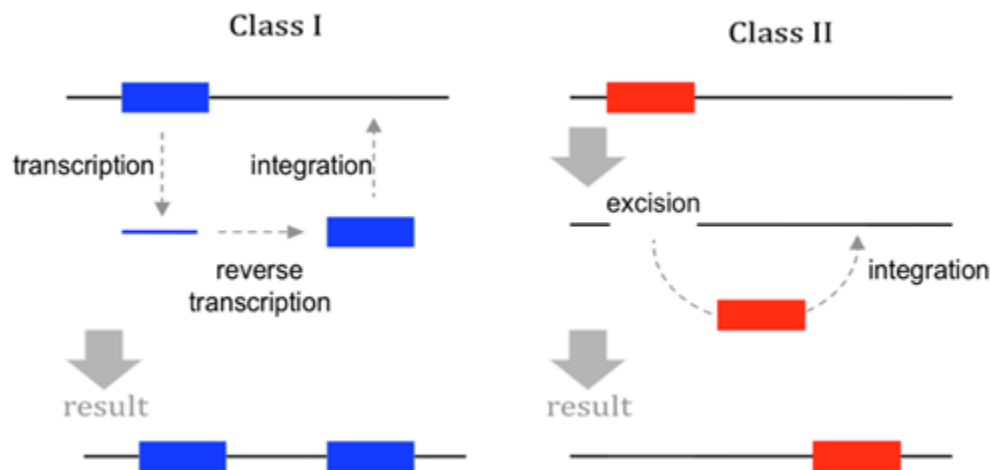


Figure 3.1.6 Diagrams of the two main types of transposable elements. (TEs) Class I elements transpose via an ssRNA intermediate, which is reverse transcribed to dsDNA prior to insertion of this copy in a new site in the genome. Class II elements do not involve an RNA intermediate; most Class II elements 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 (long intersperse nuclear elements)** and **SINEs (short interspersed nuclear elements)**. 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.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 3.1.2: Jumping Gene Caught in the Act (dnalc.org via <https://www.youtube.com/watch?v=9MPiRx3SPMM>)

Are my genes always jumping?!

If the human genome is 45% transposable elements and these elements can move around, then how does our genome stay relatively constant?

Some factors that help keep the genome (mostly) intact include:

- Many of the TE elements have acquired mutations that render them unable to transpose.
- Regulatory RNAs called piRNAs (piwi-related RNAs) appear to suppress transposition.
 - In which tissue(s) would you expect piRNAs to be most active?

Mutations of Chemical Origins

Many chemical compounds, whether natural or synthetic, can react with DNA to cause mutations. In some of these reactions the chemical structure of particular bases may change, so that they are misread during replication. In others the chemical mutagens distort the double helix causing it to be replicated inaccurately, while still others may cause breaks in chromosomes that lead to deletions and other types of aberrations. The following are examples of two classes of chemical mutagens that are important in genetics and medicine: **alkylating agents**, and **intercalating agents**.

What is an alkyl group?

An alkyl group is formed by removing one hydrogen from an alkane chain (C_nH_{2n+2}) and is described by the formula C_nH_{2n+1} . The removal of this hydrogen results in a stem change from **-ane** to **-yl**. Take a look at the following examples.



Source: https://chem.libretexts.org/Courses/can/org/03%3A_Organic_Compounds%3A_Alkanes_and_Their_Stereochemistry/3.4%3A_Alkyl_Groups

Ethane methyl sulfonate (**EMS**) is an example of an alkylating agent that is commonly used by geneticists to induce mutations in a wide range of both prokaryotes and eukaryotes. The organism is fed or otherwise exposed to a solution of EMS. It reacts with some of the G bases it encounters in a process called alkylation, where the addition of an alkyl group to G changes the base pairing properties so that the next time the alkylated DNA strand is replicated, a T instead of a C will be inserted opposite to the alkylated-G in the daughter strand. The new strand therefore bears a mutation (base substitution), which will be inherited in all the strands that are subsequently replicated from it.

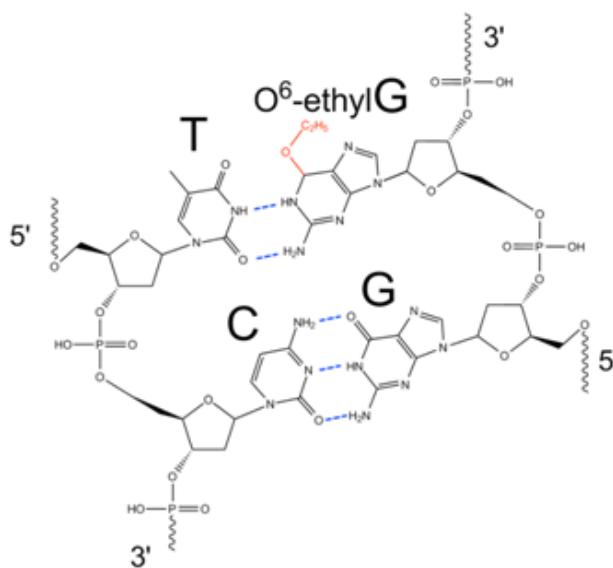


Figure 3.1.7: Alkylation of G (shown in red) allows G to bond with T, rather than with C. (Original-Deyholos-CC:AN)

Intercalating agents are another type of chemical mutagen. They tend to be flat, planar molecules like **benzo[a]pyrene**, a

component of wood and tobacco smoke, and induce mutations by inserting between the stacked bases at the center of the DNA helix. This intercalation distorts the shape of the DNA helix, which can cause the wrong bases to be added to a growing DNA strand during DNA synthesis.

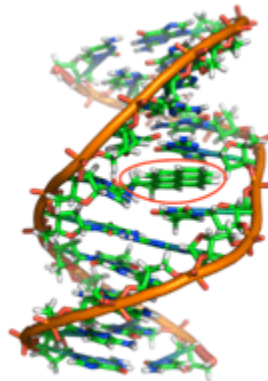


Figure 3.1.8: Benzo[a]pyrene (circled in red) is an example of an intercalating agent.(Wikipedia-Zephyris-GFDL)

There are a large number of chemicals that act as intercalating agents, can mutate DNA, and are **carcinogenic** (can cause cancer). Many of these are also used to treat cancer, as they preferentially kill actively dividing cells. Another important intercalating agent is **ethidium bromide**, the fluorescent dye that stains DNA in laboratory assays. For this reason, molecular biologists are trained to handle this chemical carefully.

Exercise 3.1.1

The chemical EMS adds an ethyl group to G, which then pairs with T instead of C. What would be the effect of exposing the DNA sequence ...CATGTCA... to EMS?

Answer available at the end of this section.

Mutations of physical origin

Anything that damages DNA by transferring energy to it can be considered a physical mutagen. Usually this involves radioactive particles, x-rays, or UV light. The smaller, fast moving particles may cause base substitutions or delete a single bases, while larger, slightly slower particles may induce larger deletions by breaking the double stranded helix of the chromosome. Physical mutagens can also create unusual structures in DNA, such as the thymine dimers formed by UV light (Figure 3.1.9). Thymine dimers disrupt normal base-pairing in the double helix, and may block replication altogether if not repaired by the cell's DNA repair enzymes.

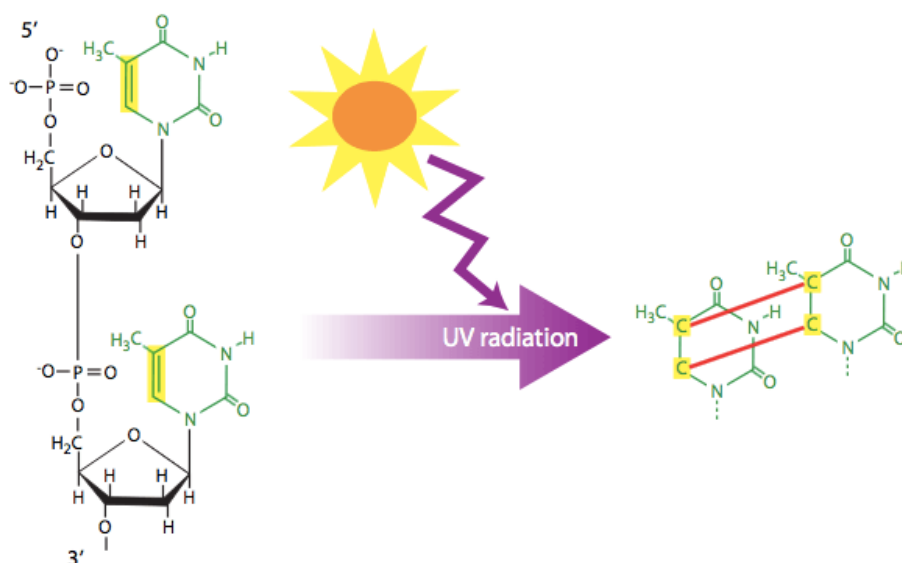


Figure 3.1.9: Ultraviolet radiation can be absorbed by some DNA and commonly causes pyrimidine cyclobutyl dimers connecting adjacent nucleotide bases. (CC Wong via [7.5: DNA Lesions](#))

Summary Table of Mutagenic Agents

Mutagenic Agents	Mode of Action	Effect on DNA	Resulting Type of Mutation
Nucleoside analogs			
2-aminopurine	Is inserted in place of A but base pairs with C	Converts AT to GC base pair	Point
5-bromouracil	Is inserted in place of T but base pairs with G	Converts AT to GC base pair	Point
Nucleotide-modifying agent			
Nitrous oxide	Deaminates C to U	Converts GC to AT base pair	Point
Ethane methyl sulfonate (EMS)	Alkylates G (which pairs with T)		
Intercalating agents			
Acridine orange, ethidium bromide, polycyclic aromatic hydrocarbons	Distorts double helix, creates unusual spacing between nucleotides	Introduces small deletions and insertions	Frameshift
Ionizing radiation			
X-rays, γ-rays	Forms hydroxyl radicals	Causes single- and double-strand DNA breaks	Repair mechanisms may introduce mutations
X-rays, γ-rays	Modifies bases (e.g., deaminating C to U)	Converts GC to AT base pair	Point
Nonionizing radiation			

Mutagenic Agents	Mode of Action	Effect on DNA	Resulting Type of Mutation
Ultraviolet	Forms pyrimidine (usually thymine) dimers	Causes DNA replication errors	Frameshift or point

Query 3.1.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Contributors and Attributions

- [Dr. Todd Nickle and Isabelle Barrette-Ng](#) (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.
- [E. V. Wong](#) Axolotl Academica Publishing (Biology) at [Axolotl Academica Publishing](#)
- Mutations. (2021, January 3). Retrieved August 6, 2021, from <https://bio.libretexts.org/@go/page/5184> " Mutations" by OpenStax, LibreTexts is licensed under CC BY ([11.5: Mutations](#)).

Answers

Exercise 3.1.1

When DNA is exposed to a mutagen it is unlikely that all the Gs in a given sequence will be affected. We will assume that only the G in the above sequence is affected (not the two Gs that will be on the complementary strand). I will write the modified sequence as ...CATG^mTCA...

- When the methylated strand is used as a template for DNA replication, the newly synthesized strand will be ...GTATAGT...
- If the cell undergoes another round of cell division, using the new strand as a template will create the strand ...CATATCA... and the former GC base pair is now an AT base pair. The cell has no way of knowing that this base pair is incorrect or whether or not it will affect the cell.

Note that in this course, you do not need to memorize the action of specific mutagens. Information about what the mutagen does will be provided.

This page titled [3.1: Origins of Mutations](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [3.1: Origins of Mutations](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

3.2: Consequences of mutations

Learning Objectives

- Classify mutations in coding regions as silent, nonsense, missense, or frameshift.

The effect of a mutation will depend upon the function of the DNA sequence. Mutations in non-protein coding regions could occur in regulatory regions that affect genes or have no effect on a cell or organism.

When mutations occur in coding sequences, we can predict the effect on the primary structure of the protein using the codon table.

- Silent mutations: do not change the encoded amino acid
- Nonsense mutations: change a codon to a STOP codon
- Missense mutations: change a codon to a codon for a DIFFERENT amino acid
- Frameshift mutations: add or remove bases to change all downstream codons

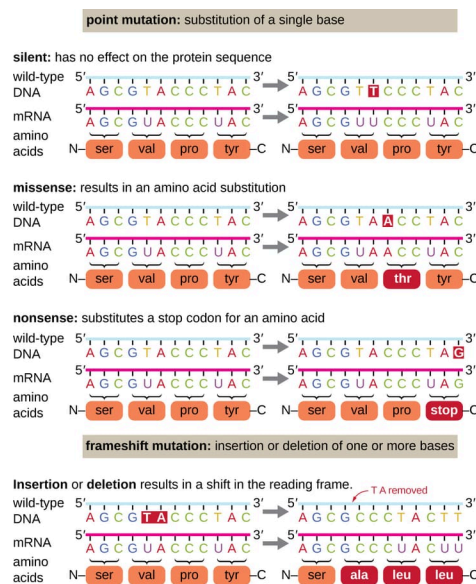


Figure 3.2.1: Mutations can lead to changes in the protein sequence encoded by the DNA. (OSC_Microbio_11_05_EffMut.jpg "Mutations" by OpenStax, LibreTexts is licensed under CC BY.)

Multiples of three

Adding or removing bases in multiples of three, will not cause a frameshift mutation. Why not?

Would mutations like this affect the protein? What additional information would you need to make a prediction about such a mutation?

Query 3.2.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Do mutations affect protein function?

Both frameshift and nonsense mutations are likely to affect protein function by producing a shortened or extremely altered peptide that is unlikely to retain enough structure to perform the intended function. However, a frameshift or nonsense mutation that occurs near the 3' end of the gene might allow the protein to retain the ability to function.

Silent mutations are least likely to affect protein function.

Missense mutations are often difficult to predict. If the change is to a similar amino acid or in a non-critical amino acid, the mutation protein may continue to function. However, if the amino acid is changed to a different type of amino acid or removes a critical amino acid (in terms of structure or activity), then the mutant protein may change in activity.

Mutations that reduce the function of a gene product are called **loss-of-function**. If the gene product is completely non-functional, the allele is termed a **null** allele. Loss-of-function and null alleles are usually (but not always) recessive in diploid organisms, because one functional allele can usually supply enough protein product.

Some mutations actually make proteins more active, for example, by altering an inhibitory domain; these mutations are called **gain-of-function**. Gain-of-function alleles are often dominant alleles because their new function or activity is produced independently of the other allele.

- [3.2: Consequences of mutations](#) by [Stefanie West Leacock](#) has no license indicated.

3.3: Example - Mutations and Cystic Fibrosis

Cystic Fibrosis (CF)

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.

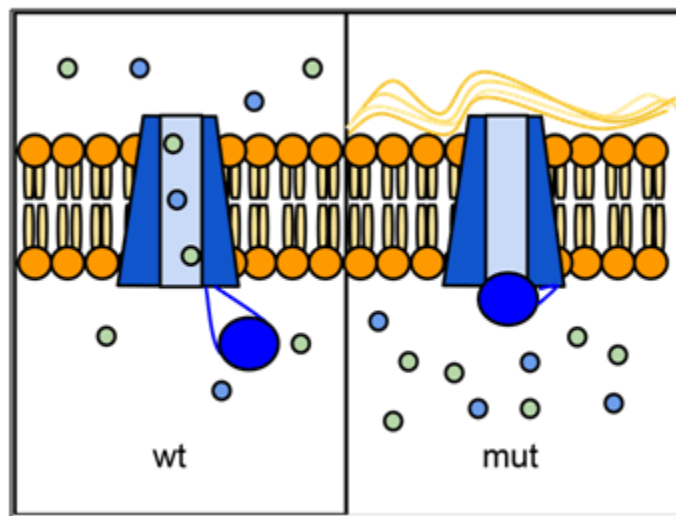


Figure 3.3.1: Wild-type and mutant forms of CFTR in the cell membrane: In wild-type, the CFTR ion channel is gated; when activated by ATP, the channel opens and allows ions to move across the membrane. In some CFTR mutants, the channel does not open, which prevents the movement of ions and water and allows mucus to build up on the lung epithelium. (CC BY via 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 (F) from position 508 of the 1480 amino acid wild-type protein. Mutation ΔF508 causes CFTR to be folded improperly in the endoplasmic reticulum (ER) and preventing CFTR from reaching the cell membrane. The allele 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 around the world 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, although little data has been uncovered to support these hypotheses. However, recent mining of health data and genetic data from over 19,000 CFTR mutant allele heterozygotes suggests that heterozygotes exhibit an increased incidence of a variety of conditions including infertility, pancreatitis, diabetes, short stature, failure to thrive, constipation and scoliosis (Miller et al., 2020).

Query 3.3.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



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® (ivacaftor) made by Vertex Pharmaceuticals, was approved by the FDA and Health Canada in 2012, decades after the CFTR gene was first mapped to DNA markers (Tsui et al. 1985; Wainwright et al. 1985; White et al., 1985) and cloned (Kerem et al. 1989; Riordan et al. 1989; Rommens et al. 1989). Kalydeco is effective on only some CFTR mutations, for example 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 the mutation by binding to CFTR and holding it in an open conformation. Kalydeco® is priced at approximately \$300,000 per patient per year, but costs may be covered by insurance. In 2015 an additional drug, Orkambi® was approved. This therapy combines ivacaftor with a second drug lumacaftor, which increases the trafficking of the $\Delta F508$ protein to the cell surface. Like Kalydeco, Orkambi costs about \$300,000 per year.

Watch this video about Cystic Fibrosis Mechanism and Treatment (from HHMI BioInteractive).

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

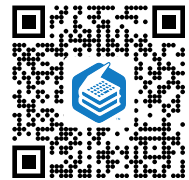
[View this YouTube Video online](#)



Query 3.3.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Contributors and Attributions

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

References

- Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science*. Sep 8; 245(4922):1073-80.
- Miller AC, Comellas AP, Hornick DB, Stoltz DA, Cavanaugh JE, Gerke AK, Welsh MJ, Zabner J, Polgreen PM. Cystic fibrosis carriers are at increased risk for a wide range of cystic fibrosis-related conditions. *Proceedings of the National Academy of Sciences* Jan 2020, 117 (3) 1621-1627; DOI: 10.1073/pnas.1914912117 (<https://www.pnas.org/content/117/3/1621>)
- Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*. Sep 8; 245(4922):1066-73.
- Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*. Sep 8; 245(4922):1059-65.
- Tsui LC, Buchwald M, Barker D, Braman JC, Knowlton R, Schumm JW, Eiberg H, Mohr J, Kennedy D, Plavsic N, et al. 1985. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 230: 1054-1057.
- Wainwright BJ, Scambler PJ, Schmidtke J, Watson EA, Law HY, Farrall M, Cooke HJ, Eiberg H, Williamson R 1985. Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* 318: 384-385.
- White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel JM, Dean M, Vande Woude G 1985. A closely linked genetic marker for cystic fibrosis. *Nature* 318: 382-384.

This page titled [3.3: Example - Mutations and Cystic Fibrosis](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [3.3.1: Example - Mutations and Cystic Fibrosis](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

3.4: Ames Test

Learning Objectives

- Explain the procedure of the Ames Test including use of histidine auxotrophs and replica plating.

Ames Test for Mutagenicity

The Ames test, developed by Bruce Ames (1928–) in the 1970s, is a method that uses bacteria for rapid, inexpensive screening of the carcinogenic potential of new chemical compounds. The test measures the mutation rate associated with exposure to the compound, which, if elevated, may indicate that exposure to this compound is associated with greater cancer risk. The Ames test uses as the test organism a strain of *Salmonella typhimurium* that is a histidine auxotroph, unable to synthesize its own histidine because of a mutation in an essential gene required for its synthesis. After exposure to a potential mutagen, these bacteria are plated onto a medium lacking histidine, and the number of mutants regaining the ability to synthesize histidine is recorded and compared with the number of such mutants that arise in the absence of the potential mutagen.

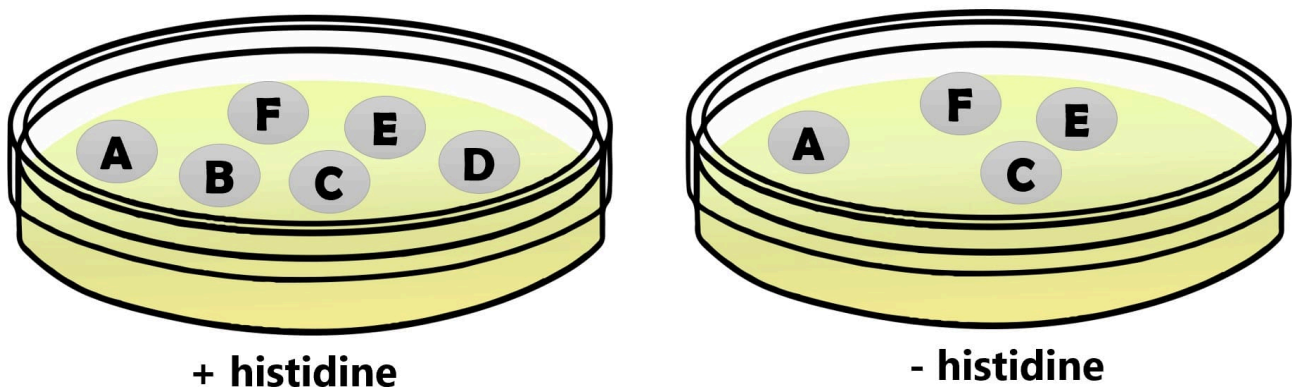
The Ames Test is possible because Ester Lederberg developed replica plating.

In replica plating, a piece of fabric is used to create an exact copy of bacterial colonies on two plates that differ in their media or environment. In the Ames test, the bacterial colonies can be grown in the presence and absence of histidine

Hear or read about how Ester Lederberg developed this critical technique in *Genetics Unzipped* (<https://geneticsunzipped.com/news/2019/3/28/powder-puffs-and-plasmids-esther-lederberg>).

Exercise 3.4.1

Histidine auxotrophs cannot synthesize their own histidine, which is required for cell viability. Histidine mutants can be identified by replica plating and identifying colonies that can grow on plates with histidine but not on plates without histidine in the media. Which colonies in the figure below could have mutations in a histidine synthesis gene (*his*⁻)?



Petri dish This Photo by Unknown Author is licensed under CC BY-SA

Answer available at the end of this section.

Chemicals that are more mutagenic will bring about more mutants with restored histidine synthesis in the Ames test. Because many chemicals are not directly mutagenic but are metabolized to mutagenic forms by liver enzymes, rat liver extract is commonly

included at the start of this experiment to mimic liver metabolism. After the Ames test is conducted, compounds identified as mutagenic are further tested for their potential carcinogenic properties by using other models, including animal models like mice and rats.

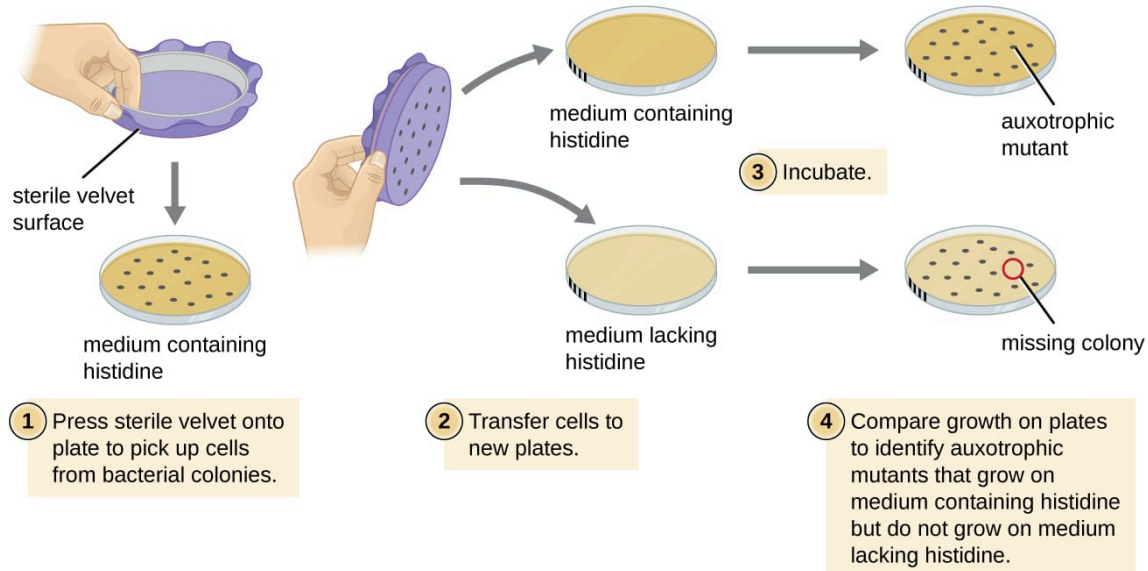


Figure 3.4.2: Identification of auxotrophic mutants, like histidine auxotrophs, is done using replica plating. After mutagenesis, colonies that grow on nutritionally complete medium, but not on medium lacking histidine, are identified as histidine auxotrophs.

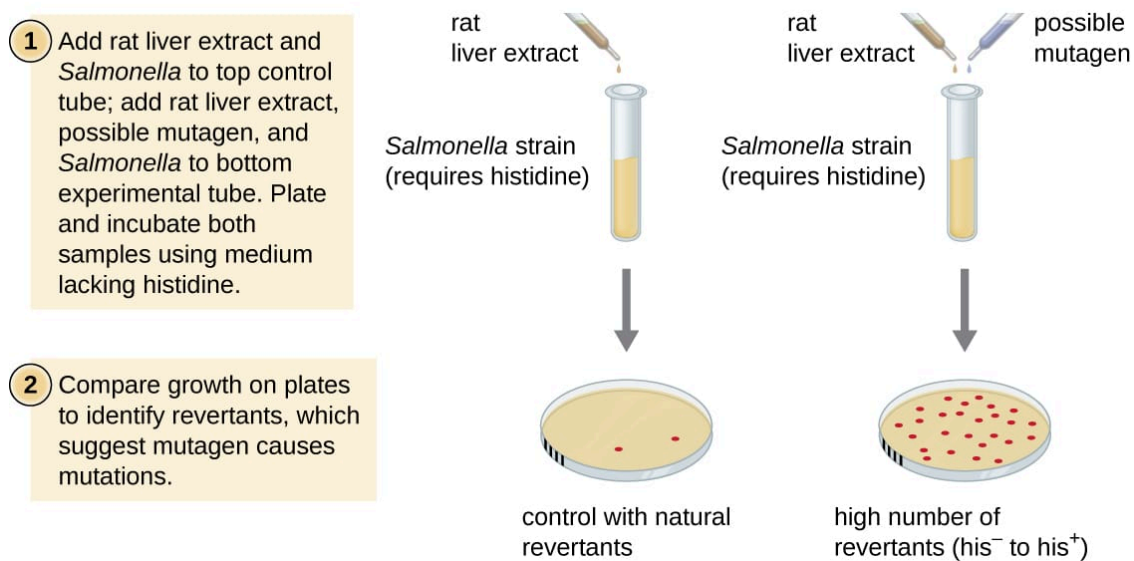


Figure 3.4.3: The Ames test is used to identify mutagenic, potentially carcinogenic chemicals. A *Salmonella* histidine auxotroph is used as the test strain, exposed to a potential mutagen or carcinogen. The number of reversion mutants capable of growing in the absence of supplied histidine is counted and compared with the number of natural reversion mutants that arise in the absence of the potential mutagen.

Questions 3.4.1

1. What mutation is used as an indicator of mutation rate in the Ames test?
2. Why can the Ames test work as a test for carcinogenicity?
3. Why do you think the Ames test is preferable to the use of animal models to screen chemical compounds for mutagenicity?

Answer available at the end of this section.

Exercise 3.4.2

Which of the following regarding the Ames test is true?

- A. It is used to identify newly formed auxotrophic mutants.
- B. It is used to identify mutants with restored biosynthetic activity.
- C. It is used to identify spontaneous mutants.
- D. It is used to identify mutants lacking photoreactivation activity.

Answer available at the end of this section.

Contributors and Attributions

- Nina Parker, (Shenandoah University), Mark Schneegurt (Wichita State University), Anh-Hue Thi Tu (Georgia Southwestern State University), Philip Lister (Central New Mexico Community College), and Brian M. Forster (Saint Joseph's University) with many contributing authors. Original content via Openstax (CC BY 4.0; Access for free at <https://openstax.org/books/microbiology/pages/1-introduction>)

Exercise 3.4.3

Zinc oxide nanoparticles have been used in food processing and packaging as well as animal farming. Mittag and colleagues tested two different sizes of zinc oxide nanoparticles for mutagenicity using the Ames Test (2021). A small portion of their results is shown in the table.

1. What is the genotype of the revertant colonies?
2. Why do revertants occur in all conditions (even the negative control)?
3. Does the data suggest that the ZnO nanoparticles are mutagenic?
4. Does this data suggest that ZnO nanoparticles are not harmful to living things?

Treatment	Average number of revertants/plate
Negative control (no nanoparticles)	148
10 µg/plate ZnO NP (<50nm)	148
10 µg/plate ZnO NP (<100nm)	143
Positive Control (a known mutagen)	1136

Source: CC BY Mittag et al (2021) via <https://www.mdpi.com/2305-6304/9/5/96/htm>

Answer available at the end of this section.

References

- Mittag A, Hoera C, Kämpfe A, Westermann M, Kuckelkorn J, Schneider T, Glei M. Cellular Uptake and Toxicological Effects of Differently Sized Zinc Oxide Nanoparticles in Intestinal Cells. *Toxics*. 2021 Apr 27;9(5):96. doi: 10.3390/toxics9050096. PMID: 33925422; PMCID: PMC8146923.
-

Answers

Exercise 3.4.1

The his⁻ colonies are B and D.

Questions 3.4.1

Exercise 3.4.2

B

Exercise 3.4.3

1. In an Ames Test, his⁻ mutants are treated with the potential mutagen. Any DNA changes that restore the ability of the cell to synthesize histidine will produce his⁺ cells that have the ability to grow on media lacking histidine.
2. Random errors that occur during DNA replication are a constant source of new mutations; therefore the Ames Test is looking to detect a rate of mutations that occurs above this baseline level.
3. No. The number of revertants does not differ substantially from the negative control. (The complete paper reports that the *p* values from statistical tests of the data are not significant.)
4. Not necessarily. Other tests suggest that the nanoparticles could increase cell death. DNA damage is only one aspect to consider when determining whether a chemical is harmful to living things.

This page titled [3.4: Ames Test](#) is shared under a license and was authored, remixed, and/or curated by Cheneir Neely.

- [4.8: Genetic screens - Inherited Phenotypes Reveal Functions of Genes](#) by [Stefanie West Leacock](#) has no license indicated.

3.5: DNA Repair

Learning Objectives

- Describe and compare direct repair, base excision repair, nucleotide excision repair, and mismatch repair pathways.
- Understand the situations in which non-homologous end-joining and homologous recombination are used for repair.

Maintaining the Integrity of the Cell's Information: DNA Repair

If DNA is the master copy of instructions for an organism, then it is important not to make mistakes when copying the DNA to pass on to new cells. Although proofreading by DNA polymerases greatly increases the accuracy of replication, there are additional mechanisms in cells to further ensure that newly replicated DNA is a faithful copy of the original, and also to repair damage to DNA during the normal life of a cell.

All DNA suffers damage over time, from exposure to ultraviolet and other radiation, as well as from various chemicals in the environment. Even chemical reactions naturally occurring within cells can give rise to compounds that can damage DNA. As you already know, even minor changes in DNA sequence, such as point mutations can sometimes have far-reaching consequences. Likewise, unrepaired damage caused by radiation, environmental chemicals or even normal cellular chemistry can interfere with the accurate transmission of information in DNA. Maintaining the integrity of the cell's "blueprint" is of vital importance and this is reflected in the numerous mechanisms that exist to repair mistakes and damage in DNA.

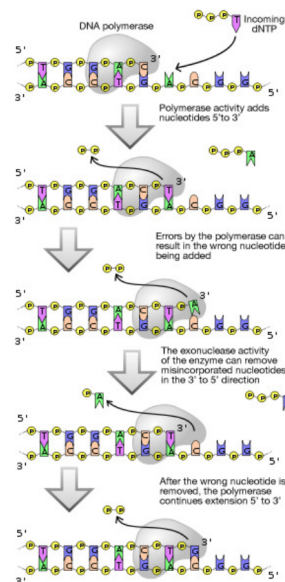


Figure 3.5.1: Example of the proofreading function of DNA polymerase, in which DNA polymerase 3' to 5' exonuclease activity removes the incorrect A paired with C. Then DNA polymerase continues extending DNA in the 5' to 3' direction.

Post-Replicative Mismatch Repair

Does proofreading eliminate all errors made during replication? No. While proof-reading significantly reduces the error rate, not all mistakes are fixed on the fly by DNA polymerases. What mechanisms exist to correct the replication errors that are missed by the proof-reading function of DNA polymerases?

Errors that slip by proofreading during replication can be corrected by a mechanism called mismatch repair. While the error rate of DNA replication is about one in 10^7 nucleotides in the absence of mismatch repair, this rate is further reduced a hundred-fold to one in 10^9 nucleotides when mismatch repair is functional.

What are the tasks that a mismatch repair system must accomplish?

- Scan newly made DNA to see if there are any mispaired bases (e.g., a G paired to a T)
- Identify and cut out the region of the mismatch
- Correctly fill in the gap created by the excision of the mismatch region.

Importantly, the mismatch repair system must have a means to distinguish the newly made DNA strand from the template strand, if replication errors are to be fixed correctly. In other words, when the mismatch repair system encounters an A-G mispair, for example, it must know whether the A should be removed and replaced with a C or if the G should be removed and replaced with a T.

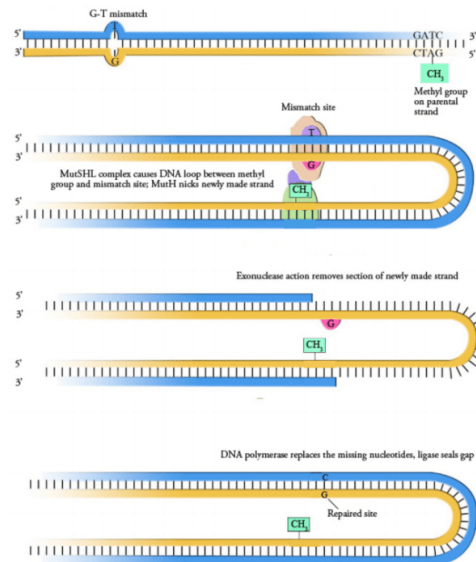


Figure 3.5.2: Mismatch repair in bacteria uses the MutSHL protein complex to recognize a mismatch, such as G-T. A loop of DNA forms between the methyl group on the parental strand and the mismatched base pair. An exonuclease removes only the newly-synthesized DNA strand from the loop. DNA polymerase re-synthesizes the removed region and ligase seals the strand.

Mismatch repair has been well studied in bacteria, and the proteins involved have been identified. Eukaryotes have a mismatch repair system that repairs not only single base mismatches but also insertions and deletions. In bacteria, mismatch repair proteins are encoded by a group of genes collectively known as the Mut (for mutator and pronounced "mute") genes. Some of the most important components of the mismatch repair machinery are the proteins MutS, MutL, and MutH (sometimes known as MutSHL complex). MutS recognizes the mismatch and recruits MutL and MutH, as well as other proteins which help cut out the region containing the mismatch. A DNA polymerase and ligase fill in the gap and join the ends, respectively.

But how does the mismatch repair system distinguish between the original and the new strands of DNA? Bacteria have a system that methylates DNA at GATC sequences to solve this problem. *E.coli* has an enzyme that adds methyl groups on the to adenines in GATC sequences. Newly-replicated DNA strands lacks this methylation and thus, can be distinguished from the template strand, which is methylated. The Mut genes use the methylation to direct removal of the non-methylated strand, keeping the methylated strand as the template to repair the region. Because methylation is the criterion that enables the mismatch repair system to choose the strand that is repaired, the bacterial mismatch repair system is described as being methyl-directed. Note that the newly-synthesized strand will eventually also be methylated by the cellular enzymes, so this repair system is only useful immediately following DNA replication, before the new strand becomes marked as well.

Although eukaryotic cells have Mut gene homologs, they do not use the same methylation mechanism to distinguish the new strand from the template. Although the mechanisms, are not yet fully understood, recent research suggests that the mismatch repair system in eukaryotes may be coupled to DNA replication so that the strand being synthesized strand can be identified via nicks or proteins asymmetrically bound to the replication machinery (reviewed in Reyes et al. 2015).

Systems to Repair Damage to DNA

Some mistakes are made when DNA is copied, if the wrong base is inserted during synthesis of the new strand. But even DNA that is not being replicated can get damaged or mutated. These sorts of damage are not associated with DNA replication, rather they can occur at any time.

Recall that some major causes of DNA damage are:

- Radiation (e.g., UV rays in sunlight, in tanning booths)
- Exposure to damaging chemicals (such as benzopyrene in car exhaust and cigarette smoke)
- Chemical reactions within the cell (such as the deamination of cytosine to give uracil).

How do cells repair such damage? Cells have several ways to remove the sorts of damage described above, with excision repair being a common strategy. Excision repair is a general term for the cutting out and re-synthesis of the damaged region of the DNA. There are a couple of varieties of excision repair:

Direct repair

Cells may possess mechanisms to repair types of damage that occur frequently, for example there are proteins that remove alkyl groups from DNA bases. Most organisms, especially those exposed to high amounts of UV radiation, have enzymes that can directly repair thymine dimers. The process, called photoreactivation or light repair, uses an enzyme, photolyase, that depends on visible light to directly breaks the bonds joining the thymine bases. Notably, placental mammals do not have photolyases capable of this type of repair and so must use an alternate pathway to repair UV-induced thymine dimers.

Nucleotide Excision Repair (NER)

This system fixes damage by chemicals as well as UV damage. In nucleotide excision repair, the damage is recognized and a cut is made on either side of the damaged region by an endonuclease enzyme (sometimes called an excision nuclease or exonuclease). A short portion of the DNA strand containing the damage is then removed and a DNA polymerase fills in the gap with the appropriate nucleotides. The newly made DNA is joined to the rest of the DNA backbone by the enzyme DNA ligase. In *E. coli*, NER is carried out by a group of proteins encoded by the *uvrABC* genes. As you can see, NER is similar, in principle, to mismatch repair. However, in NER, the distortion of the helix, caused by the DNA damage, clearly indicates which strand of the DNA needs to be removed and replaced.

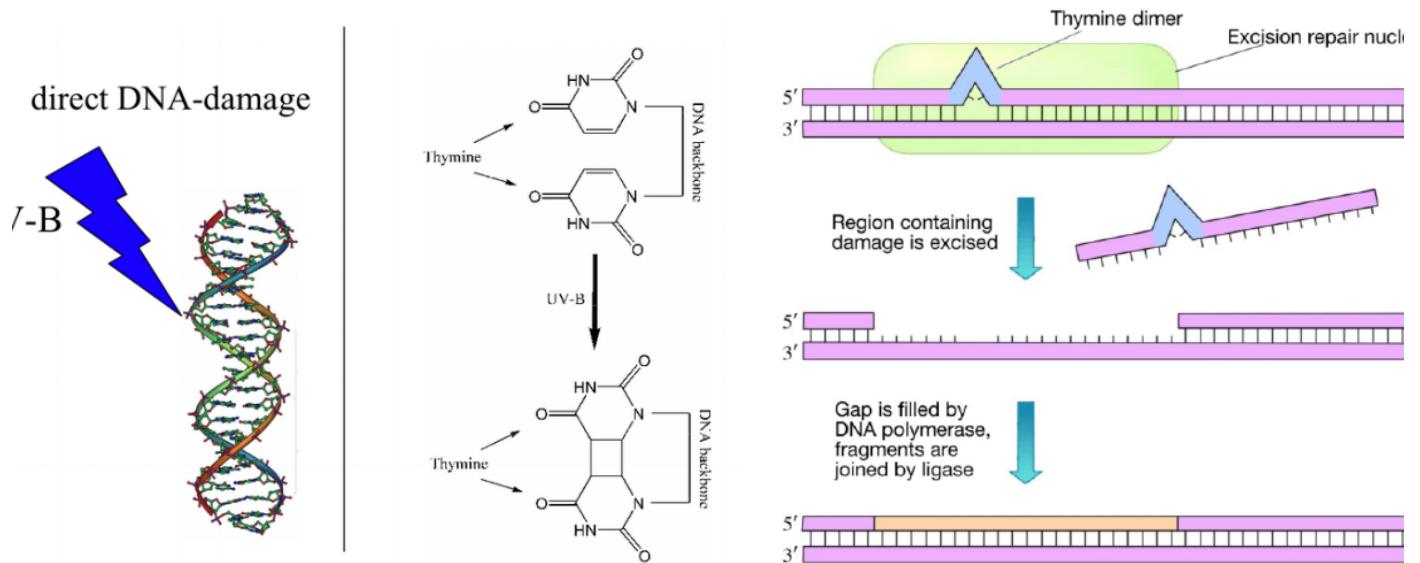


Figure 3.5.3: (Left) Thymine dimer formation can occur when UV-B light hits DNA and covalent bonds form between adjacent thymine bases. (Right) Thymine dimers can be removed by an excision repair nuclease that recognizes the dimer, excising the region of one strand of DNA containing the dimer. The gap is filled by DNA polymerase and joined to the original strand by ligase.

Base Excision Repair (BER)

BER deals with situations like the deamination of cytosine to uracil. As noted earlier, cytosines in DNA sometimes undergo deamination to form the base uracil. Because cytosines pair with guanines and uracils pair with adenine, the conversion of cytosine to uracil in the DNA would lead to the insertion of an A in the newly replicated strand instead of the G that should have gone in across from a C. To prevent this from happening, uracils are removed from DNA by base excision repair.

In base excision repair, a single base is first removed from the DNA, followed by removal of a region of the DNA surrounding the missing base. The gap is then repaired. The removal of uracil from DNA is accomplished by the enzyme uracil DNA glycosylase, which breaks the bond between the uracil and the sugar in the nucleotide. The removal of the uracil base creates a gap called an apyrimidinic site (AP site). The presence of the AP site triggers the activity of an AP endonuclease that cuts the DNA backbone. A short region of the DNA surrounding the site of the original uracil is then removed and replaced.

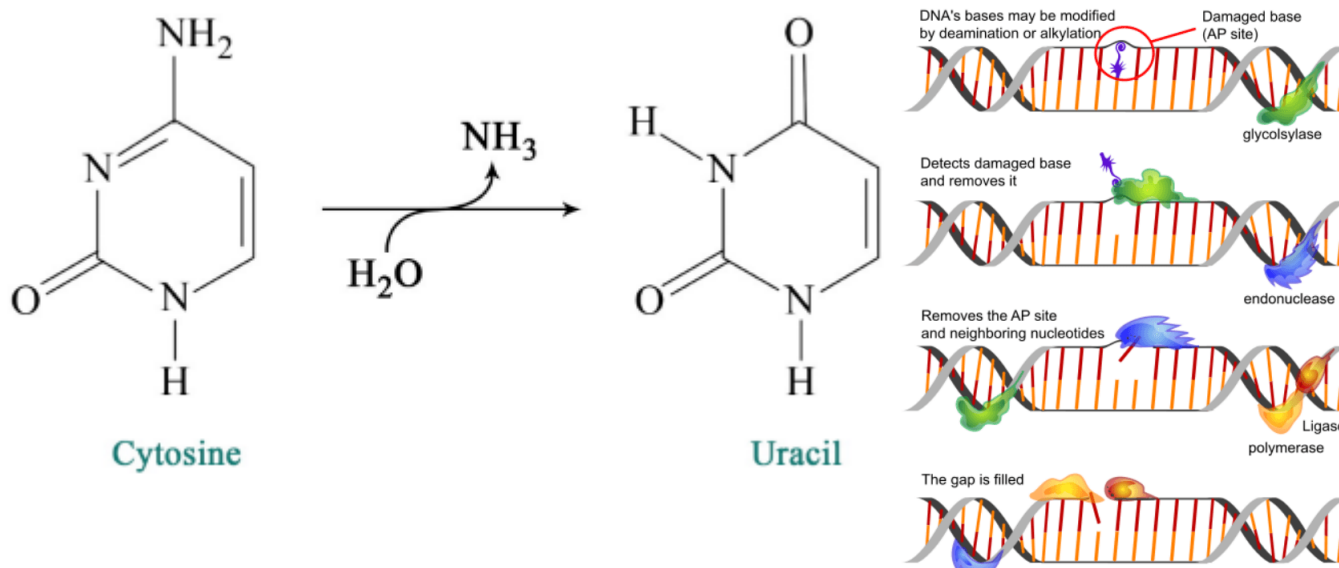


Figure 3.5.4: (Left) Deamination of cytosine removes an -NH₃ group converting cytosine to uracil. (Right) In this example a glycosylase recognizes a damaged base and removes the base (without breaking the DNA backbone) creating a apyrimidinic site. An endonuclease then cuts the DNA backbone at the site. Bases near the AP site are removed and then replaced by DNA polymerase and the repaired strand is sealed by ligase.

Nobel Prize Winning Work

Tomas Lindahl, Paul Modrich and Aziz Sancar shared the 2015 Nobel Prize in Chemistry "for mechanistic studies of DNA repair" for their characterization of the proteins and processes involved in mismatch repair, base excision repair, and nucleotide excision repair (<https://www.nobelprize.org/prizes/chemistry/2015/press-release/>).

Additional repair pathways

What if DNA damage is not isolated to the bases? Damage to DNA may also involved breakage of the sugar-phosphate backbone of the chromosome. Cells have some additional options to repair damage to large regions of DNA. The first is **non-homologous end-joining (NHEJ)**. As the name suggests, in this process the ends of two broken chromosomes are ligated together without identifying whether or not they were actually part of the same chromosome. This mechanism can be thought of as an emergency response to broken chromosomes; without repair the broken ends of DNA could be further damaged by exonucleases, so joining the two ends prevents further loss of genetic material. However, the process is imperfect and can result in insertions, deletions, or translocations (joining of two different chromosomes). A second option is **homologous recombination repair**, in which a homologous chromosome is used as a template to synthesize a broken or damaged region of DNA. The fundamental mechanism here is similar to the process of recombination used during meiosis.

Contributors and Attributions

- Dr. Kevin Ahern and Dr. Indira Rajagopal (Oregon State University)

References

Reyes, G.X., Schmidt, T.T., Kolodner, R.D. *et al.* New insights into the mechanism of DNA mismatch repair. *Chromosoma* **124**, 443–462 (2015). <https://doi.org/10.1007/s00412-015-0514-0>

This page titled [3.5: DNA Repair](#) is shared under a license and was authored, remixed, and/or curated by Cheneir Neely.

- [5.2: DNA Repair](#) by Kevin Ahern and Indira Rajagopal is licensed CC BY-NC-SA 4.0. Original source: <https://biochem.oregonstate.edu/undergraduate/educational-resources>.
- [3.5: DNA Repair](#) has no license indicated.

3.6: Chapter 3 Review Questions

3.1: Origins of Mutations

Query 3.6.3.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 3.6.3.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 3.6.3.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 3.6.3.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



3.2: Consequences of Mutations

Query 3.6.3.5

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 3.6.3.6

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 3.6.3.7

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



3.4: Ames Test

Query 3.6.3.8

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 3.6.3.9

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



3.5: DNA Repair

Query 3.6.3.10

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 3.6.3.11

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

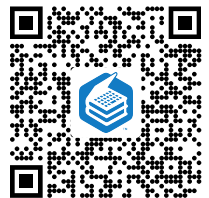
[View this Interactive Content online](#)



Query 3.6.3.12

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



4: INHERITANCE

4.1: Meiosis

4.1.1: Homologous recombination

4.1.2: CRISPR

4.2: Mendelian Genetics

4.2.1: Monohybrid Crosses and Segregation

4.2.2: Dihybrid Crosses and Independent Assortment

4.3: Pedigrees

4.4: Exceptions to autosomal inheritance

4.4.1: Inheritance patterns for X-linked and Y-linked genes

4.4.2: Organellar Inheritance

4.5: Linkage

4.5.1: Linkage and Mapping

4.5.2: GWAS

4.6: Exceptions to simple dominance

4.7: Gene Interactions

4.7.1: Complementation Test

4.8: Chapter 4 Review Questions (draft)

Template:HideTOC

This page titled [4: Inheritance](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4: Inheritance](#) by [Stefanie West Leacock](#) has no license indicated.

4.1: Meiosis

Learning Objectives

- Describe and draw the key events and stages of meiosis that lead to haploid gametes.
- Recall that homologous chromosomes separate during meiosis I (a reductional division) and that sister chromatids separate during meiosis II (an equational division).
- Compare mitosis and meiosis.
- Compare the processes of oogenesis and spermatogenesis in humans, including the chromosome complements of the gametes.

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 begins similarly to mitosis (a cell has replicated its chromosomes and grown large enough to divide), but requires two rounds of division. In the first, known as meiosis I, the homologous chromosomes separate and segregate. During meiosis II the sister chromatids separate and segregate. Note how meiosis 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.

Meiosis is divided into two stages designated by the roman numerals I (one) and II (two). 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**. 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.

Meiosis I

In meiosis I replicated, homologous chromosomes pair up, or synapse, during the pachytene stage of prophase I, line up in the middle of the cell during metaphase I, and separate 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, so that the homologous chromosomes can be separated.

Query 4.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



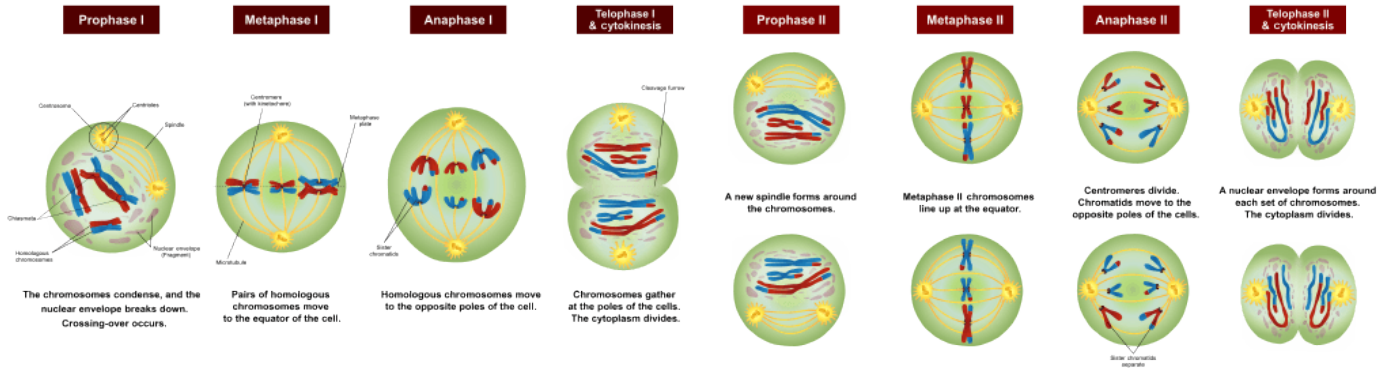


Figure 4.1.1: Meiosis is a process in which a diploid cell divides into 4 haploid cells. At the end of Meiosis there are four genetically different cells. The diagram shows Meiosis as a non-cyclic process. (CC BY SA Ali Zifan via https://commons.wikimedia.org/wiki/File:Meiosis_Stages.svg)

Chromosome condensation during meiosis

As meiosis proceeds, chromatin becomes increasingly condensed. In some organisms, the DNA becomes so condensed that it appears as a spot of DNA instead of a line under the microscope. As you might expect from condensed chromatin, little transcriptional activity occurs during these stages of meiosis, so cells must produce the needed mRNAs in advance of meiosis.

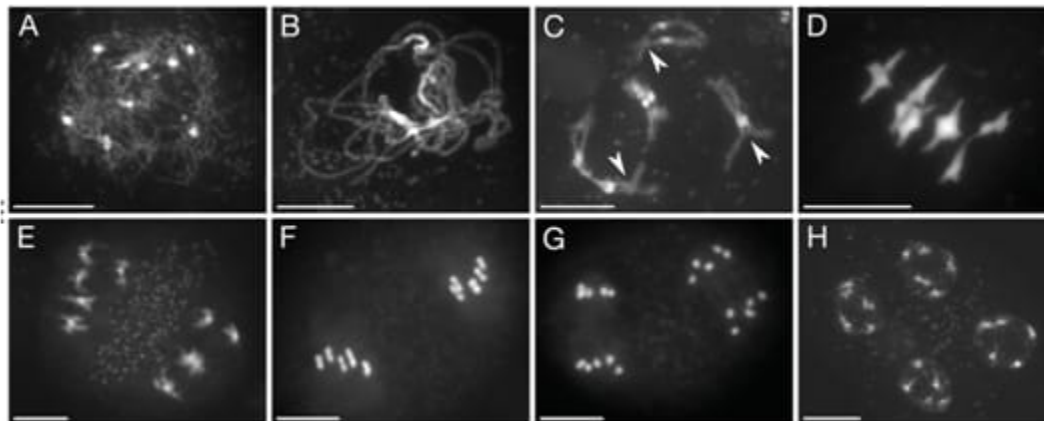


Figure 4.1.2: Staining of DNA during meiosis in Arabidopsis shows an increasing degree of chromosome condensation at different stages of prophase I (A-C). Subsequent phases of meiosis are metaphase I (D), telophase I (E), metaphase II (F), anaphase II (G), and telophase II (H). (Copyright CC by Chelysheva et al. (2008) The Arabidopsis BLAP75/Rmi1 Homologue Plays Crucial Roles in Meiotic Double-Strand Break Repair. *PLoS Genet* 4(12): e1000309.)

Homologous Recombination

Within the synaptonemal complex during prophase 1, homologous recombination, or crossing over, occurs. These are places where DNA endonucleases break two non-sister chromatids in similar locations and then covalently reattach non-sister chromatids together to create a crossover between non-sister chromatids (4.1.1: Homologous recombination). 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 that they segregate successfully; they also cause the reshuffling of allele combinations to create genetic diversity, which can have an important effect on evolution.

Meiosis II

At the completion of meiosis I there are two haploid cells, each with one, replicated copy of each chromosome ($1n$). Because only one copy of each homolog is present, bivalents are not formed. In metaphase of meiosis II, 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 II.

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 four cells have different genetic composition. There will be allelic differences among the gametes.

Query 4.1.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Outcomes of meiosis

The outcome of meiosis is a cell or cells with half the number of chromosomes as the starting cell. But what combinations of chromosomes are possible?

The homologous chromosomes separate during meiosis I, but the separation of the pairs of homologs is independent of other homologs. As an example in the figure below, for a cell with two pairs of chromosomes ($2n=4$), there can be four possible combinations of the four chromosomes. When we add recombination and additional pairs of chromosomes, there are almost infinite combinations of chromosomes in gametes.

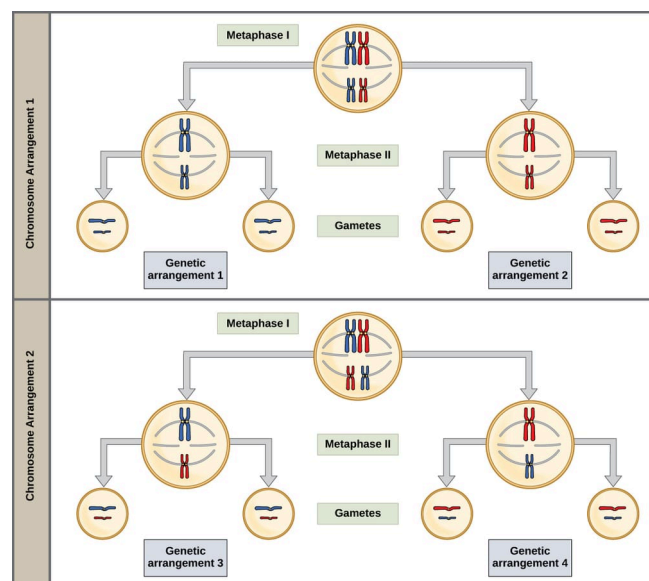


Figure 4.1.2: Random, independent assortment during metaphase I can be demonstrated by considering a cell with a set of two chromosomes ($n = 2$). In this case, there are two possible arrangements at the equatorial plane in metaphase I. The total possible number of different gametes is $2n$, where n 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. (CC-BY OpenStax [Figure_11_01_03.jpg](#))

Watch the video below to follow how 4 pair of chromosomes are passed during meiosis I and II.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 4.1.1: How to count the number of chromosomes during the different phases of meiosis is explained. (www.youtube.com/watch?v=vCyioKyYkMw&t=3s)

Gamete maturation

In animals and plants, the cells produced at the end of meiosis need to mature before they become functional gametes.

Spermatogenesis

In most male animals, the four products of meiosis are called **spermatids**. They change morphology to develop tails and become functional sperm cells. In the meiotic steps of spermatogenesis, the cell divisions are equal, with the meiotic spindle aligned with the center of the cell, and the cells have equal amounts of cytoplasm, much like an average cell that has undergone mitosis. The streamlined, minimal-cytoplasm mature sperm is a product of post-meiotic differentiation, in which it gains the flagellar tail, and ejects most of its cytoplasmic material, keeping only some mitochondria to power the flagella and an acrosomal vesicle that contains the enzymes and other molecules needed to reach and fuse with (i.e. fertilize) a mature egg.

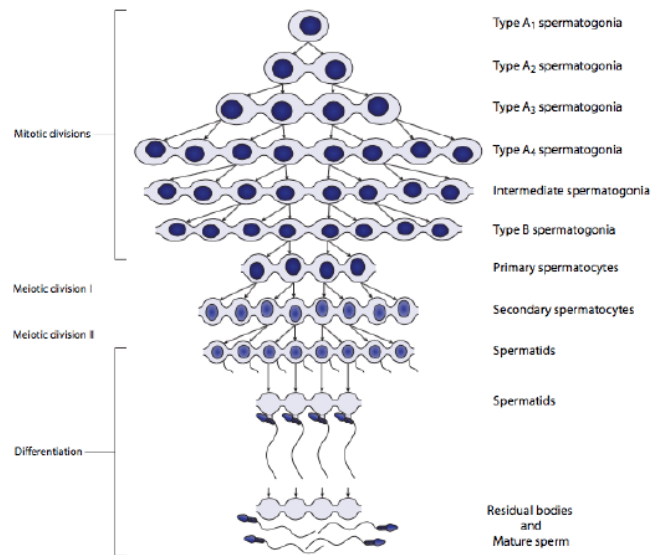


Figure 4.1.3: Spermatogenesis begins with spermatogonia that divide by mitosis to produce primary spermatocytes, which undergo meiosis I to produce secondary spermatocytes. These cells undergo meiosis II to produce spermatids that undergo differentiation to produce mature sperm and residual bodies with unneeded cellular material.

Oogenesis

In female animals, the gametes are **oocytes**. Each mature ovum (egg) will need to be as large possible to contain the maximum amount of cytoplasm including organelles, proteins, mRNAs, and nutrients to support the embryo after fertilization. To create large oocytes, only one of the four products of meiosis becomes an egg. The other three cells end up as tiny "disposable" cells called **polar bodies**, essentially little "packages" of extra DNA and very little cytoplasm. These cells are not viable and will eventually be degraded.

How can you make a really small cell?

The asymmetric distribution of cytoplasm in the first meiotic division for oocytes is due to the position of the meiotic spindle in the periphery of the cell rather than centered. During oogenesis, chromosomes do not line up in the middle of the cell during metaphase I or II. Because the center of the spindle determines the position of the contractile ring for cytokinesis, one small and one large cell are produced.

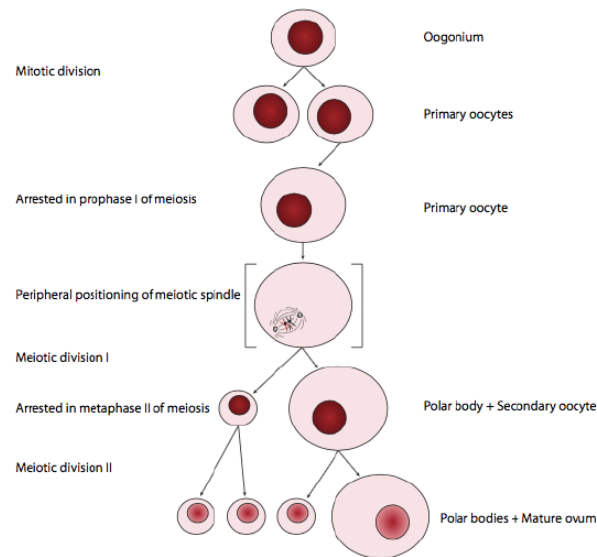


Figure 4.1.4: Representation of typical stages of oogenesis in animals. Primary oocytes divide asymmetrically by positioning the meiotic spindle near the periphery of the oocyte to produce a polar body and secondary oocyte. The secondary oocyte also divides asymmetrically to produce a polar body and mature ovum.

Timing of Spermatogenesis and Oogenesis

In addition to the differences in gamete size and number, in mammals the timing of meiosis differs between males and females. In males, germ cells are pre-meiotic at birth and do not enter meiosis until the onset of puberty. Mitosis maintains a population of precursor cells, so that sperm production can continue throughout adulthood. In females, germ cells enter meiosis I during embryonic development. These primary oocytes remain arrested ("stuck") in meiosis I until puberty. After this time, one primary oocyte per month (roughly for humans, depends upon cycle length for other mammals) completes meiosis I, enters meiosis II and is ovulated. Actually, meiosis II is only completed if the oocyte is successfully fertilized! The timing of meiotic arrest can differ between different species. For example, in the nematode *C. elegans*, oocytes are arrested in late prophase of meiosis I and only complete meiosis I and II rapidly after fertilization.

A note about plants

In plants, the products of meiosis reproduce a few times using mitosis as they develop into functional male or female gametes.

Fertilization

The purpose of gametes is to allow reproduction generation after generation. By uniting two gametes with half the number of chromosomes, the full chromosome number is restored each generation. Remember that homologous recombination and assortment of chromosomes create a genetically diverse population of gametes.

Fertilization restores the diploid number

$$\text{oocyte (n) + sperm (n) = zygote (2n)}$$

Errors in meiosis

Like any biological process, errors can occur during meiosis. If homologous chromosomes or sister chromatids are not correctly

distributed during meiosis (known as nondisjunction), gametes can have too many or too few chromosomes ([5.1: Changes in Chromosome Number](#)).

Contributors and Attributions

- [Dr. Todd Nickle and Isabelle Barrette-Ng](#) (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.
- [15.8: Meiosis](#)
- Stefanie Leacock, University of Arkansas-Little Rock

This page titled [4.1: Meiosis](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.1: Meiosis](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

4.1.1: Homologous recombination

Learning Objectives

- Diagram the process of homologous recombination to explain how meiosis generates new combinations of alleles on chromosomes.
- Understand that recombination involves a physical breakage and ligation of the DNA backbones of sister chromatids and occurs during the pachytene of prophase of meiosis I.
- Recall that Holliday junctions are linked strands of DNA generated during recombination that must be resolved by additional DNA cuts and ligations and can lead to either recombinant or nonrecombinant chromosomes.

Because each chromosome undergoes at least one recombination event during meiosis, new combinations of alleles are generated. The arrangement of alleles inherited from each parent are not preserved, but rather the new germ cells carry chromosomes with new combinations of alleles of the genes. This remixing of combinations of alleles is a rich source of diversity in a population.

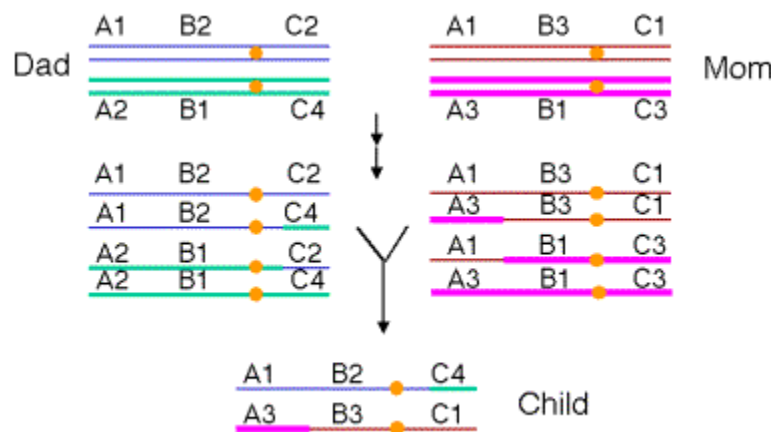


Figure 4.1.1.1: Recombination during meiosis generates new combinations of alleles in the offspring. One homologous pair of chromosomes is illustrated as a bivalent in meiotic prophase I. Each line represents double-stranded DNA in a sister chromatid. The two chromosomes in the "Dad" or sperm parent (inherited from the paternal grandparents) are blue and green; the homologous chromosomes in the "Mom" or egg parent (inherited from the maternal grandparents) are brown and pink. All chromosomes have genes A, B and C in the same order; different numbers refer to different alleles. In this illustration, a crossover on the short arm of the chromosome during development of the male germ cells links allele 4 of gene C with alleles 1 of gene A and allele 2 of gene B, as well as the reciprocal arrangement. A crossover on the long arm of the chromosome is illustrated for development of the female germ cell, making the new combination A3, B3 and C1. One possible child can have the new chromosomes A1 B2 C4 and A3 B3 C1. Note that neither of these combinations was in the father or mother. ([image012.png](#))

Over time, recombination will separate alleles at one locus from alleles at a linked locus. A chromosome is not fixed through generations, but rather it is "fluid," having many different combinations of alleles. This allows nonfunctional (less functional) alleles to be cleared from a population. If recombination did not occur, then one deleterious mutant allele would cause an entire chromosome to be eliminated from the population. However, with recombination, the mutant allele can be separated from the other genes on that chromosome. Then negative selection can remove defective alleles of a gene from a population while affecting the frequency of alleles only of genes in tight linkage to the mutant gene. Conversely, the rare beneficial alleles of genes can be tested in a population without being irreversibly linked to any potentially deleterious mutant alleles of nearby genes. This keeps the effective target size for mutation close to that of a gene, not the whole chromosome.

Contributors and Attributions

- Ross C. Hardison, T. Ming Chu Professor of [Biochemistry and Molecular Biology](#) ([The Pennsylvania State University](#))

Mechanisms of recombination

Recombination occurs when a piece of the paternal chromosome is swapped for the homologous piece of DNA on the matching maternal chromosome (or vice versa). Obviously, this kind of a DNA swap must be done carefully and with equivalence, so that the resultant DNA does not gain or lose information. To ensure this precision in recombination, the non-sister homologous chromatids are held together via proteins in a synaptonemal complex (SC) during prophase I. This ladder-like complex begins to form in the zygotene stage of prophase I and completes in pachytene. The complete SC consists of proteinaceous lateral elements (aka axial elements) that run along the length of the chromatids and a short central element composed of fibrous proteins forming the rungs of the ladder perpendicular to the two lateral elements.

Recombination may occur with or without the formation of double-strand breaks, and in fact, can occur without the formation of the synaptonemal complex, although the SC probably enhances the efficiency of recombination. In *S. pombe*, meiosis occurs without the formation of a synaptonemal complex, but there are small discontinuous structures somewhat similar to parts of the SC. In the fruit fly, *Drosophila melanogaster*, females undergo meiosis using a synaptonemal complex, but males do not undergo meiotic recombination, and their chromosomes do not form synaptonemal complexes. In most cases, recombination is preceded by the formation of recombination nodules, which are protein complexes that form at potential points for recombination.

The best studied mechanism for meiotic recombination involves a double-stranded break of one of the chromosomes initiated by the meiosis-specific endonuclease, Spo11. The 5' ends (one in each direction) of this cut are degraded slightly to form 3' single-stranded overhangs. These unpaired ends lead to the formation of **Holliday junctions** (named after Robin Holliday) with a strand from another chromatid acting as a template for synthesis of the missing portion of the chromatids, leading to two sister chromatids that are "entangled" by having one strand of DNA paired with a different chromatid. This entanglement may be resolved with or without a crossover. The recombination is initiated in pachytene and completes in diplotene, at which time the synaptonemal complex breaks down. As the chromatids begin to separate, **chiasmata** (sites where chromatids remain in contact) become apparent at some of the recombination sites. As prophase completes, the chiasmata resolve from the center of the chromosomes to the ends.

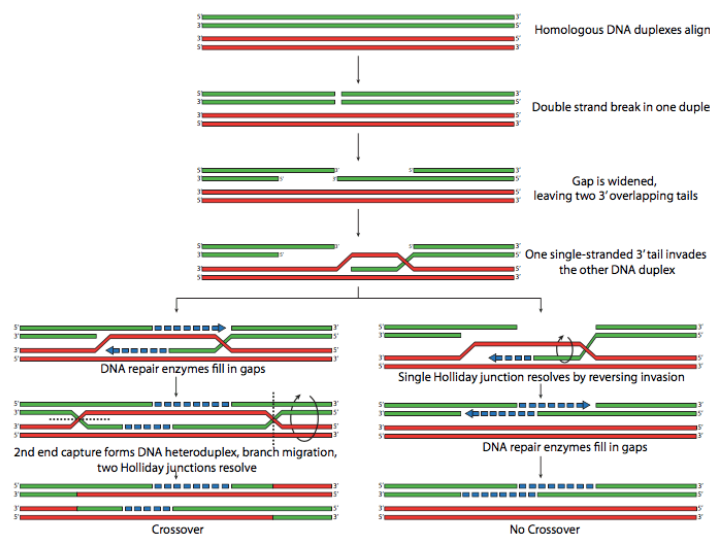


Figure 4.1.1.2. Recombination of homologous chromosomes.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 4.1.1.1: In this animation, explore how a Holliday junction is formed, and how it can subsequently be resolved. (www.youtube.com/watch?v=MvnWxN81Qps)

Recombination can also be used as a repair mechanism!

Homologous recombination occurs in meiotic cells. In most species, every chromosome will undergo at least one recombination event. However, the ability to use one chromosome as a template for a broken one can also be used in instances of DNA damage, especially DNA backbone breaks. Remember that outside of meiosis I, homologous chromosomes are not "paired up" in cells, although recent studies suggest they would be in the same region of the nucleus during interphase. The ability of the cell to use homologous recombination to repair a broken chromosome depends on whether or not it can "find" a homologous sequence to use as a template to re-synthesize a damaged chromosome.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](https://www.youtube.com/watch?v=86JCMM5kb2A)



Video 4.1.1.2: This animation shows the use of homologous recombination as a repair mechanism for a double strand break. (Oxford Academic via <https://www.youtube.com/watch?v=86JCMM5kb2A>)

Contributors and Attributions

- Wong Molecules and Mechanisms : 15.8: Meiosis

This page titled [4.1.1: Homologous recombination](#) is shared under a license and was authored, remixed, and/or curated by Cheneir Neely.

- [4.1.1: Homologous recombination](#) has no license indicated.

4.1.2: CRISPR

Learning Objectives

- Explain the natural role of CRISPR sequences in bacteria and how this system has been adapted for editing other genomes.
- Describe the roles of homologous recombination or non-homologous end-joining in CRISPR-mediated genome editing.
- Distinguish between somatic and germ-line genome editing.

We learned previously that most mutations occur at random, so while any mutation is possible, until recently scientists had limited and labor-intensive tools to create desired mutations. The discovery and elucidation of the mechanisms of a bacterial defense system changed this reality.

What is CRISPR?

Viruses are infectious particles of genetic material that affect virtually every type of organism, and therefore defense mechanisms against viruses have evolved. In the case of prokaryotes, the CRISPR locus is a region of genomic DNA to which viral genome sequences can be added to serve as a "memory" of previous infections for future defense against the same virus. The name **CRISPR** stands for Clustered Regularly Interspaced Short Palindromic Repeats. The short regions (20-50 nucleotides) of viral DNA are the "spacers" and are used to transcribe CRISPR RNAs (crRNAs). The crRNAs interact with an additional RNA (tracrRNA) and a CRISPR-associated (Cas) protein, which cleaves complementary DNA that enters the cell, thus breaking the viral replication cycle.

EXPERIMENT: Do spacer sequences confer bacteriophage resistance?

A wild-type strain of *Streptococcus thermophilus* was exposed to two types of bacteriophages (phage 858 and phage 2972) and mutant bacteria resistant to each strain were identified. The spacer sequences in the CRISPR locus in the mutants were then determined and compared to the two bacteriophage genome sequences. Mutants with spacer sequences that corresponded to a sequence in the bacteriophage genome were resistant to specifically to that bacteriophage (Barrangou et al. 2007). Even a few mismatched bases did not confer resistance.

Modifying CRISPR/Cas systems for genome editing

The discovery that an RNA could guide an endonuclease to a specific genomic position to cut DNA presented an opportunity to modify this system to edit genomes. Scientists discovered that they could combine the two RNAs (tracrRNA and crRNAs) into a single guide RNA (sgRNA). The Cas enzyme can be provided as DNA via a plasmid or RNA encoding the protein, which is produced by the cellular gene expression machinery, or directly as purified protein. The Cas9 protein does require a consensus sequence to interact with DNA (5' NGG 3'), but this motif is relatively common, allowing targeting of most loci.

But how can creating double-stranded breaks in genome benefit an organism? The answer lies in cellular DNA repair. Double-stranded breaks are typically repaired by non-homologous end-joining (NHEJ) or homologous recombination. NHEJ frequently makes small insertions or deletions, which could for example, disrupt the coding sequence of a protein to inactivate it. Homologous recombination mediated repair can occur either by using the homologous chromosome as a template for repair or by providing exogenous DNA with the desired edit, effectively "tricking" the cell into making the repair based on the donor DNA provided. Importantly, at this time, the efficiency of CRISPR gene editing is not 100% so correctly edited cells must be identified. Additionally, scientists are continuing to study the incidence of off-target (non-intended) edits when using CRISPR-based systems.

Exercise 4.1.2.1

Order the steps in CRISPR gene editing.

1. Cellular DNA repair
2. Cas endonuclease binds genomic DNA at location determined by the complementary guide RNA
3. Identification of edited cells
4. Cleavage of genomic DNA
5. Cas endonuclease and guide RNA are introduced into cells
6. Select a guide RNA sequence

For help, use this link for a review of the steps <https://media.hhmi.org/biointeractive/click/CRISPR/> (Click-N-Learn by HHMI Biointeractive).

Answer available at the end of this section.

Nobel Prize-Winning Work

In 2020, Emmanuelle Charpentier from the Max Planck Unit for the Science of Pathogens, Berlin, Germany and Jennifer Doudna from the University of California, Berkeley, USA shared the Nobel Prize in Chemistry “for the development of a method for genome editing” having characterized the CRISPR system in prokaryotes and shown that it could be used in other genomes as well (<https://www.nobelprize.org/prizes/ch...press-release/>).

Inherited or not?

In terms of human gene editing, the distinction between somatic and germline gene editing may be important to both legal and ethical approval of future CRISPR genetic modifications. Any edits made to germ cells will potentially be passed on to offspring during meiosis. However, gene edits created in somatic cells will not be inherited by offspring.

Somatic gene editing is currently being tested in clinical trials, such as for treatment of sickle cell disease.

Human germline gene editing has only been reported by a Chinese scientist who performed gene editing on embryos to create a mutation that is associated with resistance to HIV in 2018.

Questions surround human gene editing

The ability to do something may not necessarily mean that it should be used without restriction. All people, non-scientists and scientists, are considering what CRISPR means for human treatments and gene editing in the future.

A report was released in 2020 with a number of recommendations regarding Heritable Human Gene Editing. You can find them at this link to the National Academies of Sciences, Engineering and Medicine (<https://www.nationalacademies.org/our-work/international-commission-on-the-clinical-use-of-human-germline-genome-editing>).

References

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007 Mar 23;315(5819):1709-12. doi: 10.1126/science.1138140. PMID: 17379808.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012 Aug 17;337(6096):816-21. doi: 10.1126/science.1225829. Epub 2012 Jun 28. PMID: 22745249; PMCID: PMC6286148.

Thurtle-Schmidt DM, Lo TW. Molecular biology at the cutting edge: A review on CRISPR/CAS9 gene editing for undergraduates. *Biochem Mol Biol Educ*. 2018 Mar;46(2):195-205. doi: 10.1002/bmb.21108. Epub 2018 Jan 30. PMID: 29381252; PMCID: PMC5901406.

Answers

Exercise 4.1.2.1

6, 5, 2, 4, 1, 3

This page titled [4.1.2: CRISPR](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.1.2: CRISPR](#) by [Stefanie West Leacock](#) has no license indicated.

4.2: Mendelian Genetics

Learning Objectives

- Define and identify examples of: homozygous, heterozygous, allele, gene, locus, dominant, recessive, genotype, phenotype.
- Construct genetic crosses for P₀, F₁, and F₂ generations and predict the genotypes and phenotypes of offspring.
- Explain Mendel's laws of segregation and independent assortment, and how they predict the 3:1 dominant-to-recessive phenotypic ratio among the F₂ of a monohybrid cross, or the 9:3:3:1 phenotypic ratio in a dihybrid cross, respectively. Relate the key events of meiosis that explain Mendel's first and second laws.
- Be able to draw chromosomes during meiosis with alleles labeled.
- Interpret phenotypic ratios of progeny in experimental organisms to infer how particular traits are inherited.
- Predict genotypic and phenotypic ratios or probabilities of outcomes among progeny of single factor and multifactor crosses using simple rules of probability (sum rule and product rule).
- Cite the most common molecular explanations for dominant and recessive alleles.

Introduction to Gregor Mendel and his Work

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](https://youtu.be/XgXLVaBjQqw)



Video 4.2.1: Shirley Tilghman describes Gregor Mendel's Famous Genetics Experiment. (CC BY NC XBio via Wonder Collaborative on youtube.com <https://youtu.be/XgXLVaBjQqw>)

Mendel Studied Visible Character Traits in Pea Plants

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. Recalling that genes contain information needed to make proteins, we now understand that alleles are differences in gene sequence. If these differences alter the production, structure, or function of the protein, an observable or measurable change in the organism may occur. For example, Mendel identified two forms of a gene for seed color: one allele gave green seeds and the other gave yellow seeds.

Seed		Flower	Pod		Stem	
Form	Color	Color	Form	Color	Place	Size
						
Round	Yellow	White	Full	Yellow	Axial pods, Flowers along	Long (6-7ft)
						
Wrinkled	Green	Purple	Constricted	Green	Terminal pods, Flowers top	Short $\frac{1}{2}$ -1ft
1	2	3	4	5	6	7

Figure 4.2.1: Seven traits Mendel studied in peas. (Wikipedia-Mariana Ruiz-PD)

Representing genes and alleles

Alleles are forms of genes, if genes are DNA sequences, then alleles are variations in the sequence of a gene.

In genetics, you may encounter different ways of representing alleles. Traditionally, genes are represented as letters when studying Mendelian genetics. This representation is usually easy to follow in problems involving crosses; however, using letters does not reflect our modern understanding of the genetic differences between alleles, which often involves knowing whether or not a product is functional or how the allele was identified.

- A plus can be used to indicate that the gene product of an allele is functional.
- A minus can be used to indicate that the gene product is not functional.
- If sequence information is known, the nucleotide or amino acid change can be identified.
- In model organisms, alleles are often given numbers when they are identified as mutants and these numbers can be used to identify different alleles of the gene.

The table shows some examples of how we might represent genes and alleles.

Examples of ways genes can be represented	Examples of representing wild-type alleles	Examples of representing the mutant alleles
<i>C</i> representing a hypothetical gene	<i>C</i> or <i>C</i> ⁺	<i>c</i> or <i>C</i> ⁻ or <i>C</i> ¹
<i>white</i> or <i>w</i> gene in <i>Drosophila</i>	<i>white</i> or <i>white</i> ⁺ or <i>w</i> ⁺	<i>w</i> ⁻ or <i>w</i> ¹¹¹⁸
<i>ced-1</i> gene in <i>C. elegans</i>	<i>ced-1</i> or <i>ced-1</i> (+) or <i>ced-1</i> ⁺	<i>ced-1</i> (-) or <i>ced-1</i> (e1754)
<i>CFTR</i> gene in humans <i>Cftr</i> gene in mouse	<i>CFTR</i> or <i>CFTR</i> ⁺	<i>CFTR</i> (delF508) or <i>CFTR</i> (482G-A)

Heterozygous and homozygous

Mendel's work and discoveries are especially remarkable because he made his observations and conclusions (in 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**.

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 because all produce 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**.

wild type vs wild-type

The noun **wild type** (without a hyphen) means the the most common form of a gene, phenotype, or organism under standard conditions.

- Example: The mutant worms produce fewer offspring than **wild type**.

The adjective **wild-type** (with a hyphen) can be used to describe a gene, allele, organism, phenotype, or trait as what is most commonly present among a population or under standard conditions.

- Example: A **wild-type** worm produces about 300 offspring.

Contributors and Attributions

- [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 [4.2: Mendelian Genetics](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.2: Mendelian Genetics](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

4.2.1: Monohybrid Crosses and Segregation

True Breeding Lines

Mendel used **true-breeding** lines of pea plants, which are in-bred populations of plants or animals in which all parents and their offspring (over many generations) have the same phenotypes with respect to a particular trait. True breeding lines are useful, because they are typically assumed to be homozygous for the alleles that affect the trait of interest. When two individuals that are homozygous for the same alleles are crossed, all of their offspring will all also be homozygous. The continuation of such crosses constitutes a true breeding line or strain. A large variety of different strains, each with a different, true breeding character, can be collected and maintained for genetic research.

Monohybrid Crosses

A **monohybrid cross** is one in which both parents are heterozygous (or a hybrid) for a single (mono) trait. The trait might be petal color in pea plants. When conducting crosses, the first generation is called P (or P_0), the second generation is F_1 (F is for filial), and the next generation is F_2 .

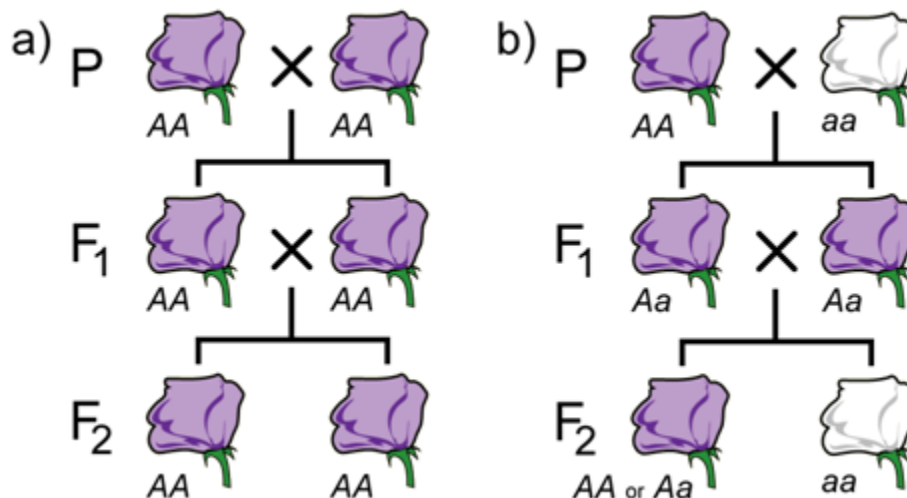


Figure 4.2.1.1: (a) A true-breeding line (b) A monohybrid cross produced by mating two different pure-breeding lines. (Original-Deholos-CC:AN)

Using monohybrid crosses, Mendel observed that although different alleles could influence a single trait, they remained indivisible and could be inherited separately. Additionally, the allele could be present but invisible in one generation, only to reappear in the next generation.

Definition: Mendel's First Law

The Law of Segregation states that during gamete formation, the two alleles at a gene locus segregate from each other; each gamete has an equal probability of containing either allele.

Punnett Squares

Given the genotypes of any two parents, we can predict the genotypes of gametes that will be produced during meiosis. Using that information, 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 genotype and phenotype 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. Punnett squares can also be used to calculate the frequency of types offspring that are expected.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Figure 4.2.1.1: Top: A Punnett square shows the possible offspring of a monohybrid cross between Pp and Pp parents. Bottom Video: How to set up and complete the square (Copyright CC BY Leacock)

Query 4.2.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



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 homozygote based on phenotype alone. 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, because you know the genotype of the plant is homozygous recessive aa. Depending on the genotype of the purple-flowered parent, you will observe different phenotypic ratios in the F₁ generation. If the purple-flowered parent was a homozygote, all of the F₁ progeny will be purple. If the purple-flowered parent was a heterozygote, the F₁ progeny should segregate purple-flowered and white-flowered plants in a 1:1 ratio.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 4.2.1.2: How to use a test cross (Copyright CC BY Leacock)

Query 4.2.1.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Contributors and Attributions

- [Dr. Todd Nickle and Isabelle Barrette-Ng](#) (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.
- Stefanie Leacock, UA-Little Rock

Molecular basis of dominant and recessive alleles

What determines whether alleles are dominant or recessive?

The alleles have different DNA sequences. Because the sequence of DNA contains information to make products, different sequences can lead to different products. One advantage of diploid species is that there are two copies of every sequence. If one sequence makes a "faulty" or non-functional product, it would be called a **loss-of-function** allele. However, there is likely another sequence that produces a "correct" or functional product. For most genes a single wild-type (usually the normal / functional) allele is capable of producing enough product for the cell, resulting in a **dominant** phenotype. However, if both copies of the gene are loss-of-function alleles, there will not be any functional protein and the **recessive** phenotype will be observed.

An example from pea plants

Mendel studied pea plants in which the peas could be round or wrinkled, with wrinkled being the recessive characteristic. The gene that determines this trait (originally represented as R) has since been identified and named SBE1 (for starch-branching enzyme). The wild-type sequence encodes an enzyme that catalyzes a chemical reaction of carbohydrates to form branched chains of starch in plants. When SBE1 protein is present and functional, the peas produce branched starch and exhibit a round shape. When SBE1 protein is absent, the amount of branched starch is reduced, but the levels of disaccharides are higher, resulting in increased water absorption. Later, this water will be lost and the pea will become wrinkled.

Molecular studies of DNA in round (*RR*) and wrinkled (*rr*) plants revealed an insertion of about 800 base pairs in an exon of the SBE1 gene in the *rr* plants (Bhattacharyya et al, 1990). The insertion is derived from a transposon and disrupts the SBE1 protein-

coding sequence. Therefore, the *RR* plants have two copies of a functional SBE1 allele (*SBE1+/+*) to make functional SBE1 enzyme, while *rr* plants have two copies of the SBE1 gene in which the sequence is disrupted (*SBE1-/-*) and cannot produce any functional SBE1 enzyme.

So what happens in a heterozygous *Rr* (*SBE1+/-*) plant? These plants have one chromosome with the *SBE1+* allele and one chromosome with the *SBE1-* allele. The *SBE1-* allele is still transcribed, but does not code for the functional protein during translation. The *SBE1+* allele is transcribed and translated into a functional enzyme, starch-branching occurs, and the peas have the round phenotype. Although the total amount of SBE1 enzyme is less than homozygous dominant peas, it is sufficient for the round phenotype and therefore dominant. In this case, molecular discoveries 125 years after Mendel's work reveal the reason for the dominance of the *R* allele over the *r* allele.

References

Bhattacharyya MK, Smith AM, Ellis TH, Hedley C, Martin C. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*. 1990 Jan 12;60(1):115-22. doi: 10.1016/0092-8674(90)90721-p. PMID: 2153053.

This page titled [4.2.1: Monohybrid Crosses and Segregation](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.2.1: Monohybrid Crosses and Segregation](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

4.2.2: Dihybrid Crosses and Independent Assortment

Mendel's Second Law

Before Mendel, it had not yet been established that heritable traits were controlled by discrete factors. Therefore an important question was whether distinct traits were controlled by separate factors that were inherited independently of one another. 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. When either of these traits was studied individually, the phenotypes segregated in the classical 3:1 ratio among the progeny of a monohybrid cross, with $\frac{3}{4}$ of the seeds green and $\frac{1}{4}$ yellow in one cross, and $\frac{3}{4}$ round and $\frac{1}{4}$ wrinkled in the other cross. Would this be true when both were in the same individual?

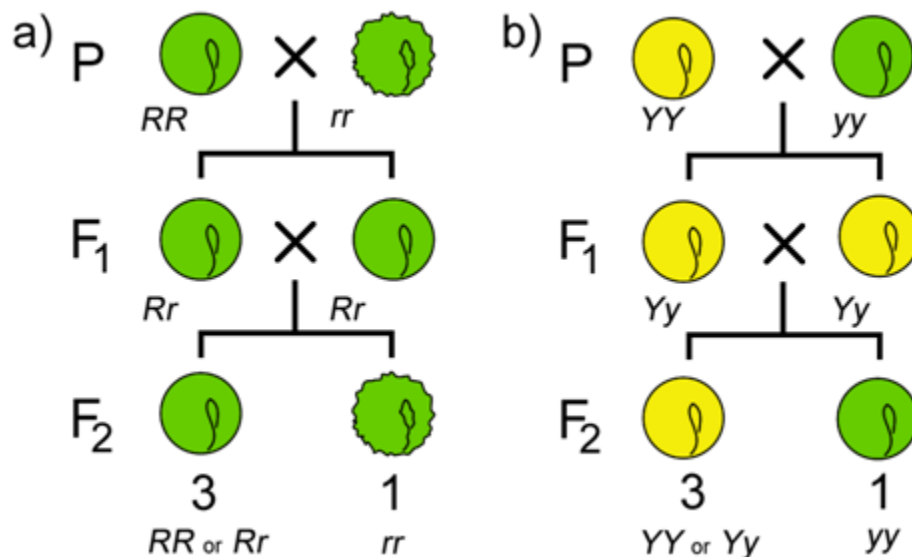


Figure 4.2.2.1: Monohybrid crosses involving two distinct traits, shape which can be round or wrinkled and color which can be yellow or green.
a) A monohybrid cross between RR (round peas) and rr (wrinkled peas) plants and b) A monohybrid cross between YY (yellow peas) and yy (green peas) plants. (Original-Deyholos-CC:AN)

To analyze the segregation of both traits at the same time in the same individual, Mendel 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 yellow and round, and which were also **dihybrids**; they carried two alleles at each of two loci. If the alleles for the two genes for pea shape and pea color cannot be separated from each other, then in the F_2 generation, the offspring should be only green, round pea plants or yellow, wrinkled plants, like the P generation plants.

If the genes controlling shape and color can be inherited independently, then what is the probability of phenotypes in the F_2 generation? Using the product rule, we can multiply the individual probabilities of obtaining a round phenotype ($\frac{3}{4}$) with the probability of obtaining a yellow phenotype ($\frac{3}{4}$), then $\frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$ of the progeny would be both round and green. Likewise, $\frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$ of the progeny would be both round and yellow, and so on. By applying the product rule to all 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.

The probability of two or more independent events occurring is the product obtained by multiplying the probabilities of the individual events.

If you can use the word "and" to describe the occurrence of the events, then you can usually use the **product rule**.

For example, the probability of an organism with

phenotype
1
and
phenotype
2
is
the
product
of
the
probability
of
phenotype
1
times
the
probability
of
phenotype
2.

The
dihybrid
crosses
that
Mendel
performed
consistently
revealed
the
9:3:3:1
ratio
in
dihybrid
crosses,
leading
him
to
conclude
that
the
factors
controlling
the
traits
are
inherited
independent
of
one
another,
a
rule
commonly
known
as
the

Law
of
Independent
Assortment.
Importantly,
the
 F_2
generation
includes
two
phenotype
classes
(in
this
example,
round,
green
peas
and
yellow,
wrinkled
peas)
that
are
new
combinations
of
phenotypes,
distinct
from
the
parental
generation
plants.
Thus,
independent
assortment
contribute
to
genetic
diversity.

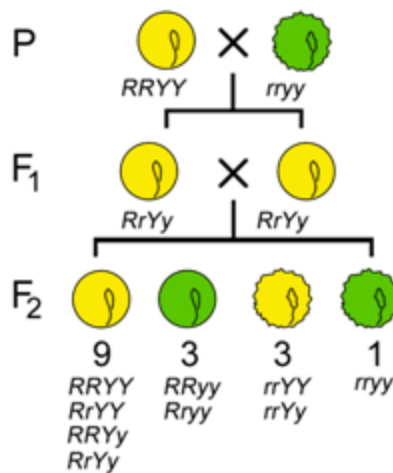


Figure 4.2.2.2: Pure-breeding lines *RRYY* and *rryy* are crossed to produce dihybrids in the *F*₁ generation. The cross of the *F*₁ *RrYy* dihybrids produces four phenotypic classes in a ratio of 9:3:3:1. (Original-Deyholos-CC:AN)

Definition: Mendel's Second Law

The **Law of Independent Assortment** states that two loci assort independently of each other during gamete formation.

Table 4.2.2.1: Phenotypic classes expected in monohybrid and dihybrid crosses for two seed traits in pea.

Frequency of phenotypic crosses within separate monohybrid crosses:

seed shape: $\frac{3}{4}$ round $\frac{1}{4}$ wrinkled

seed color: $\frac{3}{4}$ yellow $\frac{1}{4}$ green

Frequency of phenotypic crosses within a dihybrid cross:

$\frac{3}{4}$ round × $\frac{3}{4}$ yellow = $\frac{9}{16}$ round & yellow

$\frac{3}{4}$ round × $\frac{1}{4}$ green = $\frac{3}{16}$ round & green

$\frac{1}{4}$ wrinkled × $\frac{3}{4}$ yellow = $\frac{3}{16}$ wrinkled & yellow

$\frac{1}{4}$ wrinkled × $\frac{1}{4}$ green = $\frac{1}{16}$ wrinkled & green

The 9:3:3:1 phenotypic ratio calculated using the product rule can also be obtained using Punnett Square. 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.

	<i>RY</i>	<i>Ry</i>	<i>rY</i>	<i>ry</i>
<i>RY</i>	<i>RRYY</i>	<i>RRYy</i>	<i>RrYY</i>	<i>RrYy</i>
<i>Ry</i>	<i>RRYy</i>	<i>RRyy</i>	<i>RrYy</i>	<i>Rryy</i>
<i>rY</i>	<i>RrYY</i>	<i>RrYy</i>	<i>rrYY</i>	<i>rrYy</i>
<i>ry</i>	<i>RrYy</i>	<i>Rryy</i>	<i>rrYy</i>	<i>rryy</i>

Figure 4.2.2.3: A Punnett Square showing the results of a dihybrid cross. 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. 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.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 4.2.2.1: Watch the video to see how to set up and complete a dihybrid Punnett 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 $RrYy \times RrYy$. 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.

Query 4.2.2.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Applying the product rule

Mendel's conclusions about the segregation of alleles and independent of assortment of genes continue to hold true for inheritance in diploid organisms, in which meiosis produces gametes that have one copy of thousands genes - not just one or two. However, making a Punnett Square for more than two genes becomes tricky and cumbersome. The Product Rule can be used

to predict outcomes when considering more than two genes at a time. Find the probability of each genotype or phenotype and multiple each probability.

Exercise 4.2.2.1

If two organisms with genotype $GgHhJi$ are crossed, what percent of offspring are expected to be homozygous dominant for all three genes?

Hint: GG and HH and JJ

Answer available at the end of this section.

Contributors and Attributions

- [Dr. Todd Nickle and Isabelle Barrette-Ng](#) (Mount Royal University) The content on this page is licensed under CC SA 3.0 licensing guidelines.
- Stefanie Leacock, UA-Little Rock

Answers

Exercise 4.2.2.1

(Probability GG : $1/4$) $\times$ (Probability HH : $1/4$) $\times$ (Probability JJ : $1/4$) = $1/64$

This page titled [4.2.2: Dihybrid Crosses and Independent Assortment](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.2.2: Dihybrid Crosses and Independent Assortment](#) by Stefanie West Leacock is licensed [CC BY-SA 4.0](#).

4.3: Pedigrees

Learning Objectives

- Draw a pedigree based on information in a story problem.
- Using pedigrees, distinguish between dominant and recessive autosomal inheritance.

When working with plants, animals, or humans, pedigrees can be used to follow traits among family groups.

Pedigrees 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. 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**. Historically, a standardized set of symbols has been used to represent an individual's sex (circle for female, square for male), family relationships (lines between individuals) and phenotype (unfilled or filled symbol). Matings that represent a union of gametes (not necessarily marriage or other relationships) are drawn as a line joining a two symbols, while a consanguineous mating (a mating between closely related individuals) is two lines.

A modern understanding of assisted reproduction, complex family relationships, disorders of sex determination (DSD), and gender identity requires us to consider inclusive symbols and notations. When gender is unknown, convention has dictated that a diamond be used. The diamond has also been suggested as substitute symbol to indicate a transgender or non-binary individual. However, there is not yet an official standard for indicating trans or non-binary persons; a recent survey of genetic counselors presents several alternatives (Sheehan et al., 2020). These possibilities including a square with a circle inside for a trans male or circle with a square inside for a trans female, the appropriate symbol for the persons gender identify with text or abbreviations can be used to provide additional information, for example trans male, trans female, FTM (female to male), or FTNB (female to non-binary). A circle or square symbol with the gender identify could be used with the sex chromosomes indicated could also be used. Updates and standardization of these symbols are likely to continue to occur. Other circumstances that might need to be noted are assisted reproduction, including sperm or egg donors as well as surrogates, as these may include individuals who contribute genes to the offspring, but may not have a role in the family or family group.

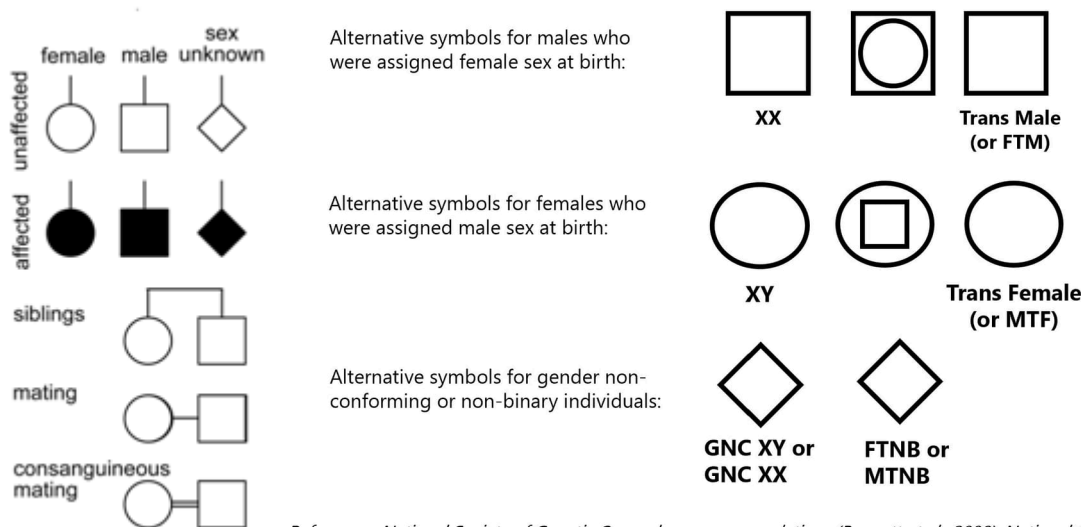


Figure 4.3.1: Left: Traditional symbols used in drawing a pedigree are circles for females and squares for males (Copyright: Original-Deyholos-CC:AN). Right: Possible approaches for depicting trans people in pedigree charts compiled from a survey of genetic counseling professionals (Sheehan et al., 2019) as well as non-official proposals to present diversity in biology classes (Hale, 2020).

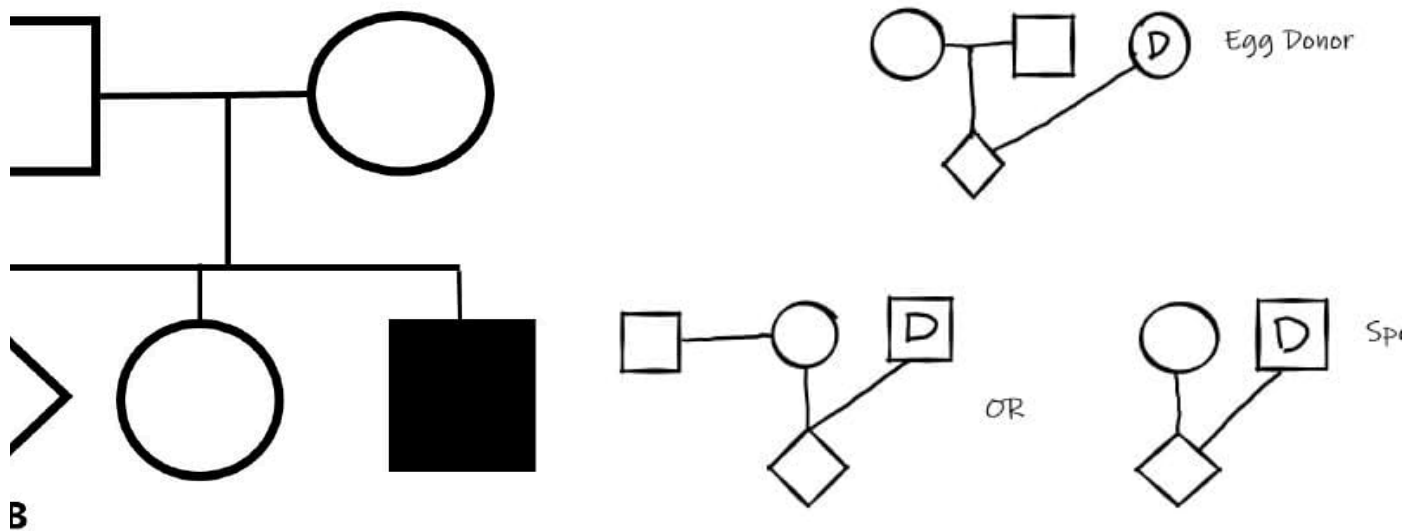


Figure 4.3.2: Left: Example pedigree in which an unaffected male and female have three offspring, one female to non-binary, one female, and one male who is affected with a disease or phenotype. Right: Example designation of egg or sperm donors as a symbol with a D for donor connected by a diagonal line to the offspring.

Each pedigree represents all of the available information about the inheritance of a single trait (most often a disease) within a family. If an individual is known to have symptoms of the disease or trait (**affected**), the symbol is filled in. Sometimes a half-filled in symbol is used to indicate a known **carrier** of a disease; carriers are unaffected but can pass an allele to subsequent generations because they are heterozygous. Note that when a pedigree is constructed, it is often unknown whether a particular individual is a carrier or not, so carriers are not always explicitly indicated in a pedigree. Unless additional information is provided, assume that the pedigrees presented in this section are accurate, and represent fully penetrant traits.

The pedigree is 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. 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.

References

- Bennett, R.L., French, K.S., Resta, R.G. and Doyle, D.L. (2008), Standardized Human Pedigree Nomenclature: Update and Assessment of the Recommendations of the National Society of Genetic Counselors. *J Genet Counsel*, 17: 424-433. <https://doi.org/10.1007/s10897-008-9169-9>
- Hales KG. Signaling Inclusivity in Undergraduate Biology Courses through Deliberate Framing of Genetics Topics Relevant to Gender Identity, Disability, and Race. *CBE Life Sci Educ*. 2020 Jun;19(2):es2. doi: 10.1187/cbe.19-08-0156. PMID: 32357097.
- Sheehan E, Bennett RL, Harris M, Chan-Smutko G. Assessing transgender and gender non-conforming pedigree nomenclature in current genetic counselors' practice: The case for geometric inclusivity. *J Genet Couns*. 2020 Dec;29(6):1114-1125. doi: 10.1002/jgc.4.1256. Epub 2020 Mar 30. PMID: 32232917.

Using pedigrees to visualize patterns 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.

Important to remember about autosomes & allosomes (sex chromosomes)

- Remember that in humans the autosomes are the same in both sexes. Both sperm and oocytes contain one copy of each autosome.
- Remember that sex chromosomes may differ between individuals of different sexes. Both sperm and oocytes normally contain one copy of a sex chromosome. In oocytes this chromosome is almost always an X; in sperm, it could be X or Y.

Autosomal Dominant Inheritance

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 and the trait often appears in every generation; this pattern is often called a vertical inheritance pattern, as in from top to bottom of the pedigree. A pedigree with affected individuals in every generation is typical of autosomal dominant 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 autosomal dominant 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 autosomal dominant diseases are usually rare in populations, therefore affected individuals with autosomal dominant 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 autosomal dominant mode of inheritance.

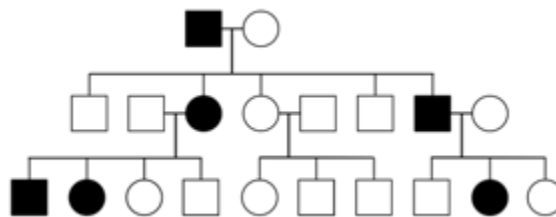


Figure 4.3.3: A pedigree consistent with autosomal dominant inheritance. (Original-Deyholos-CC:AN)

Autosomal Recessive Inheritance

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 autosomal recessive traits, many individuals in a pedigree can be carriers, probably without knowing it. Compared to pedigrees of dominant traits, autosomal recessive pedigrees tend to show fewer affected individuals and are often described as "skipping" generations. Thus, the major feature that distinguishes autosomal recessive from dominantly inherited traits is that unaffected individuals can have affected offspring.

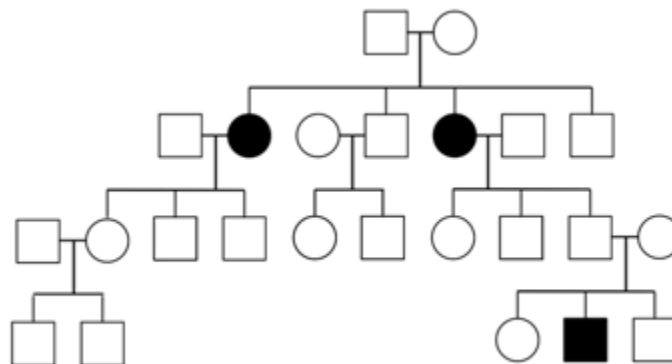


Figure 4.3.4: A pedigree consistent with autosomal recessive inheritance. (Original-Deyholos-CC:AN)

Contributors and Attributions

- [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 [4.3: Pedigrees](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.3: Pedigrees](#) by [Stefanie West Leacock](#) has no license indicated.

4.4: Exceptions to autosomal inheritance

Learning Objectives

- Compare different modes of sex determination, such as mammalian XY, fruit flies, insects, or environmental.
- Explain why dosage compensation is needed and the role of the non-coding RNA *Xist* in mammalian dosage compensation.
- Predict the number of Barr bodies in a particular cell.
- Predict offspring phenotypes for X-linked genes.

For loci on autosomes, in which all individuals have two copies of the chromosome and either copy is equally likely to be passed to offspring, the alleles follow the normal Mendelian pattern of inheritance. However, for loci on the allosomes (also called sex chromosomes) individuals may have zero, one, or two copies of a chromosome. Because homologous chromosomes contain the same genes, individuals may have zero, one, or two copies of some genes. In humans, the allosomes are the X and Y chromosomes, which contain different genes, even though they act as a homologous pair during meiosis. Instead, genes on the X chromosome will follow an X-linked pattern of inheritance, while genes on the Y chromosome will follow a Y-linked pattern of inheritance.

X-Linked inheritance example: 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.

Note about nomenclature

In genetics, genes are often named after their mutant (frequently, but not always, loss-of-function) phenotypes. Remember, genetics has its origin in observing phenotypes. The *white* allele described here was identified in 1910 (Morgan, T.H.) long before gene sequences could be determined, in fact, even before DNA had been identified as the unit of heredity. The DNA sequence of this locus was not published until 1984 (O'Hare et al). We know now that this gene encodes a transporter protein that deposits pigment and that the mutant allele does not produce a functional transporter, but the gene name *white* continues to be used.

The process of identifying mutant phenotypes and then later determining which gene is involved is called **forward genetics**.

The alternative, to identify a candidate gene and manipulate it to observe the phenotype, is called **reverse genetics**, and was only made possible by the advent of modern molecular biology tools.

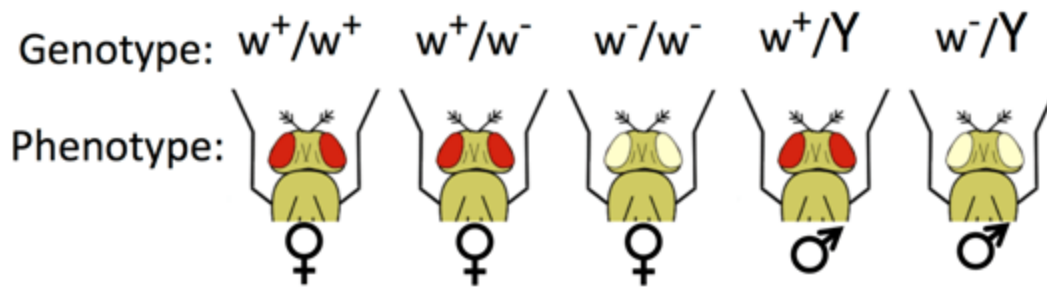


Figure 4.4.1: Relationship between genotype and phenotype for the white gene on the X-linked gene in *Drosophila melanogaster*. The Y chromosome is indicated with a capital Y because it does not have a copy of the *white* gene. (Original-Harrington/Locke-CC:AN)

A researcher may not know beforehand whether a novel mutation is sex-linked. The definitive method to test for sex-linkage is **reciprocal crosses**, in which a male and a female that have different phenotypes are crossed. In a second cross, the phenotypes are reversed relative to the sex of the parents in the first cross. 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 are expected to show the same genotypes and phenotypes 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).

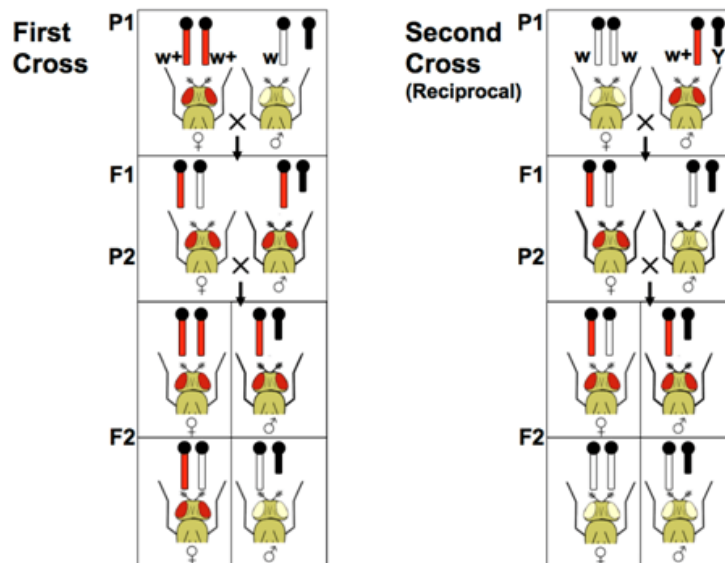


Figure 4.4.2: Reciprocal crosses involving an X-linked gene in *Drosophila melanogaster*. In a cross (left) between a w^+w^+ red-eyed female fly and a wY white-eyed male fly, all of the offspring in the F_1 have red eyes. The F_2 offspring are all red-eyed females and half red-eyed males. In a reciprocal cross (right) between a ww white-eyed female fly and a w^+Y red-eyed male fly, all of the female F_1 offspring have red eyes and the male F_1 offspring all have white eyes. The F_2 offspring are half red-eyed males and females. Thomas H. 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)

Diversity in Sex Determination

Sex determination refers to the process by which reproductive organs are specified to become ovaries or testes. Mechanisms for sex determination in plants and animals are remarkably diverse, and include sex chromosomes, chromosome dosage, and environment among others.

Some organisms used chromosomal sex determination. For example in humans and other mammals XY embryos commonly

develop testes while XX embryos develop ovaries. This difference in development is usually initiated by the presence of a single gene, the *SRY* gene (sex-determining region on Y), on the Y-chromosome. *SRY* encodes a transcription factor that activates expression of genes that promote testes differentiation. Without *SRY*, these genes are not activated and genes that promote ovary development are transcribed. Notably, *SRY* initiates the process of testes development, but a series of genes are required to complete testes and reproductive organ development. If the downstream genes are not fully functional, development of these structures may not occur despite the presence of the Y chromosome and *SRY*.

Query 4.4.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Although *Drosophila melanogaster* also has a Y chromosome, there is not an *SRY* gene and the X to Autosome (X:A) ratio determines male or female differentiation. Individuals with two autosome sets and two X chromosomes (2A:2X) will develop ovaries, while those with only one X chromosome (2A:1X) will develop testes. The presence/absence of the Y-chromosome and its genes are not significant. For example, both an XO and XY fruit fly will develop testes and other male phenotypes, if they both have two sets of autosomes.

In other animal species, the number of chromosome sets can determine sex. For example the haploid-diploid system is used in bees, ants, and wasps. Typically haploid individuals, which develop from unfertilized eggs, develop testes, while diploid individuals, which develop from fertilized eggs, form ovaries. For this system to work, meiosis does not occur during sperm production in haploids; instead sperm are identical copies produced by mitosis because chromosome number does not need to be reduced by half.

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

Dosage Compensation

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. 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. Comparison of X and Y chromosomes has led to the view that Y chromosomes have evolved from autosomes that degenerated, slowly mutating and losing loci. In modern day mammals, the Y chromosomes have very few genes, 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.

Human X and Y Compared

	X chromosome	Y chromosome
Number of base pairs	156,040,895	57,227,415
Number of coding genes	857	64
Contains essential genes for viability?	Yes	No

Chromosomal sex determination requires dealing with the gene dosage difference between individuals with different chromosome complements: e.g. XX females have two copies ("doses") of X-chromosome genes while XY males only have one copy of these genes. For all the autosomes, XX and XY individuals have the same number of copies of genes. Producing too much of some proteins can be problematic, even lethal, for organisms. This difference in gene dosage needs to be compensated in a process called **dosage compensation**. In *Drosophila* and many other insects, individuals with single X chromosome express the genes on it at twice the normal rate. This mechanism of dosage compensation restores a balance of proteins encoded by X-linked genes and those made by autosomal genes.

X-chromosome Inactivation in Mammals

In mammals the dosage compensation system operates in any cell with more than one X chromosome. In XX embryos, one X in each cell is randomly chosen and marked for inactivation by expressing a long non-coding RNA *Xist* (X inactive specific transcript). The *Xist* RNA stays in the nucleus and "coats" the X chromosome, recruiting proteins that affect chromosome packing. This X chromosome becomes highly compact heterochromatin. The expression of *Xist* is repressed on the other chromosome, so it remains more loosely packed euchromatin. 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 and the chemical modifications that regulate chromatin are passed on following each round of cell division.

Make a prediction

What would happen to XY or XX mouse embryos that lack the *Xist* gene?

Answer available at the end of this section.

From this point forward this chromosome will be inactive, abbreviated X_{inactive} (X_i). The chromosome appears as a condensed mass within interphase nuclei called the **Barr body**. With the inactivation of genes on one X chromosome, XX individuals have the same number of functioning X-linked genes as XY individuals. The other X chromosome, the X_{active} (X_a), is unaffected. The mechanisms that determine X inactivation can also be used in other situations, such as XXX, in which cells inactivate two X chromosomes, or XXY individuals, in which cells inactivate one X chromosome. This remarkable system always leaves one X chromosome active, such that every mammalian cell expresses the correct dose of X chromosome genes.

Counting chromosomes

The exact mechanism cells use to "count" their X chromosomes is not entirely clear. Essentially all mammalian cells leave only one X chromosome active, even when abnormal numbers of X chromosomes are present.

Examples:

- Females with an extra X (XXX genotype) have 2 Barr bodies and 1 active X.
- Males with an extra X (XXY) have 1 Barr body and 1 active X.

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. 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. 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.

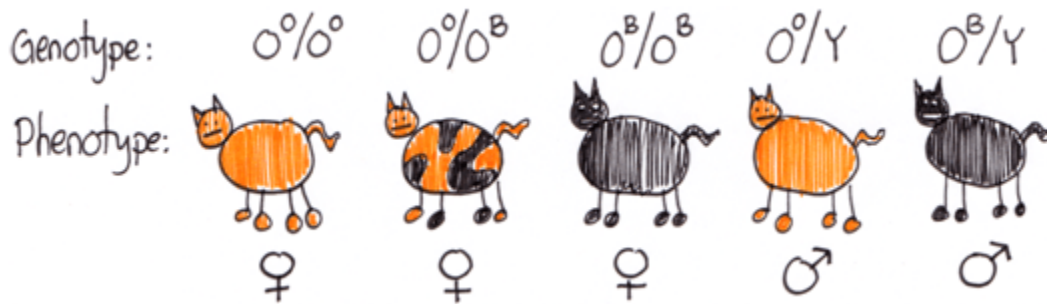


Figure 4.4.3: Relationship between genotype and phenotype for an X-linked gene in cats. The O^O allele = orange while the O^B allele = black. (Original-Harrington-CC:AN)

Another example is the *F8* gene in humans that encodes Factor VIII, a protein needed for blood clotting that is produced in liver cells. If an XY individual is hemizygous for a mutant allele ($F8^-/F8^-$) the result is hemophilia type A. An XX individual homozygous for mutant alleles will also have hemophilia. Heterozygous $F8^+/F8^-$ individuals 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. 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.

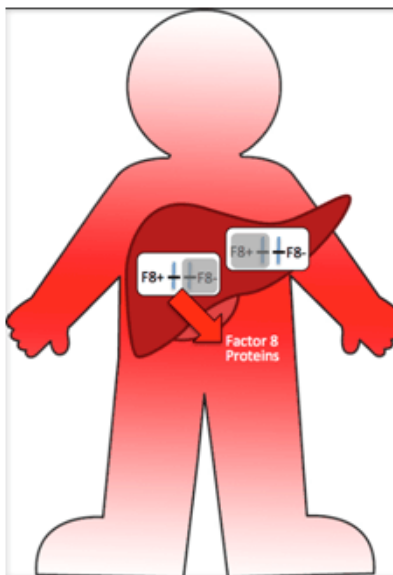


Figure 4.4.4: Females heterozygous for an *F8* mutation have a mixture of liver cells in which the $F8^+$ or $F8^-$ chromosome is inactivated. Because people with the $F8^+/F8^-$ genotype have the same phenotype, normal blood clotting, as $F8^+/F8^+$ people the $F8^-$ mutation is classified as recessive. (Original-Harrington/Locke-CC:AN)

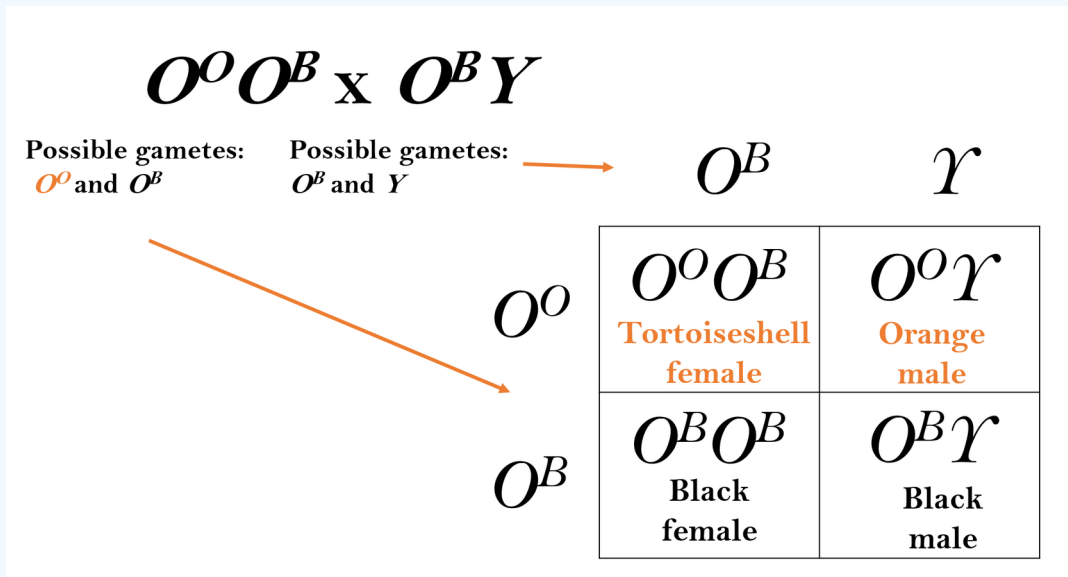
Example 4.4.1

Predicting genetic crosses with X-linked genes follows that of autosomal genes, remembering that each gamete contains only one sex chromosome, unless otherwise specified.

A tortoiseshell female cat and a black male cat have 8 offspring. How many are expected to be tortoiseshell?

Solution

Using a Punnett square reveals that 25% (or two) offspring are predicted to be tortoiseshell.



Exercise 4.4.1

If you were able to clone a tortoiseshell cat, would the pattern of orange and black markings be identical in the clone? Why or why not?

Answer available at the end of this section.

Exercise 4.4.2

A heterozygous individual (genotype $F8^+/F8^-$) is a carrier for the hemophilia allele. Her partner is an XY male with normal blood clotting. What is the probability their son will have hemophilia? What about a daughter?

Answer available at the end of this section.

Z-linked genes

The ZW system is distinguished from the XY system based on which sex has two different sex chromosomes, also called the heterogametic sex. In this case, male birds commonly have ZZ chromosomes, whereas female birds commonly have ZW chromosomes. Thus, female birds may have only one copy of Z-linked genes. For example, a Z-linked gene influences feather color 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 4.4.5). Only turkeys with two Z chromosomes, commonly males, can be heterozygous for this locus. Heterozygous turkeys (Ee) are 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.

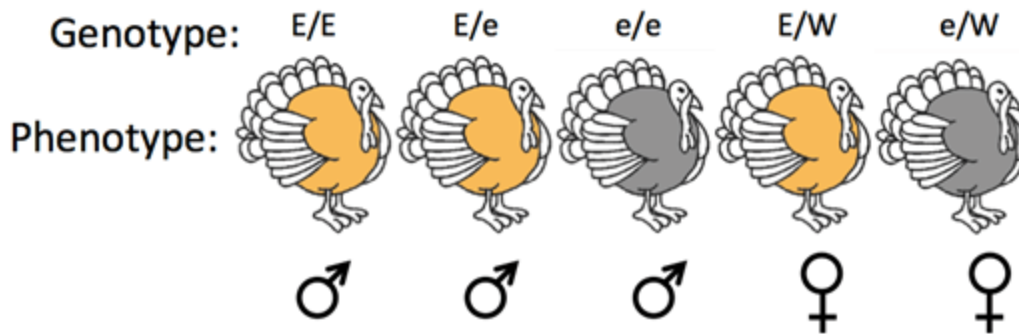


Figure 4.4.5: Relationship between genotype and phenotype for a Z-linked gene in turkeys. The W chromosome does not have an E/e-gene so it is just indicated with a capital W. (Original-Harrington/Locke-CC:AN)

Contributors and Attributions

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

References

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, et al. (2014) Sex Determination: Why So Many Ways of Doing It? PLoS Biol 12(7): e1001899. <https://doi.org/10.1371/journal.pbio.1001899>

Answers

Make a prediction

XY mouse embryos with a deletion of the *Xist* gene were viable and fully fertile. The XX mouse embryos died during embryonic development. (Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 1997 Jan 15;11(2):156-66. doi: 10.1101/gad.11.2.156. PMID: 9009199.)

Exercise 4.4.1

The pattern is not likely to be identical because X chromosome inactivation happens during embryogenesis. When the cloned cat is in the early stages of development, the X chromosomes with the orange and black alleles will be randomly inactivated and passed on during each mitosis as more cells are formed.

Exercise 4.4.2

A son will be either $F8^{+}Y$ or $F8^{-}Y$ (50% with hemophilia). A daughter will be either $F8^{+}/F8^{-}$ or $F8^{+}/F8^{+}$ (0% with hemophilia).

This page titled 4.4: Exceptions to autosomal inheritance is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stefanie West Leacock.

- 4.4: Exceptions to autosomal inheritance by Stefanie West Leacock is licensed CC BY-SA 4.0.

4.4.1: Inheritance patterns for X-linked and Y-linked genes

X-linked dominant inheritance

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 XX individuals, who tend to be more frequently affected than XY individuals in the population because they have two copies of the X chromosome. However, not all pedigrees provide sufficient information to distinguish X-linked and autosomal dominance. One definitive indication that a trait is inherited on an autosome, and not the X chromosome, is that an affected XY sperm parent passes the disease to an XY child; this type of transmission is not possible with X-linked dominance, because XY individuals inherit their X chromosome from the XX egg parent.

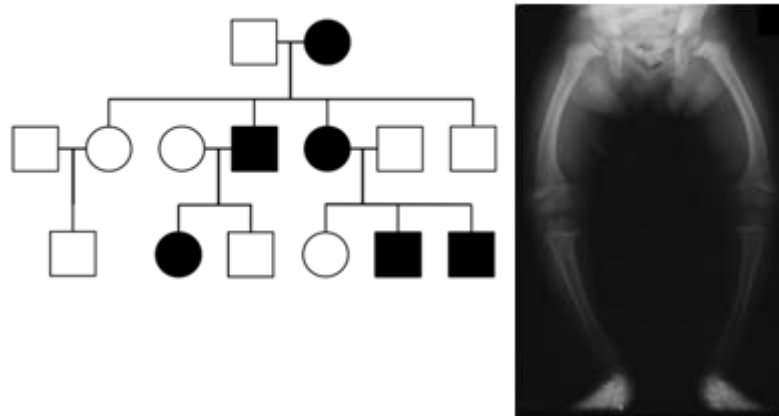


Figure 4.4.1.1: Left: A pedigree consistent with X-linked dominant inheritance. Right: X-ray of an individual with rickets, which can be caused by an X-linked gene. (Wikipedia-Mrsh-CC:AS)

X-linked recessive

Any XY individual that inherits an X-linked recessive disease allele will be affected by it (assuming complete penetrance), because they do not have a second copy of the X chromosome to provide a dominant allele. Therefore, in X-linked recessive inheritance, XY individuals, commonly males, tend to be affected more frequently than XX individuals, commonly females, in a population. In autosomal dominant and recessive traits, the sex of the individual does not change the probability of being affected, because all individuals have two copies of autosomes. 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 establish that an inheritance pattern is not X-linked recessive is the presence of an affected XX offspring from unaffected parents; because the sperm-parent would have a mutant X and therefore would also have been affected if the trait was X-linked recessive.

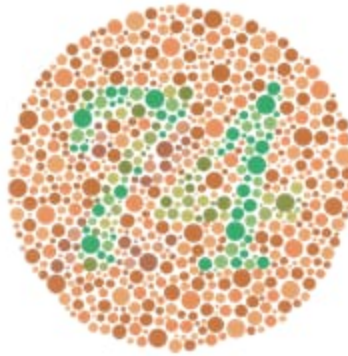


Figure 4.4.1.2: Some forms of colour blindness are inherited as X-linked recessive traits. Color blindness is diagnosed using tests such as this.
Ishihara Test. (Wikipedia-unknown-PD)

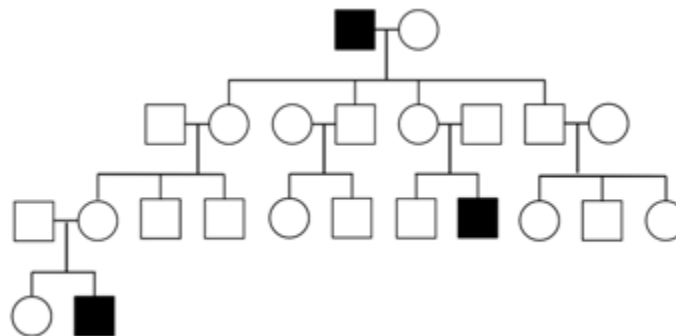


Figure 4.4.1.3\): A pedigree consistent with X-linked recessive inheritance. (Original-Deyholos-CC:AN)

Y-linked Inheritance

Only individuals with a Y chromosome are affected in mammalian Y-linked inheritance and transmission can only occur from a sperm-producing parent. 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. In 1960, the **hairy-ear-rim** phenotype seen in some Indian families was described as a Y-linked trait (Gates, RR <https://science.sciencemag.org/content/132/3420/145.1.long>) based on 20 pedigrees showing father to son transmission. In addition, the daughters of men with the hairy-ear-rim phenotype were unable to pass on the trait. However, questions remained about the accuracy of this conclusion and in 2004 a genotyping study found a lack of molecular evidence for Y linkage of this trait (Lee et al <https://www.nature.com/articles/5201271.pdf>).

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, or about 0.5% world wide (Zerjal et al, 2003).

Thinking about Y linkage

Aside from the size and small gene number on the Y chromosome, why are Y linked traits rare?

What functions do you think genes on the Y chromosome perform?

References

Lee AC, Kamalam A, Adams SM, Jobling MA. Molecular evidence for absence of Y-linkage of the Hairy Ears trait. *Eur J Hum Genet*. 2004 Dec;12(12):1077-9. doi: 10.1038/sj.ejhg.5201271. PMID: 15367914.

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhon T, Dashnyam B, Mehdi SQ, Tyler-Smith C. The genetic legacy of the Mongols. *Am J Hum Genet*. 2003 Mar;72(3):717-21. doi: 10.1086/367774. Epub 2003 Jan 17. PMID: 12592608; PMCID: PMC1180246.

This page titled [4.4.1: Inheritance patterns for X-linked and Y-linked genes](#) is shared under a license and was authored, remixed, and/or curated by Cheneir Neely.

- [4.4.1: Inheritance patterns for X-linked and Y-linked genes](#) has no license indicated.

4.4.2: Organellar Inheritance

Learning Objectives

- Explain why genetic information in organelles is passed independently of nuclear DNA.

In eukaryotes, DNA and genes also exist outside of the nuclear chromosomes. Both the **chloroplast** and **mitochondrion** have circular chromosomes (Figure 4.4.2.1). These organellar genomes are often present in multiple copies within each organelle. In most sexually-reproducing species, **organellar chromosomes** are inherited from only one parent, usually the one that produces the largest gamete. Thus, in mammals, angiosperms, and many other organisms, mitochondria and chloroplasts are inherited only through the oocyte.

These organelles are likely the remnants of prokaryotic endosymbionts that entered the cytoplasm of ancient progenitors of today's eukaryotes (**endosymbiont theory**). These endosymbionts had their own, circular chromosomes, like most bacteria that exist today. Chloroplasts and mitochondria typically have circular chromosomes that behave more like bacterial chromosomes than eukaryotic chromosomes, i.e. these organellar genomes do not undergo mitosis or meiosis.

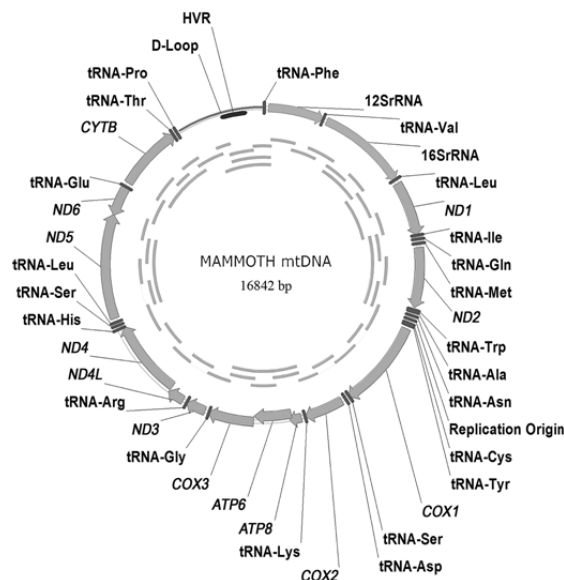


Figure 4.4.2.1: A map of the complete mitochondrial chromosome of the woolly mammoth (*Mammuthus primigenius*). The mtDNA that was used to produce this map was obtained from tissue of a mammoth that lived approximately 32,000 years ago. The map shows the position of enzymes encoded on the chromosome including components of the NADH dehydrogenase (ND) complex and cytochrome oxidases (COX), all of which function during energy metabolism in the mitochondrion. The mitochondrial chromosome also encodes various tRNAs and rRNAs used in translation of the genes encoded on this chromosome. Other proteins required by the mitochondrion are encoded in the nuclear genome, and are translated in the cytoplasm and imported into the organelle. (From Rogaeu et al, 2006). mtDNA work indicates that mammoths are more closely related to Indian elephants than to either of the African species (Rohland et al, 2010).

Implications of mitochondrial inheritance

As with nuclear DNA, organellar DNA can be mutated. Cells can have a mixture of hundreds to thousands of organelles with different alleles for genes. Because there are not simply one or two organelles in a cell, terms like heterozygous and homozygous do not apply to this situation and patterns of inheritance can be unpredictable. Some patterns of inheritance that are usually observed for mitochondrial inheritance are:

- Traits can be passed via egg to offspring
- Traits are not passed via sperm to offspring
- Variable penetrance and expressivity are often observed are often due to different proportions of wild type and mutant organelles in the organism of even differing proportions between different tissues in the same organism.

Mitochondrial DNA polymorphisms are also used to investigate evolutionary lineages, both ancient and recent.

Current Research in Plant Genetics:

Although organelles are most often inherited through oocytes, exceptions have been identified. Recent studies of cucumber plants (*Cucumis sativus* var. *sativus*) identified SNPs in true-breeding lines and performed reciprocal crosses. The results showed that the chloroplasts were inherited from the maternal parent but that mitochondria are inherited from the male parent.

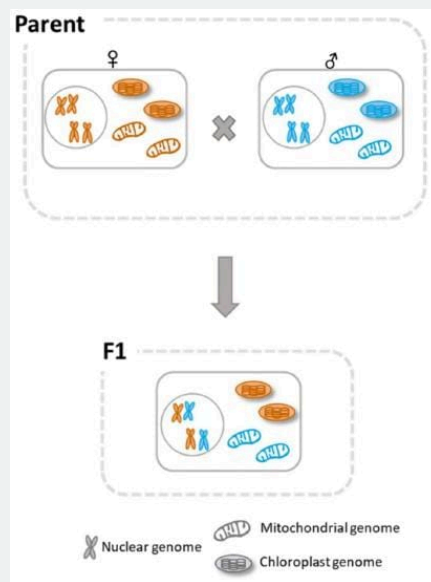


Figure 4.4.2.2: Inheritance of organelles and nuclear chromosomes in cucumbers. (CC BY (<http://creativecommons.org/licenses/by/4.0/>) Park et al. (2021) via <https://doi.org/10.1038/s41598-021-81988-w>)

Reference: Park, HS., Lee, W.K., Lee, SC. et al. Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F₁ hybrid combinations. *Sci Rep* **11**, 2506 (2021). <https://doi.org/10.1038/s41598-021-81988-w>

Contributors and Attributions

- 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 [4.4.2: Organellar Inheritance](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.4.2: Organellar Inheritance](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

4.5: Linkage

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 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 and current efforts to find loci that contribute to complex traits.

Contributors and Attributions

- [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 [4.5: Linkage](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.5: Linkage](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

4.5.1: Linkage and Mapping

Learning Objectives

- Understand that linked genes do not exhibit independent assortment because recombination will not always occur between the loci.
- Distinguish between parental and recombinant chromosomes, gametes, and offspring, and identify them in crosses.
- Calculate the map distance between loci given the phenotypes of offspring or predict phenotypes of offspring given the recombination frequency between loci. Use the distance to construct genetic maps based on data from two-point or three-point testcrosses.

Loci are locations of genes on chromosomes

In many genetic crosses involving one or two genes, the gene can be representing by a name or a letter. However, when considering linked genes the location of each gene and allele often needs to be represented. For example, a dihybrid $BbEe$ can have one chromosome with both dominant alleles (BE) or one chromosome with a dominant allele for one gene and recessive allele for the other (Be for example). The locations of the genes on the chromosomes are loci. For this section, remember that is the distance between loci that influences how often homologous recombination occurs between them between meiosis.

If loci B and E are on the same chromosome, a cell with genotype $BbEe$ can have chromosomes with alleles:

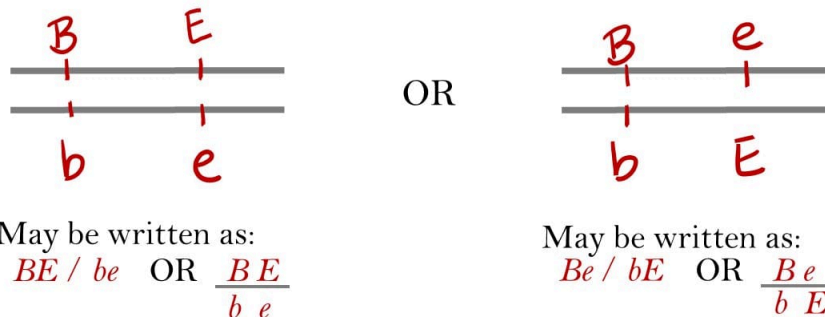


Figure 4.5.1.1: A cell with genotype $BbEe$ can have two possible arrangements of alleles on the chromosomes. (S. Leacock)

Exercise 4.5.1.1

In Figure 4.5.1.1, why is a chromosome with Bb or Ee not shown as a possible arrangement?

Answer available at the end of this section.

Effect of recombination on gamete possibilities

If loci B and E in the above example (Figure 4.5.1.1) were on two different chromosomes, we would expect to obtain four gamete genotypes (25% each): BE , Be , bE , and be , as observed by independent assortment. However, if B and E in the above example were so close that homologous recombination (crossing over) never occurs between them during meiosis, then all types of gametes

will not be observed. An organism with chromosomes BE / be could produce only gametes BE and be (50% each). An organism with chromosomes Be / bE could produce only gametes Be and bE (50% each).

Homologous recombination during meiosis I breaks and rejoins pieces of homologous chromosomes. If homologous recombination occurs between B and E then all four gametes will be possible. But at what frequency will each gamete be observed? The answer depends on how far apart they are! Recombination frequency is the percent of meioses in which homologous recombination exchanges two loci. In genetic mapping, this number expresses distance in **map units (m.u.)** or **centiMorgans (cM)** (named after geneticist Thomas Hunt Morgan).

How do geneticists know if recombination has occurred? Use a testcross. As shown in the next video, the map distance between loci B and E is determined by the number of recombinant offspring.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Remember:

- The # of recombinant offspring / total # of offspring $\times 100\%$ = recombination frequency
- Recombination frequency = map units = centiMorgan (cM)

Example 4.5.1.1

In the above example, number of recombinant offspring was used to calculate map distance. However, the map distance can also be used to predict recombinant offspring. What percentage of fg / fg offspring will be produced from a cross between Fg / fG and fg / fg if loci F and G are 30 map units apart?

Solution

The homozygous recessive parent can only transmit an fg chromosome, so determine the percentage of fg gametes from the dihybrid parent. The parental chromosomes are Fg and fG . The map distance (30 m.u.) is equal to the recombination frequency, so 30% of gametes will be recombinant, but there are two types of recombinants, so 15% will be FG and 15% will be fg . Therefore, we predict 15% of offspring will be fg / fg .

Exercise 4.5.1.2

Two hypothetical SNPs in humans are 4 map units (m.u.) apart. The allele at SNP 1 can be A or T; the allele at SNP 2 can be C or G.

A male with genotype AG / TC and a female with genotype AC / AC have a child.

Male (sperm-producing) parent

A G

T C

Female (oocyte-producing) parent

A C

A C

What is the probability they have a child with genotype AG / AC?

Answer available at the end of this section.

Query 4.5.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.5.1.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.5.1.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Double crossovers

Especially for large chromosomes, multiple crossover events can occur on the same chromosome. A double crossover occurs when fragments of the chromosome are exchanged in two places. The result of a double crossover is that the two ends of the chromosome are parental, but a region between the crossovers has been "swapped" for another sister chromatid sequence; this is depicted in the video.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Three factor crosses

Distances between multiple loci can be determined using three factor testcrosses. Again, we will cross a heterozygous parent to

a parent homozygous recessive for all three genes. When solving three factor test crosses, remember that in the heterozygote the dominant and recessive alleles can be on the same or different chromosomes. You will know which chromosomes are parental because they will be the most abundant offspring from the testcross. In contrast, the double crossover offspring will be the least abundant, because the double crossover events between the genes of interest are more rare than single crossovers.

Note: Steps for solving a three factor cross

- Identify the parental offspring (the most abundant).
- Identify the double crossover offspring (the least abundant).
- Determine which locus is in the middle (the one that "swaps places" in the double crossover).
- Determine the recombination frequency between one locus and the middle locus.
- Determine the recombination frequency between the other locus and the middle locus.
- Draw the genetic map.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Note that other factors that might influence recombination and double crossovers, such as the position along the chromosome or whether recombination at one site suppresses recombination nearby. These might make the number of observed recombinants different from expected, but we will not consider these factors at this time.

Exercise 4.5.1.3

Consider three genes on the same chromosome, *geneD*, *geneE* and *geneF*. To determine the distance between and order of these three genes, you conduct a test cross between a heterozygote *DEF / def* and a *def / def* homozygous recessive. The offspring produced from the cross are shown in the table. What is the distance between the genes? Draw a map showing the position and distance between the three loci.

Offspring class	Number of offspring
<i>DEF</i>	84
<i>def</i>	86
<i>Def</i>	41
<i>dEF</i>	43
<i>DEf</i>	25
<i>deF</i>	27
<i>DeF</i>	5
<i>dEf</i>	5

Answer available at the end of this section.

Answers

Exercise 4.5.1.1

The loci are the locations of the genes on chromosomes. The *B* gene must be on the same position on each homologous chromosome in pair. The *B* gene is not located at two positions on one chromosome.

Exercise 4.5.1.2

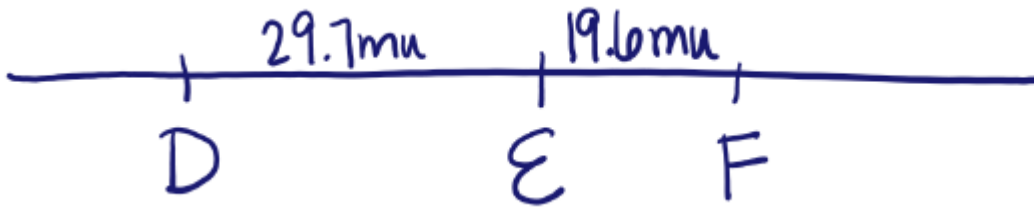
The female will produce eggs with an AC chromosome. (Recombination still occurs in during meiosis in this female, but with or without recombination, the outcome is the same for these two SNPs.) To have chromosomes AG / AC, child must inherit the AG parental chromosome from the father.



The map distance (4 m.u.) indicates that 4% of gametes will be recombinant. In the male sperm, 4% of gametes will contain a recombinant (AC or TG) chromosome, and 96% of gametes will be parental: 48% of gametes will have the AG chromosome and 48% will have the TC chromosome. Therefore, the probability of an AG / A C child is 48%.

Exercise 4.5.1.3

The distance between D and E is 29.7 m.u.; the distance between E and F is 19.6 m.u.



- [4.5.1: Linkage and Mapping](#) by [Stefanie West Leacock](#) has no license indicated.

4.5.2: GWAS

Learning Objectives

- Explain how molecular markers and linkage can be used to identify genes that contribute to complex traits through GWAS.
- Interpret a Manhattan plot, understanding that GWAS identify chromosomal regions that are linked to phenotypes.
- Explain how scientists use this information from GWAS to perform experiments on genes in candidate loci to determine if those genes have a role in the phenotype.

What are genome wide association studies?

We saw that linkage could be used to map genes. However, this approach is not always appropriate for tackling important questions in genetics. Consider these factors:

- In some cases, especially human genetics, directed crosses (test crosses) are not practical or ethical!
- For complex traits, the phenotype may vary along a spectrum.
- For complex traits, alleles of genes on multiple chromosomes likely contribute to the phenotype.

Instead, **genome wide association studies (GWAS)** compare single nucleotide polymorphisms across the genome and use them as markers to compare many individuals with and without the phenotype to ask which **SNPs** (loci) are likely to be inherited with the phenotype.

Important terms for understanding GWAS

- SNP: single nucleotide polymorphisms are base pairs in particular positions (loci) that vary among individuals
- Linkage: loci are sufficiently close to each other along a chromosome such that homologous recombination occurs between them less often than predicted by independent assortment
- Haplotype: a set of SNPs that can be inherited together on the same chromosome

How to genotype SNPs on a genomic scale

Depending on the species being studied, there may be several options for technologies to genotype the samples. An example is the Illumina platform, in which DNA is attached to slides and probes adjacent to each SNP are provided. DNA polymerase can then extend the probe by adding a base complementary to the next base in the sample. The free nucleotides are tagged and visualized by fluorescence.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 4.5.2.1: Illumina Advances Genomic Research with the Infinium Assay. (Illumina via <https://www.youtube.com/watch?v=IVG04dAAyY>)

P values and Manhattan Plots

The genotype data is analyzed to identify SNPs that have a statistically significant association with the trait of interest. Without covering the details of the calculations, in statistical analysis, a P value represents the likelihood that two events are associated by random chance. A low P value indicates a low likelihood that two things are associated by chance and therefore, the two things are correlated or connected one another. The datasets in GWAS are very large, for example, a study of human ear morphology genotyped approximately 670,000 SNPs in 4,919 individuals (Adhikari et al, 2015). For GWAS studies, a P value threshold may vary depending on the analysis, but it is often less than 0.00001 (1×10^{-5}) or even less than 0.00000005 (5×10^{-8}). The differences between such small P values can be difficult to visualize graphically, so scientists convert the P values to negative logarithms ($-\log_{10}(P \text{ value})$).

negative logs and P values

Remember that a log is an exponent, for example, $\log_{10}(100) = 2$ because $10^2 = 100$.

A significant P value for GWAS is $P < 5 \times 10^{-8}$ and the $\log_{10}(0.00000005) = -7.3$, so the $-\log = 7.3$, so any SNPs that are significantly associated should have a $-\log_{10}(P \text{ value}) > 7.3$.

The $-\log_{10}(P)$ are plotted on a graph called a Manhattan plot because the peaks at various positions across the genome resemble buildings such as in a city skyline. The peaks identify SNPs that are associated with the trait. The significant SNPs are not necessarily the cause of the disease or phenotype, but due to linkage may be inherited with the causal mutation.

Research example: A genome-wide association study identifies multiple loci for variation in human ear morphology

Craniofacial features are examples of complex traits that are likely controlled by many genes. In a GWAS in 2015, researchers examined 10 traits associated with human ear shape: ear protrusion, lobe size, lobe attachment, tragus size, antitragus size, helix rolling, folding of antihelix, crus helix expression, superior crus of antihelix expression and Darwin's tubercle (Adhikari et al, 2015).

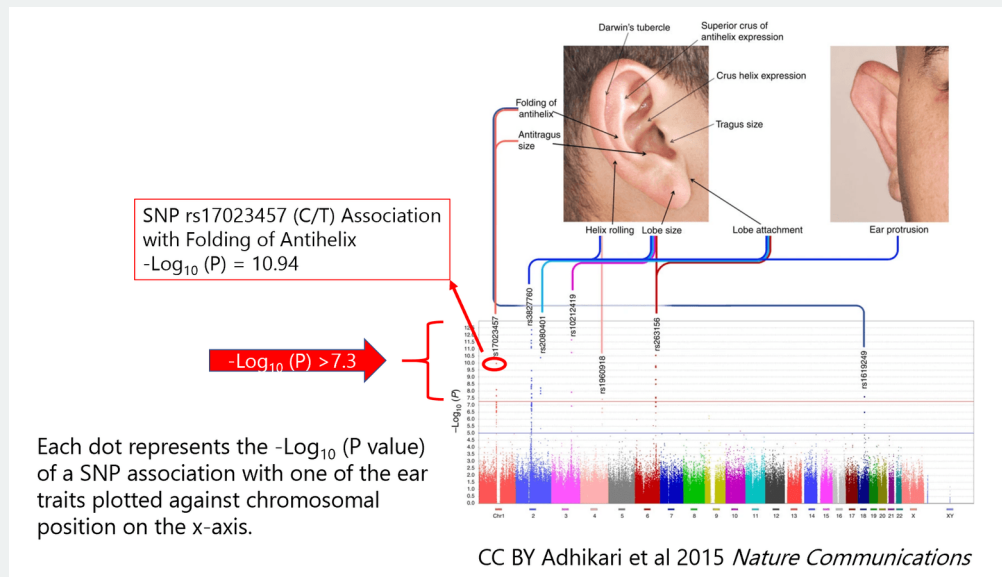


Figure 4.5.2.1: Manhattan plot of Genome-wide association of pinna (ear) morphological traits was plotted in a graph in which each dot represents the $-\log(P)$ value with the red line showing the threshold of 7.3, the x-axis represents the chromosomal positions across the human genome. (Figure annotated from CC BY Adhikari et al via *Nature Communications* <https://rdcu.be/cnqCp>)

This study identified SNPs on chromosomes 1 and 2 as the strongest candidates loci with potential to influence human ear morphology. A significant SNP on chromosome 2 is associated with a missense mutation in the EDAR protein, which has been implicated in human ectodermal derivatives including teeth and sweat glands. Additional studies performed in mice suggest that this protein has an additional role in ear formation. A significant SNP on chromosome 1 has been identified as a binding site for a transcriptional regulatory protein for a nearby gene, TBX15, which functions in cartilage development, suggesting this region is part of the regulatory environment that influences ear shape.

COVID GWAS research

A 2020 paper in the New England Journal of Medicine reports on a "Genomewide Association Study of Severe Covid-19 with Respiratory Failure" (<https://www.nejm.org/doi/full/10.1056/NEJMoa2020283>).

Consider the following questions about the paper:

- How many patients were involved and what information about patients was collected?
- Were there any ethical considerations about the study? If so, explain why.
- How many SNPs were genotyped and what was the method (see video above) and how many SNPs (or variants) showed statistical significance?
- How do the scientists move from an associated SNP (or other variant) and potential genes?
- What should the next experiments be to examine the role of the candidate gene(s)?

References

Adhikari, K., Reales, G., Smith, A. et al. A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nat Commun* **6**, 7500 (2015). <https://doi.org/10.1038/ncomms8500>

Ellinghaus D, Degenhardt F, Bujanda L, et al. Severe Covid-19 GWAS Group. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med*. 2020 Jun 17;NEJMoa2020283. doi: 10.1056/NEJMoa2020283. Epub ahead of print. PMID: 32558485; PMCID: PMC7315890.

This page titled [4.5.2: GWAS](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.5.2: GWAS](#) by [Stefanie West Leacock](#) has no license indicated.

4.6: Exceptions to simple dominance

Learning Objectives

- Categorize interactions between alleles of single genes as **completely dominant**, **incompletely dominant**, or **codominant** and explain the biochemical basis of these relationships. Predict phenotypes of offspring using each mode of inheritance.
- Understand that multiple alleles (not just two) may exist for a gene.
- Recognize examples of **incomplete penetrance**, **variable expressivity**, and **pleiotropy**.
- Predict ratios of offspring when certain genotypes are **lethal alleles**.

Incomplete Dominance

During Mendel's time, people believed in a concept of blending inheritance whereby offspring demonstrated intermediate phenotypes between those of the parental generation. This was refuted by Mendel's pea experiments that illustrated a Law of Dominance. Despite this, non-Mendelian inheritance can be observed in sex-linkage and co-dominance where the expected ratios of phenotypes are not observed clearly. **Incomplete dominance** superficially resembles the idea of blending inheritance, but can still be explained using Mendel's laws with modification. In this case, alleles do not exert full dominance and the offspring resemble a mixture of the two phenotypes.

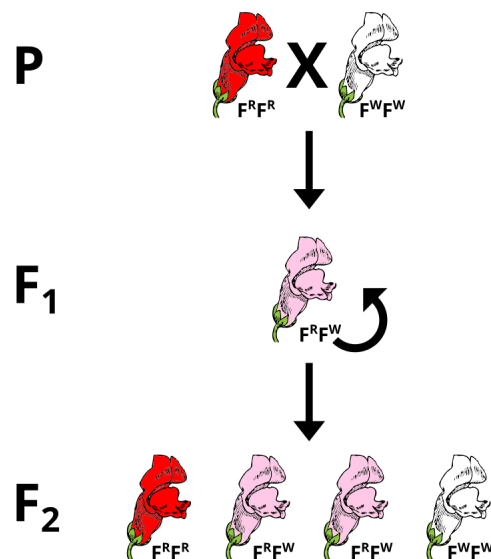


Figure 4.6.1: Alleles for red flowers (F^R) and alleles for white flowers (F^W) in snapdragons are incompletely dominant. The F_1 heterozygote exhibits an intermediate phenotype - pink. The round arrow indicates that the F_1 is self-fertilized. Image Credit: Jeremy Seto (CC-BY-NC-SA)

The most obvious case of a two allele system that exhibits incomplete dominance is in the snapdragon flower. The alleles that give rise to flower coloration (Red or White) both express and the heterozygous genotype yields pink flowers. There are different ways to denote this. In this case, the superscripts of R or W refer to the red or white alleles, respectively. Since no clear dominance is in effect, using a shared letter to denote the common trait with the superscripts (or subscripts) permit for a clearer denotation of the ultimate genotype to phenotype translations.

Exercise 4.6.1

If pink flowers arose from blending inheritance, then subsequent crosses of pink flowers with either parental strain would continue to change the phenotype. Using a Punnett Square, perform a cross between a heterozygous plant and a parental to predict the phenotypes of the offspring.

Answer available at the end of this section.

Thinking about the biochemical basis of incomplete dominance

Refer back to the biochemical basis of dominant and recessive alleles. Why do many loss-of-function alleles act as recessive? What must be different about incompletely dominant alleles? How would you test your idea?

The alleles that determine curly or straight hair in dogs provide another example of incomplete dominance. Using GWAS, the KRT71 gene, which encodes a keratin protein, was identified as the locus that determines this phenotype (Cadieu et al, 2010). Dogs with C alleles for a SNP in the KRT71 gene exhibit straight hair, while dogs with curly hair have two T alleles for the SNP that alter an amino acid, which is a missense mutation. When curly and straight haired dogs are bred, such as the cross of poodles and Labrador Retrievers that create Labradoodles, the offspring are heterozygous for the SNP (C/T) and have an intermediately wavy hair.

Query 4.6.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Codominance and Multiple Alleles

Codominance occurs when phenotypes associated with two dominant alleles occur simultaneously. The prototypical case for this is the human ABO blood grouping. The ABO gene (also known as the *I* locus) on human chromosome 9 encodes a glycosyltransferase enzyme that attaches a carbohydrate molecule to proteins present on the surface of red blood cells.

The I^A and I^B alleles have several DNA sequence differences that produce amino acid changes. As a result, the I^A allele produces an enzyme that adds an acetylated galactosamine (GalNAc) to a surface protein. The I^B allele produces an enzyme that adds galactose (Gal) instead. These alleles are codominant to each other because each the enzymes catalyze reactions that occur independently of the other enzymes that are present. On the surface of red blood cells in heterozygous individuals ($I^A I^B$), a mixture of proteins with GalNAc and Gal will be present.

A third allele of the gene, designated *i*, also exists but contains a single nucleotide deletion that causes a frameshift, completing disrupting the protein (Yamamoto et al, 1990). The protein produced by the *i* allele does not catalyze the attachment of any sugar. This allele is recessive to I^A and I^B because the protein has no activity and does not influence the activity of the proteins encoded by the dominant alleles. If an I^A or I^B allele is also present, those enzymes will still function, so the A or B antigens will be present, producing the dominant phenotype.

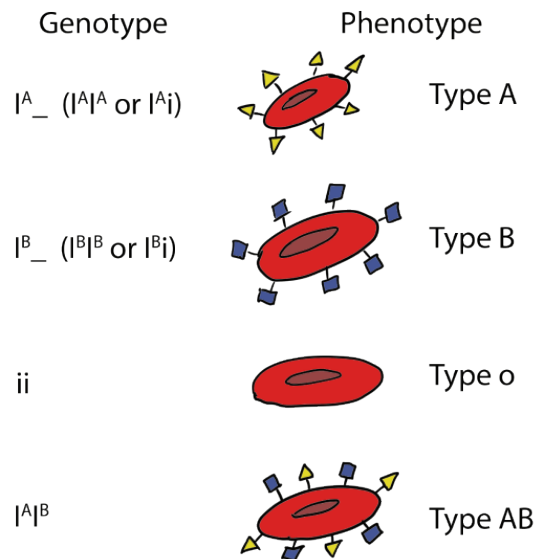


Figure 4.6.2: The three alleles of ABO system can produce 6 genotypes and four blood phenotypes: A, B, O and AB. (CC0 Wikimedia Commons)

Query 4.6.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.6.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Importance of ABO Blood Types

The alleles of the *ABO* gene determine features that are present on the surface of blood cells that are used when cells identify foreign cells. An individual with blood type A has proteins with the GalNac modification, but does not have proteins with the Gal modification produced by the I^B allele. The individual will produce antibodies that recognize the Gal modification that is present in Type B blood. Therefore, if Type B blood should not be introduced into a Type A individual.

Blood type influences blood donors and recipients

The table shows how blood type influences blood donations. If an individual receives a blood type that they make antibodies against, the blood will be recognized as foreign and the immune system will attack the donated cells. Because individuals with Type O blood have cells without any sugar modifications, their blood can be donated to any other blood type; they are sometimes called universal donors. In contrast, individuals with Type AB blood recognize either the GalNAc or Gal modification and can receive any type of blood; they are sometimes called universal recipient.

Blood type	Antibodies against:	Can donate blood to:	Can receive blood from:
A	B	A or AB	A or O
B	A	B or AB	B or O
AB	none	AB	A, B, AB, or O
O	A and B	A, B, AB, or O	O

There are additional genes that influence human blood type, for example, blood type might also be described as O positive or AB negative. Whether blood type is positive or negative is determined by a second locus, the *RHD* gene, which encodes a surface protein called the Rh Blood Group D Antigen. As Mendelian inheritance predicts, the two genes, *ABO* on chromosome 9 and *RHD* on chromosome 1, are inherited independently during meiosis.

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. For example, all pea plants that are homozygous for the allele for white flowers actually have white flowers, this genotype is completely penetrant. In contrast, many human genetic diseases are incompletely penetrant, because 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. Many human genetic diseases provide examples of broad expressivity, because 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), although often it remains unknown which of these cause(s) are most responsible for individual phenotypes.

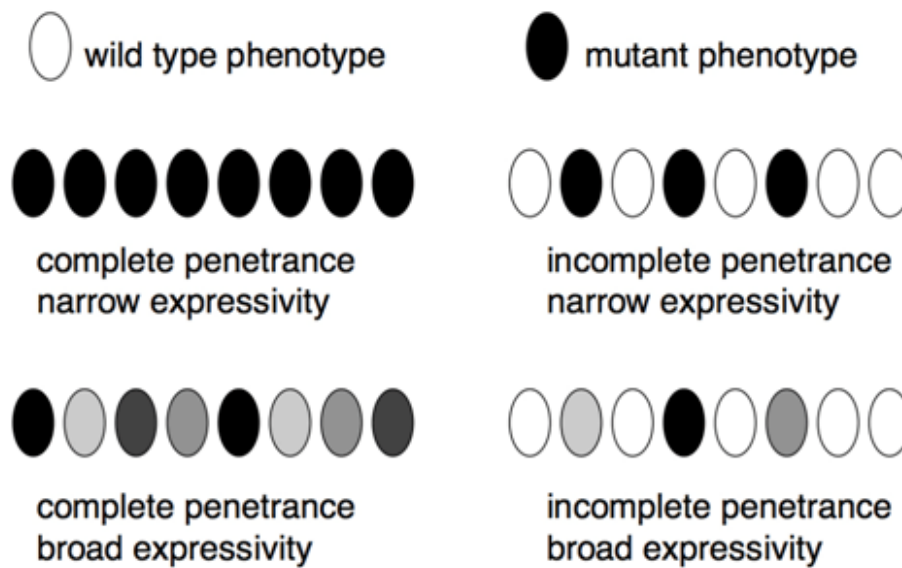


Figure 4.6.3: Relationship between penetrance and expressivity in eight individuals that all have a mutant genotype. Penetrance can be complete (all eight have the mutant phenotype) or incomplete (only some have the mutant phenotype). Among individuals with the mutant phenotype, the expressivity can be narrow (very little variation) to broad (lots of variation). (Original-Locke-CC:AN)

Example of variable penetrance and expressivity from nematodes

Mutations in the gene *meg-1* in *C. elegans* provide an example of variable penetrance as well as broad expressivity. This gene encodes a protein (MEG-1) that functions to establish the reproductive system in cells during early embryonic development. Mutant worms that lack *meg-1* therefore produce sterile offspring. However, the sterile phenotype is not fully penetrant, approximately 20-25% of mutant offspring will be fertile, despite the lack of MEG-1. These sterile adults also exhibit variable expressivity; some sterile adults have severely-reduced germ-line development with few identifiable germ cells, whereas other sterile adults have an identifiable germ line but lack gametes.

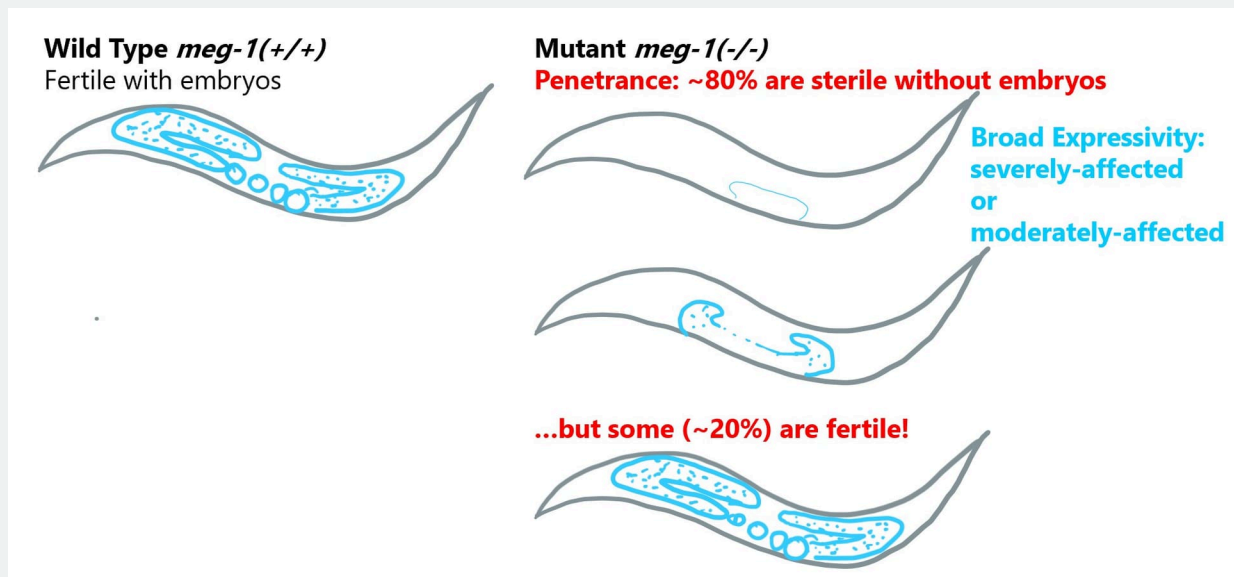


Figure 4.6.4: Incomplete penetrance and variable expressivity in *meg-1* mutants in nematodes. Worms with the wild-type alleles have a large u-shaped bilateral germ line (blue) that produces embryos (circles). Mutants with the genotype *meg-1*(-/-), but some are fertile, illustrating in complete penetrance. Sterile worms that lack embryos can be severely-affected with few identifiable germ cells or moderately-affected with a small but identifiable number of germ cells that fails to produce gametes or embryos. (CC BY Leacock)

Pleiotropy

While some genes are associated with a single trait, many genes encode proteins that function in multiple ways or are the basis for other traits. Pleiotropy occurs when a single genotype affects multiple phenotypes. For example, the human gene *CFTR* is associated with the disease cystic fibrosis. However, variations in this gene are also associated with other phenotypes, such as congenital bilateral absence of the vas deferens (a feature that causes male infertility) and pancreatitis.

Pleiotropy is often observed when a gene is expressed in more than one tissue or organ. The CFTR protein plays a role in ion balance in the lungs, where lack of the protein is associated with the thick lung mucus characteristic of cystic fibrosis. However, CFTR is expressed and functions in ion balance in other tissues as well. Experiments that assay transcripts in multiple tissues, show that the CFTR mRNA is present in a number of tissues, including pancreas and testis. This expression pattern is consistent with some of the observed phenotypes that occur when CFTR is mutated.

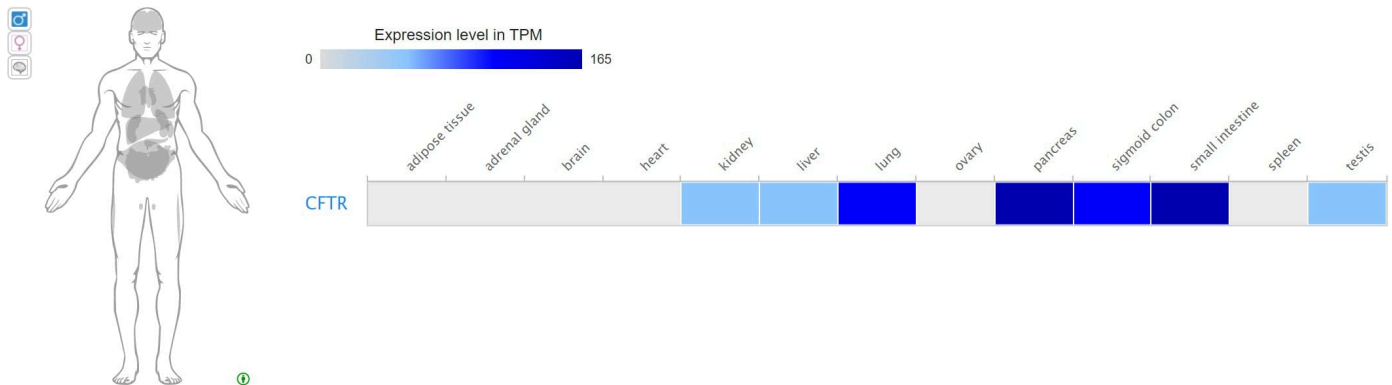


Figure 4.6.5: Expression level of CFTR in human tissues in TPM (transcripts per million) shows expression in kidney, liver, and testis in light blue, and higher expression in lung, pancreas, colon, and small intestine in dark blue (data from Lin et al 2014). (CC BY [Expression Atlas update: from tissues to single cells](#) (*Nucleic Acids Research*, 30 October 2019))

Lethal Alleles

Mutations that cause loss-of-function of a protein that is essential for organismal viability could result in lethality; often these alleles are recessive because the second copy of the gene will produce sufficient protein for survival. When two heterozygotes are crossed, 25% of offspring are expected to be homozygous for the lethal allele. Thus, the ratio of homozygous wild type to heterozygotes among the surviving offspring will be 2:1 and the offspring may all have the same phenotype, depending on the relationship between the two alleles.

	emb^{+}/emb^{-}	X	emb^{+}/emb^{-}
	emb^{+}		emb^{-}
emb^{+}	emb^{+}/emb^{+}		emb^{+}/emb^{-}
emb^{-}	emb^{+}/emb^{-}		emb^{-}/emb^{-}

Figure 4.6.6: Example of a cross with an embryonic lethal allele of a gene designated *emb*. The emb^{+} allele is wild type and the emb^{-} allele is a recessive lethal, so those offspring will not survive, leaving a ratio of 2:1 emb^{+}/emb^{-} heterozygotes to emb^{+}/emb^{+} homozygotes among the surviving offspring.

For some alleles, the heterozygote may have a different phenotype than the homozygote, in which case the phenotypic ratios of the progeny may differ from the expected ratios. The *agouti yellow* (A^Y) allele in mice is one example of an allele that produces a dominant yellow phenotype in heterozygotes, as well as pleiotropic effects including obesity and increased tumor incidence. Mammals typically produce two types of pigments, pheomelanin (a yellow pigment) and eumelanin (a black-brown pigment). The protein encoded by *agouti* is a signaling protein that suppresses a receptor controlling eumelanin production. When agouti protein is present transiently during hair production, individual hairs have eumelanin at the base and tip with a central band of pheomelanin, giving an overall brown fur appearance. In A/A^Y heterozygotes agouti is constantly expressed, suppressing eumelanin completely and the phenotype is yellow. The pleiotropic effects observed are likely due to interactions of the Agouti signaling protein with similar receptors that control weight and feeding behavior.

The A^Y allele is lethal when homozygous due to the nature of the mutation. The A^Y allele is a large (170kb) deletion that removes a part of an upstream gene *Raly*, which stands for RNA-binding protein associated with the *lethal yellow* mutation, and the *agouti* promoter (Michaud et al 1993; Duhl et al 1994). Now the *Raly* promoter controls the transcription of agouti, such that it is overproduced, suppressing melanin production and leading to yellow fur. Due to the size and position of the deletion, the coding sequences of the Raly protein are absent in the A^Y allele. In heterozygous embryos a functional copy of *Raly* remains, but Raly protein, which is normally expressed in early mouse embryos (preimplantation), is absent in homozygous A^Y/A^Y embryos. When translation of the *Raly* mRNA was blocked in wild-type mouse embryos, only 4% of embryos developed to the blastocyst stage, consistent with the deletion of *Raly*, and not the effect of *agouti*, as the cause of the embryonic lethality (Duhl et al. 1994).

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 4.6.1: The *Raly* and *agouti* genes are adjacent in the mouse genome. The A allele results in agouti protein transiently inhibiting eumelanin production producing black and yellow fur, whereas the A^Y deletion results in expanded agouti production and yellow fur as well as obesity. The deletion also removes the *Raly* protein-coding sequence which is lethal in the homozygous state. (Images CC BY Ensembl and Jirtle and Dolinoy; Leacock)

Example 4.6.1

The mouse gene *Agouti* (*A*) influences the pattern of pigment in mouse fur. The mutant allele A^Y produces fur that is entirely yellow, instead of the wild-type pattern of black hairs with a central band of yellow pigment. However, $A^Y A^Y$ homozygous embryos die in early stages of development. You cross two yellow mice. What ratio of offspring genotypes and phenotypes do you expect?

Solution

The yellow mice must have the heterozygous genotype AA^Y . The fertilized embryos will have the genotype ratio $1 A^Y A^Y : 2 AA^Y : 1 AA$.

However the $A^Y A^Y$ embryos will not survive. The ratio of the surviving offspring will be $2 AA^Y$ yellow : $1 AA$ agouti.

References:

Cadiou E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A, Wong A, Mosher DS, Elkahoul AG, Spady TC, André C, Lark KG, Cargill M, Bustamante CD, Wayne RK, Ostrander EA. Coat variation in the domestic dog is governed by variants in three genes. *Science*. 2009 Oct 2;326(5949):150-3. doi: 10.1126/science.1177808. Epub 2009 Aug 27. PMID: 19713490; PMCID: PMC2897713.

Duhl DM, Stevens ME, Vrieling H, Saxon PJ, Miller MW, Epstein CJ, Barsh GS. Pleiotropic effects of the mouse lethal yellow (A^Y) mutation explained by deletion of a maternally expressed gene and the simultaneous production of agouti fusion RNAs. *Development*. 1994 Jun;120(6):1695-708. PMID: 8050375.

Howe K, et al Ensembl 2021. *Nucleic Acids Res*. 2021, vol. 49(1):884–891. PubMed PMID: 33137190.

Leacock SW, Reinke V. MEG-1 and MEG-2 are embryo-specific P-granule components required for germline development in *Caenorhabditis elegans*. *Genetics*. 2008 Jan;178(1):295-306. doi: 10.1534/genetics.107.080218. PMID: 18202375; PMCID: PMC2206079. (<https://doi.org/10.1534/genetics.107.080218>)

Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman WC, Gingeras TR, Ecker JR, Snyder MP. Comparison of the transcriptional landscapes between human and mouse tissues. *Proc Natl Acad Sci U S A*. 2014 Dec 2;111(48):17224-9. doi: 10.1073/pnas.1413624111. Epub 2014 Nov 20. PMID: 25413365; PMCID: PMC4260565.

Michaud EJ, Bultman SJ, Klebig ML, et al. A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (A^Y) mutation. *Proc Natl Acad Sci U S A*. 1994 Mar 29;91(7):2562-6. doi: 10.1073/pnas.91.7.2562. PMID: 8146154; PMCID: PMC43409. (<https://www.pnas.org/content/91/7/2562.long>)

Papatheodorou I, Moreno P, Manning J, et al. Expression Atlas update: from tissues to single cells. *Nucleic Acids Res*. 2020 Jan 8;48(D1):D77-D83. doi: 10.1093/nar/gkz947. PMID: 31665515; PMCID: PMC7145605.

Yamamoto *et al*. Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229–233 (1990). <https://doi.org/10.1038/345229a0>

Contributors and Attributions

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

Answers

Exercise 4.6.1

A cross between a heterozygote $F^R F^W$ and a red parental $F^R F^R$ would produce 1/2 red-flowered and 1/2 pink-flowered offspring.

	F^R	F^R
F^R	$F^R F^R$ (red)	$F^R F^R$ (red)
F^W	$F^R F^W$ (pink)	$F^R F^W$ (pink)

This page titled [4.6: Exceptions to simple dominance](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.6: Exceptions to simple dominance](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

4.7: Gene Interactions

Learning Objectives

- Describe the results of crosses involving two genes that affect the same trait as exhibiting **complementation**, **epistasis**, or **redundancy**.
- Predict phenotypes of offspring of crosses involving two genes, given information about biochemical pathways or functions of the genes.

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, 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 4.7.1: 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**.

	<i>EB</i>	<i>Eb</i>	<i>eB</i>	<i>eb</i>
<i>EB</i>	<i>EEBB</i>	<i>EEBb</i>	<i>EeBB</i>	<i>EeBb</i>
<i>Eb</i>	<i>EEBb</i>	<i>EEbb</i>	<i>EeBb</i>	<i>Eebb</i>
<i>eB</i>	<i>EeBB</i>	<i>EeBb</i>	<i>eeBB</i>	<i>eeBb</i>
<i>eb</i>	<i>EeBb</i>	<i>Eebb</i>	<i>eeBb</i>	<i>eebb</i>

Figure 4.7.2: 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)

Biochemical Basis for the genetic interaction between *E* and *B* in dog coat color

Why is this genetic interaction observed? The *B* locus encodes the protein *TYRP1* (Tyrosinase-Related Protein 1). The enzyme catalyzes a step in the production of melanin within melanocytes. Predicted loss-of-function alleles of *TYRP1* were identified in brown dogs, for example a premature stop codon (Schmutz et al. *TYRP1* and *MC1R* genotypes and their effects on coat color in dogs. *Mamm Genome*. 2002;13(7):380-387). Dogs homozygous or heterozygous for functional *TYRP1* produce sufficient melanin to have a black coat.

The *E* locus encodes the protein *MC1R* (melanocortin receptor 1), which is a membrane bound receptor for melanocyte-stimulating hormone (MSH) on the surface of the melanocytes. Activation of this receptor is transduced into a transcriptional response to produce proteins needed for eumelanin production. Recessive loss-of-function alleles of *MC1R* result in reduced or absent receptor activity and therefore loss of the intracellular response of proteins needed for eumelanin production. When *MC1R* is not activated, cells can still produce the yellow pheomelanin, resulting in a yellow coat.

In the Punnett square above, any genotype that includes a homozygous recessive for *MC1R* (*ee*) results in a yellow coat, even when the dog has a dominant allele *B* that is predicted to produce black fur. Without a functional *MC1R* receptor, the downstream events needed to produce *TYRP1* will not be activated, so the *ee* genotype "masks" the production of brown or black fur.

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. 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. 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 4.7.3: Green, yellow, and white fruits of squash. (Flickr-Unknown-CC:AN)

	<i>AB</i>	<i>Ab</i>	<i>aB</i>	<i>ab</i>
<i>AB</i>	<i>AABB</i>	<i>AABb</i>	<i>AaBB</i>	<i>AaBb</i>
<i>Ab</i>	<i>AABb</i>	<i>AAbb</i>	<i>AaBb</i>	<i>Aabb</i>
<i>aB</i>	<i>AaBB</i>	<i>AaBb</i>	<i>aaBB</i>	<i>aaBb</i>
<i>ab</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBb</i>	<i>aabb</i>

Figure 4.7.4: Genotypes and phenotypes among the progeny of a dihybrid cross of squash plants heterozygous for two loci affecting fruit color. (Original-Deyholos-CC:AN)

Redundant genes

When a dihybrid cross produces progeny in two phenotypic classes in a 15:1 ratio, this can be because the two loci's gene products have the same (**redundant**) functions within the same biological pathway. Yet another pigmentation pathway, in this case in wheat, provides an example of gene redundancy, sometimes known as duplicate gene action. The biosynthesis of red pigment near the surface of wheat seeds involves many genes, two of which we will label *A* and *B*. Normal, red coloration of the wheat seeds is maintained if function of either of these genes is lost in a homozygous mutant (e.g. in either *aaB_* or *A_bb*). Only the doubly recessive mutant (*aabb*), which lacks function of both genes, shows a phenotype that differs from that produced by any of the other genotypes. A reasonable interpretation of this result is that both genes encode the same biological function, and either one alone is sufficient for the normal activity of that pathway.



Figure 4.7.5: Red (left) and white (right) wheat seeds. (cropwatch.unl.edu-pending)

	<i>AB</i>	<i>Ab</i>	<i>aB</i>	<i>ab</i>
<i>AB</i>	<i>AABB</i>	<i>AABb</i>	<i>AaBB</i>	<i>AaBb</i>
<i>Ab</i>	<i>AABb</i>	<i>Aabb</i>	<i>AaBb</i>	<i>Aabb</i>
<i>aB</i>	<i>AaBB</i>	<i>AaBb</i>	<i>aaBB</i>	<i>aaBb</i>
<i>ab</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBb</i>	<i>aabb</i>

Figure 4.7.6: Genotypes and phenotypes among the progeny of a dihybrid cross of a wheat plants heterozygous for two loci affecting seed color. (Original-Deyholos-CC:AN)

Complementary gene action

The progeny of a dihybrid cross may produce just two phenotypic classes, in an approximately **9:7** ratio. An interpretation of this ratio is that the loss of function of either *A* or *B* gene function has the same phenotype as the loss of function of both genes, due to **complementary gene action** (meaning that the functions of both genes work together to produce a final product). For example, consider a simple biochemical pathway in which a colorless substrate is converted by the action of gene *A* to another colorless product, which is then converted by the action of gene *B* to a visible pigment. Loss of function of either *A* or *B*, or both, will have the same result: no pigment production. Thus *A_bb*, *aaB_*, and *aabb* will all be colorless, while only *A_B_* genotypes will produce pigmented product. The modified 9:7 ratio may therefore be obtained when two genes act together in the same biochemical pathway, and when their loss of function phenotypes are indistinguishable from each other or from the loss of both genes.

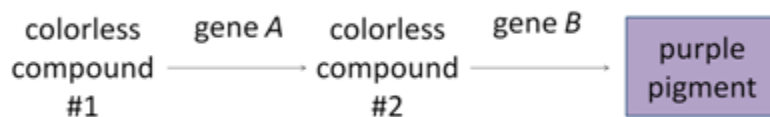


Figure 4.7.7: A simplified biochemical pathway showing complementary gene action of *A* and *B*. Note that in this case, the same phenotypic ratios would be obtained if gene *B* acted before gene *A* in the pathway. (Original-Deyholos-CC:AN)

	<i>AB</i>	<i>Ab</i>	<i>aB</i>	<i>ab</i>
<i>AB</i>	<i>AABB</i>	<i>AABb</i>	<i>AaBB</i>	<i>AaBb</i>
<i>Ab</i>	<i>AABb</i>	<i>AAbb</i>	<i>AaBb</i>	<i>Aabb</i>
<i>aB</i>	<i>AaBB</i>	<i>AaBb</i>	<i>aaBB</i>	<i>aaBb</i>
<i>ab</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBb</i>	<i>aabb</i>

Figure 4.7.8: Genotypes and phenotypes among the progeny of a dihybrid cross of a hypothetical plant heterozygous for two loci affecting flower color. (Original-Deyholos-CC:AN)

Contributors and Attributions

- [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 [4.7: Gene Interactions](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [4.7: Gene Interactions](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

4.7.1: Complementation Test

If two mutant organisms have the same phenotype and the mutant sequence has not been identified, how do you know whether the two organisms have mutations in the same gene or in two different genes that both are required for the same process?

The answer can be determined by a special test cross known as a complementation test. For this test, two homozygous recessive organisms are crossed. If the mutations are in the same gene then both copies of the gene will be mutant in the F1 offspring and they will exhibit the same phenotype as their parents. In this case, the exact mutations may be different (for example, a missense in one mutant and a nonsense in the other mutant), but both alleles of a single gene are mutant and therefore the recessive phenotype will be observed. In contrast, if the two mutants have homozygous recessive mutant alleles for two different genes, *gene1*^{-/-} and *gene2*^{-/-} for example, then the offspring will inherit a good copy of *gene1*⁺ from the *gene2*^{-/-} mutant and a wild-type copy of *gene2* from the *gene1*^{-/-} mutant parent.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 4.7.1.1: How to conduct and interpret a complementation test. (CC BY SW Leacock)

Summary: Interpreting a complementation test

Cross Mutant 1 x Mutant 2

If the F1s all have the wild-type phenotype, the two mutants **complement** each other and the mutations are in two different genes.

If the F1s all have the mutant phenotype, the two mutants **fail to complement** each other and the mutations are likely in the same gene.

Query 4.7.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



4.8: Chapter 4 Review Questions (draft)

4.1: Meiosis

Query 4.8.4.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.8.4.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.8.4.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.8.4.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.8.4.5

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 4.8.4.6

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



5: CHROMOSOME VARIATION

5.1: Changes in Chromosome Number

5.2: Changes in Chromosome Structure

5.3: Chapter 5 Review Questions (Draft)

Template:HideTOC

This page titled [5: Chromosome variation](#) is shared under a license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [5: Chromosome variation](#) by [Stefanie West Leacock](#) has no license indicated.

5.1: Changes in Chromosome Number

Learning Objectives

- Describe nondisjunction and its outcomes.
- Distinguish between aneuploidy (example: nullisomy or trisomy) and polyploidy.
 - Give an example of a human trisomy and explain why trisomies of some chromosomes are not tolerated (viable in offspring).
 - Explain why organisms with an odd number of chromosome sets are usually sterile or not even viable.
 - Recall that polyploidy is common in plants but rare in animals. Give an example of an advantage of polyploidy.

Single Chromosomes can be Lost or Gained by Nondisjunction

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 abnormalities 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. 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. 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 5.1.1, 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 a 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.

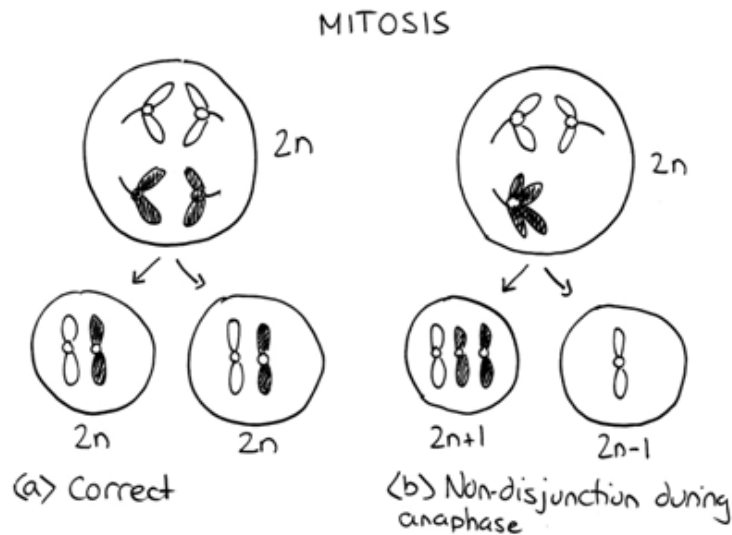


Figure 5.1.1: Mitosis done successfully (a) and unsuccessfully (b). The cell is diploid and the homologs of one chromosome are shown in white and black. (Original-Harrington-CC:AN)

If a nondisjunction event occurs during either meiosis I or meiosis II, the result is an unbalanced gamete. 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.

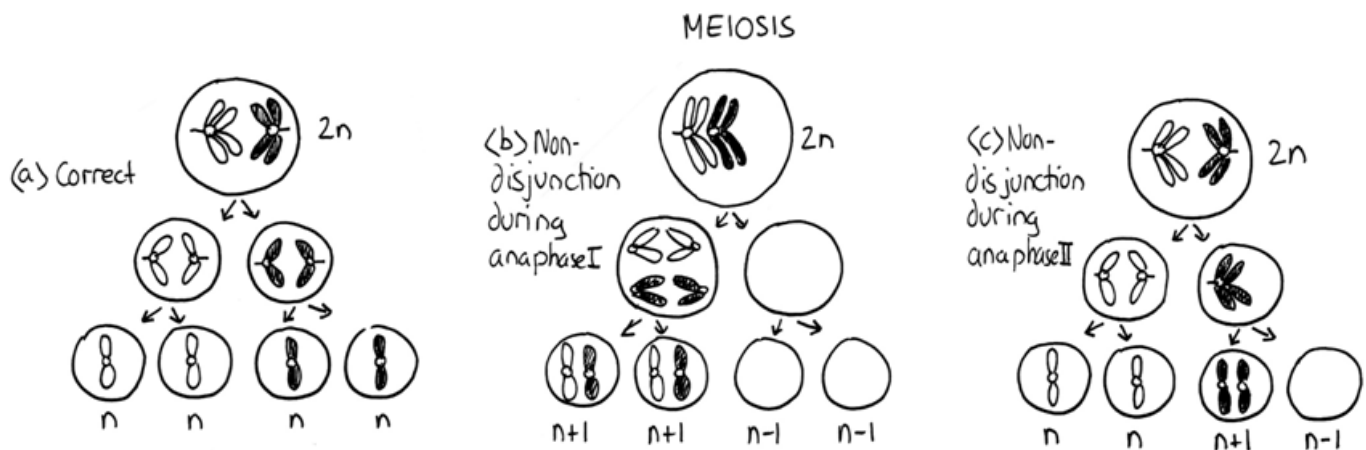


Figure 5.1.2: Meiosis done successfully (a) and unsuccessfully (b and c). (Original-Harrington-CC:AN)

Changes in sets of chromosomes

Humans, like most animals and most eukaryotic genetic model organisms, are diploids because they have two copies of each autosome. This means that most of their cells have two homologous copies of each chromosome. In contrast, many plant species and even a few animal species are **polyploids**. This means they have more than two chromosome sets, and so have more than two homologs of each chromosome in each cell.

When the nuclear content changes by a whole chromosome set we call it a change in ploidy. Gametes are haploid (n) and thus most animals are diploid ($2n$), formed by the fusion of two haploid gametes. However, some species can exist as monoploid ($1x$), triploid ($3x$), tetraploid ($4x$), pentaploid ($5x$), hexaploid ($6x$), or higher.

Notation of ploidy

As we have already seen, the letter n is the number of chromosomes in a gamete, and $2n$ is the number of chromosomes following fertilization.

To distinguish polyploids, the letter x is used to represent chromosome sets. For example:

- For a tetraploid $2n=4x$ and a gamete has $n=2x$ chromosomes.
- For a hexaploid $2n=6x$ and a gamete has $n=3x$ chromosomes.

Query 5.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Male Bees are Monoploid

Monoploids, with only one set of chromosomes, are usually inviable in most species; however, in many species of *Hymenoptera* (bees, wasps, ants) the males are monoploid and develop from unfertilized eggs. These males don't undergo meiosis for gametes; mitosis produces sperm. Females are diploid (from fertilized eggs) and produce eggs via meiosis. Female bees are diploid ($2n=32$) and are formed when an egg ($n=16$) is fertilized by a sperm ($n=16$). If an egg is not fertilized, it can still develop and the result is a $n=16$ male drone. Males are described as haploid (because they have the same number of chromosomes as a gamete) or monoploid (because they have only one chromosome set). Females produce eggs by meiosis while males produce sperm by mitosis. This form of sex determination produces more females, workers, which do the work than males, who are only needed for reproduction.



Figure 5.1.3: A worker caste European honey bee, which is female and diploid. Male drones are haploid. (Wikipedia-J. Severns-PD)

Polyploids can be stable or sterile

Like diploids ($2n=2x$), stable polyploids generally have an even number of copies of each chromosome: tetraploid ($2n=4x$), hexaploid ($2n=6x$), and so on. The reason for this is clear from a consideration of meiosis. Remembering that the purpose of meiosis is to reduce the sum of the genetic material by half, meiosis can equally divide an even number of chromosome sets, but not an odd number. Thus, polyploids with an odd number of chromosomes (e.g. triploids, $2n=3x$) tend to be sterile, even if they are otherwise healthy.

The mechanism of meiosis in stable polyploids is essentially the same as in diploids: during metaphase I, homologous chromosomes pair with each other. Depending on the species, all of the homologs may be aligned together at metaphase, or in multiple separate pairs. For example, in a tetraploid, some species may form tetravalents in which the four homologs from each chromosome align together, or alternatively, two pairs of homologs may form two bivalents. Note that because that mitosis does

not involve any pairing of homologous chromosomes, mitosis is equally effective in diploids, even-number polyploids, and odd-number polyploids.

Many Crop Plants are Hexaploid or Octoploid

Polyloid plants tend to be larger and healthier than their diploid counterparts. The strawberries sold in grocery stores come from octoploid (8x) strains and are much larger than the strawberries formed by wild diploid strains.

Bread wheat is a hexaploid ($2n=6x$) strain. This species is derived from the combination of three other wheat species, *T. monococcum*, *T. searsii*, and *T. tauschii*. Each of these chromosome sets has 7 chromosomes so the diploid species are $2n=2x=14$ and bread wheat is $2n=6x=42$ and has 14 chromosomes from each species. Bread wheat is viable because each chromosome behaves independently during mitosis. The species is also fertile because during meiosis I the equivalent chromosomes from each species can pair each other during meiosis I. Thus, even in a polyploid, homologous chromosomes can segregate equally and gene balance can be maintained.

Bananas, Watermelons, and Other Seedless Plants are Triploid

The bananas found in grocery stores are a seedless variety called Cavendish. They are a triploid variety (chromosome sets = AAA) of a normally diploid species called *Musa acuminata* (AA). Cavendish plants are viable because mitosis can occur. However they are sterile because the chromosomes cannot pair properly during meiosis I. During prophase I there are three copies of each chromosome trying to “pair” with each other. Because proper chromosome segregation in meiosis fails, seeds cannot be made and the result is a fruit that is easier to eat because there are no seeds to spit out.

Why are triploids sterile?

Remember that during meiosis in a diploid there are four sister chromatids for each chromosome and 4 cells (gametes) produced. Each gamete inherits one copy of each chromosome. In a triploid after DNA replication (S phase), there will be six sister chromatids for each chromosome, but meiosis I and II still undergo the same process.

- How many chromatids will be in each gamete?
- Will the number of copies of each chromosome be the same in a single gamete?

For a cell with three types of chromosomes ($n=3$), draw the chromosomes in a triploid cell. Show how those chromosomes might segregate during meiosis. What possibilities exist for number of chromosomes in each gamete?

Exercise 5.1.1

Seedless watermelons are another common triploid. The haploid number (n) of standard watermelons is 11. How many chromosomes are present in a single triploid watermelon cell? The thin white seeds in these triploids are the result of unbalanced gametes, but rare black viable seeds are sometimes found in a seedless watermelon. What would have to happen for these viable seeds to be produced?



Figure 5.1.4: Seedless watermelon is triploid, with the white, aborted seeds within the flesh. (Flickr-Darwin Bell-CC:AN)

Answer available at the end of this section.

If triploids cannot make seeds, how do we obtain enough triploid individuals for cultivation? The answer depends on the plant species involved. In some cases, such as banana, it is possible to propagate the plant asexually; new progeny can simply be grown from cuttings from a triploid plant. On the other hand, seeds for seedless watermelon are produced sexually: a tetraploid watermelon plant is crossed with a diploid watermelon plant. Both the tetraploid and the diploid are fully fertile, and produce gametes with two ($1n=2x$) or one ($1n=1x$) sets of chromosomes, respectively. These gametes fuse to produce a zygote ($2n=3x$) that is able to develop normally into an adult plant through multiple rounds of mitosis, but is unable to compete normal meiosis or produce seeds.

Polyploids are often larger in size than their diploid relatives. Food plant cultivation takes advantage of this property. For example, most strawberries you eat are not diploid, but octoploid ($8x$).

Polyploidy in Animals

Polyploidy in animals is rare, essentially limited to lower forms, although some vertebrates (fish) have been identified as triploid. Cultivated oysters are an example of a viable and commercially relevant triploid animal. In the 1990s, biologists explored ways to alter the number of chromosome sets in the oyster *Crassostrea gigas* (Guo et al, 1992).

Triploid oysters

Triploid oysters are cultivated for human consumption (<https://www.theatlantic.com/technology/archive/2014/09/todays-oysters-are-mutants/380858/>). What are the advantages of triploidy in oysters? How are triploid oysters similar to triploid plants? Based on the observation of rare fertilized seeds in seedless watermelons, would you expect to find rare fertile triploid oysters?

Triploidy in humans is not a viable condition, occurring in as much as 20% of miscarriages that show chromosomal abnormalities (reviewed in Kolarski et al, 2017). Triploidy in humans may be caused by abnormal meiosis that produces a gamete with two sets of chromosomes or by fertilization of a single egg by two sperm.

References

Guo, Ximing, et al. "Genetic Consequences of Blocking Polar Body I with Cytochalasin B in Fertilized Eggs of the Pacific Oyster, *Crassostrea Gigas*: I. Ploidy of Resultant Embryos." *Biological Bulletin*, vol. 183, no. 3, 1992, pp. 381–386. *JSTOR*, www.jstor.org/stable/1542013. Accessed 7 July 2020.

Kolarski M, Ahmetovic B, Beres M, et al. Genetic Counseling and Prenatal Diagnosis of Triploidy During the Second Trimester of Pregnancy. *Med Arch*. 2017;71(2):144-147. doi:10.5455/medarh.2017.71.144-147

Contributors and Attributions

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

Answers

Exercise 5.1.1

A triploid watermelon has 3 sets of 11 chromosomes: $3 \times 11 = 33$.

Although most gametes will have unbalanced chromosomes (example: one copy of five chromosomes but two copies of six chromosomes), rare balanced gametes can be produced. For example, by chance a balanced gamete with one copy of all eleven chromosomes is fertilized by another balanced gamete with two copies of all chromosomes.

This page titled [5.1: Changes in Chromosome Number](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [5.1: Changes in Chromosome Number](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

5.2: Changes in Chromosome Structure

Learning Objectives

- Distinguish between the 4 types of chromosomal rearrangements and their effects on gene dosage and meiosis.

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 and consequences of chromosome structural abnormalities are described here. They all involve breaks in the chromosomal DNA backbone.

Cause #1: Incorrect Repair of Double Strand DNA Breaks During Interphase

A chromosome is a very long but very thin molecule. In the phosphodiester 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. 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. 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).

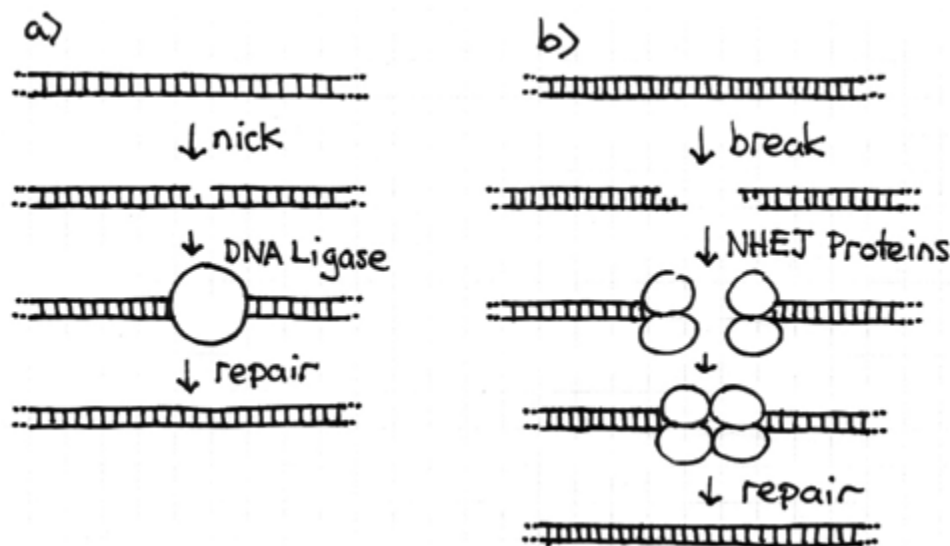


Figure 5.2.1: Repair of single-strand nicks and double-strand breaks in DNA. (Original-Harrington-CC:AN)

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. However, if there are two (or more) double-strand breaks at the same time there will be four broken ends in total. The NHEJ system proteins recognize the breaks but do not have the ability to identify exactly which ends should be joined to restore the original chromosome. Thus, NHEJ may join the ends together correctly, but if they do not the result is a **chromosome rearrangement**.

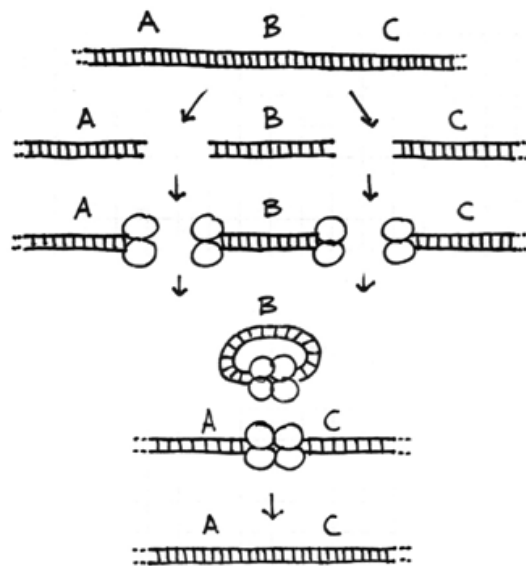


Figure 5.2.2: Errors during DNA repair can cause a chromosome deletion. In this diagram A, B, and C are genes on the same chromosome. There have been two breaks in the DNA backbone, recruitment of NHEJ proteins, and repair. After the repairs are completed the small piece of DNA with gene B is lost and 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. First double-strand DNA break occurs in one site; a second DNA break occurs and the NHEJ proteins mend the damage incorrectly by joining the ends.

There are four major types of rearrangements:

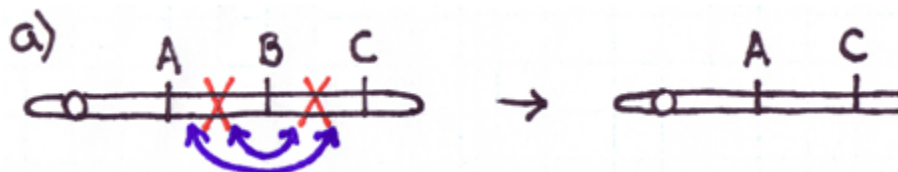


Figure 5.2.3a: **Deletions** arise when both breaks are on one chromosome. If the outer ends are joined the middle piece of DNA does not have a centromere and will be lost during the next cell division, losing that genetic material permanently.



Figure 5.2.3b: **Inversions** also occur when both breaks are on one chromosome when the fragment between the two breaks is joined to opposite chromosome ends. 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). No genetic material has been lost in this rearrangement.

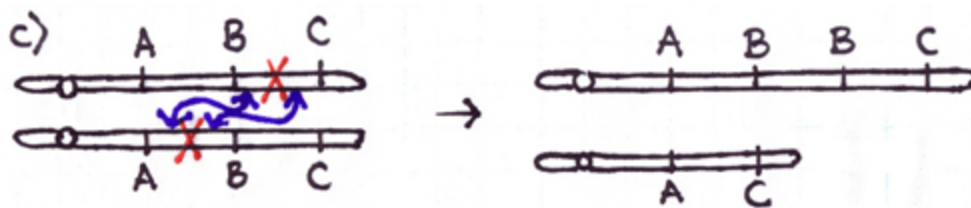


Figure 5.2.3c: **Duplications** can occur from when DNA breaks at different places in sister chromatids. If the ends are joined to the other chromatids, one chromosome will have a duplication and one will have a deletion.

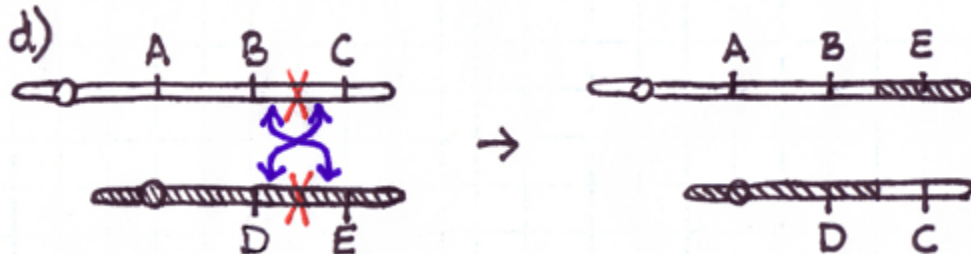


Figure 5.2.3d: **Translocations** result from two breaks on non-homologous chromosomes and incorrect rejoining. In a **reciprocal translocation**, two chromosomes have 'swapped' arms, and traded genetic information, but no genetic information is lost. **Robertsonian translocations** are those rare situations 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. They also allow recombination to occur between linked genes. 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. 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. 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.

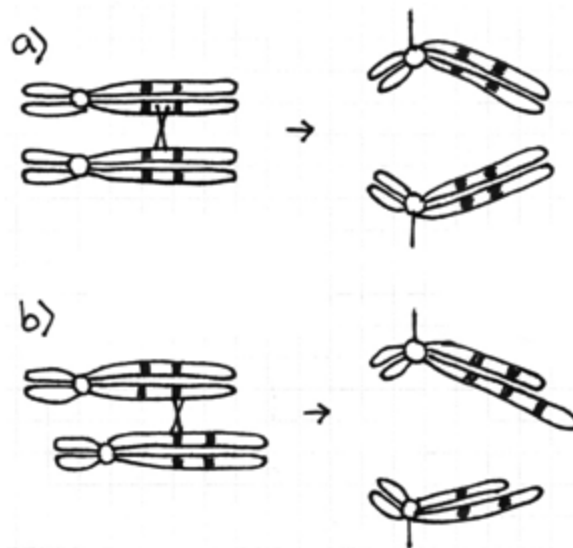


Figure 5.2.4: Errors during meiotic crossovers can cause duplications and deletions. This diagram shows homologous chromosomes pairing in prophase I and then separating in anaphase I. The shaded boxes are alu transposable elements. a) The homologous chromosomes pair properly, a crossover occurs, and 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: Abnormal pairing at Meiosis

DNA forms loops to achieve pairing when chromosomes are rearranged

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 feature can be used to locate the deletion cytologically. The deleted region is also **pseudo-dominant**, in that the cell no longer has two copies of the genes in this region, therefore mutant recessive alleles may produce a phenotypic change.

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 dicentric product that will break into deletion products and produce unbalanced gametes. Similarly, with a **pericentric inversion**, a crossover event leads to duplicate/deletion products that are unbalanced.

If unbalanced gametes are joined with a normal gamete at fertilization, they will result in an unbalanced zygote, which are usually lethal. As a consequence, crossover products (recombinants) are lost and the offspring do not appear in the next generation. Thus, if the number of recombinants was determined it would be a lower number than expected with inversions appearing 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.

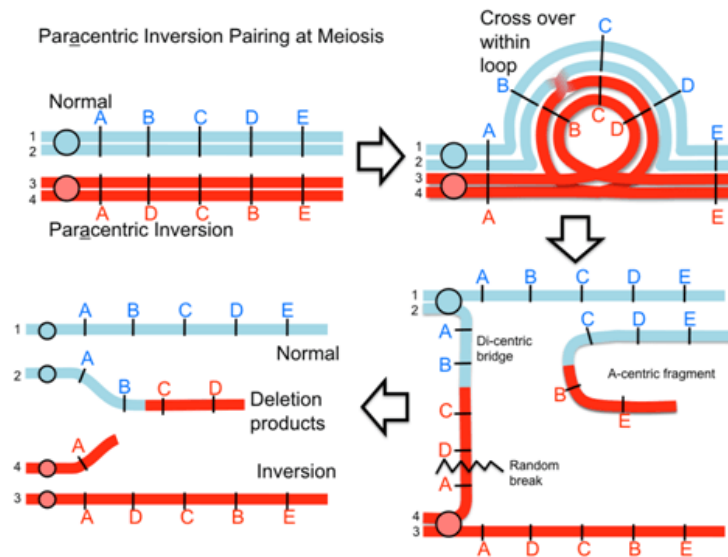


Figure 5.2.5: A paracentric inversion pairing at meiosis. A crossover within the loop causes the production of an acentric and a dicentric chromatids which leads to deletion product.(Original-Locke-CC:AN)

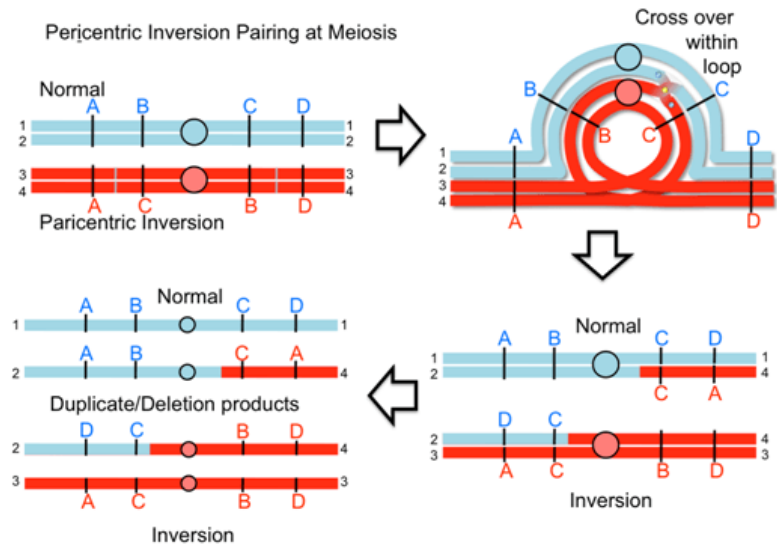


Figure 5.2.6: A pericentric inversion pairing at meiosis. A crossover within the loop causes the production of duplicate and deletion products. (Original-Locke-CC:AN)

Video explanation: Crossing over in inversions

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



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

Translocated chromosomes pair with more than one homologous chromosome

For **translocations**, a consequence for the two chromosomes involved is that when they pair at meiosis both replicated chromosome pairs will be together. During meiosis I, homologous chromosomes separate. The four chromosomes can segregate in different ways, depending on which centromeres are pulled to each pole during anaphase. This set of paired, replicated chromosomes can segregate as Alternate (balanced) where both non-translocated (N1 and N2) chromosomes go to one pole and both translocated chromosomes go to the other pole. A gamete that contains the two translocated (T1 and T2) chromosome still contains one full copy of both chromosomes. However, the chromosomes may segregate as Adjacent-1 (unbalanced) where a cell gets one normal and one translocation chromosome; each gamete will contain extra copies of some genes and lack copies of other genes. Each of these possibilities occur with approximately equal frequency and thus about half the gametes will be unbalanced.

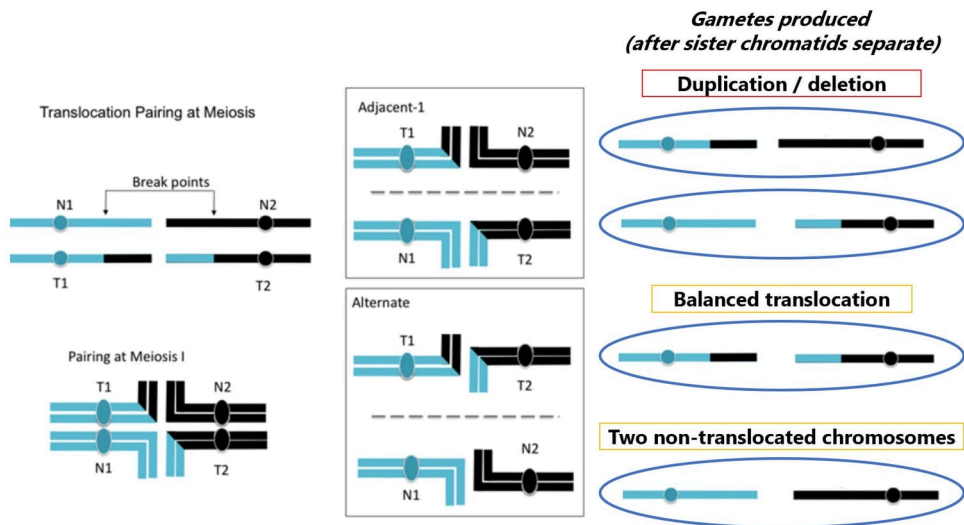


Figure 5.2.6: A reciprocal translocation pairing at meiosis. There are two main avenues for segregation: Adjacent-1 and Alternate. Adjacent-1 results in duplication and deletion for part of the chromosome segments. Alternate segregation . (Modified by SWL from Original-Locke-CC:AN)

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.

Note about gene dosage

Having too much or too little gene activity 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. There are different ways they might pair during prophase I, 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**.

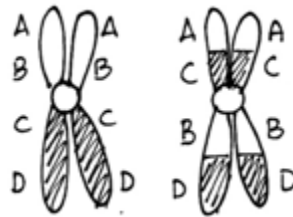


Figure 5.2.7: An unarranged chromosome (left) and a homolog with a pericentric inversion (right). (Original-Harrington-CC:AN)

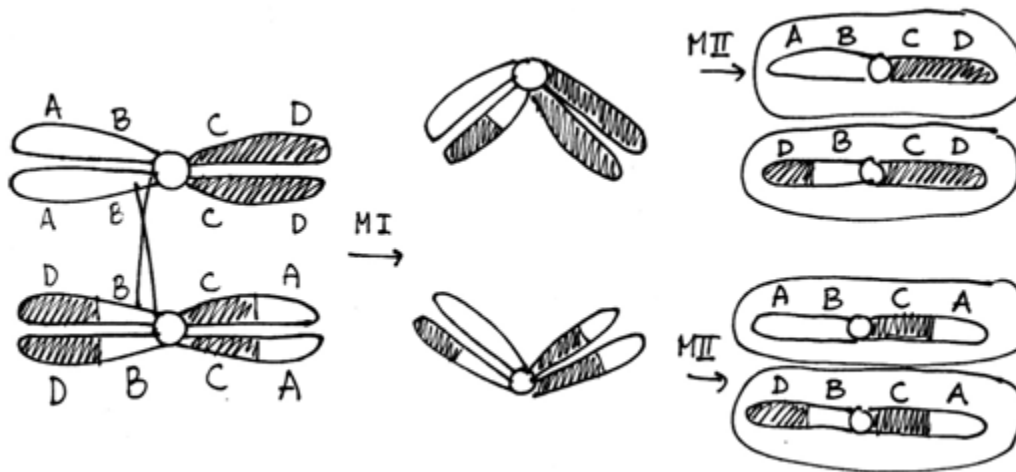


Figure 5.2.8: Meiosis in a cell heterozygous for the chromosomes shown in Figure 9.11. Note that of the four gametes one has a deletion of the A gene and a duplication of the D gene while another gamete has a duplication of A and a deletion of D. (Original-Harrington-CC:AN)

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, that result in a lot many unbalanced gametes 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. One example of such a translocation t(9;22) is the cause of chronic myelogenous leukemia (CML). Association of this translocation in cancer was first suggested in a brief letter 1960 (Nowell and Hungerford, *Science* 132:1497). The translocation creates a fusion gene with the BCR gene promoter and 5' region and the ABL 3' end. When transcribed and translated, this fusion protein is a constitutively active kinase.

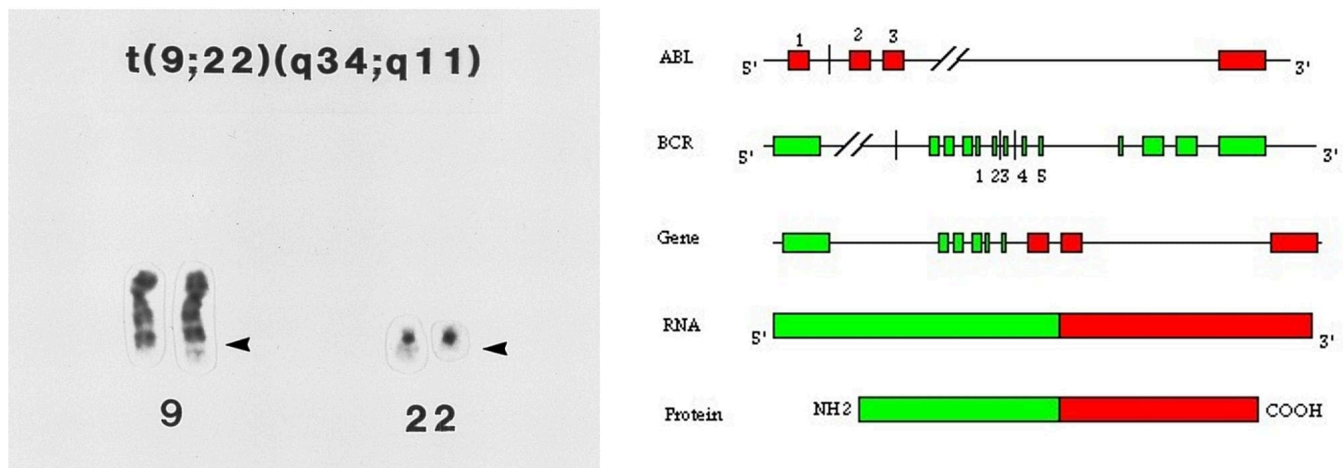


Figure 5.2.9: Left: Translocation between chromosomes 9 and 22, known as the Philadelphia chromosome (The Armed Forces Institute of Pathology (AFIP) / Public domain) Right: The translocation breaks and joins two genes that then produces a fusion protein known as BCR-ABL. (WikimediaCommons Creative Commons 3.0).

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. For example, multiple copies of the globin genes are found in mammals and have unique functions.

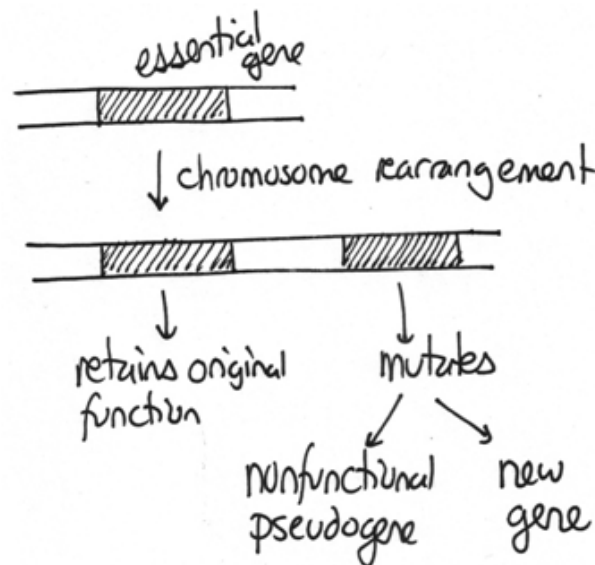


Figure 5.2.9: Duplicated genes can mutate without compromising the viability of the organism. Occasionally the result is a new gene. (Original-Harrington-CC:AN)

Chromosome rearrangements that decrease fertility are also important for the origin of new species. If a rearrangement 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.

Contributors and Attributions

- [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 [5.2: Changes in Chromosome Structure](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [5.2: Changes in Chromosome Structure](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

5.3: Chapter 5 Review Questions (Draft)

5.1: Changes in Chromosome Number

Query 5.3.5.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 5.3.5.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 5.3.5.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 5.3.5.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



5.2: Changes in Chromosome Structure

Query 5.3.5.5

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



6: 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**. Prokaryotes and other unicellular organisms regulate genes too, changes in temperature or chemicals in the environment can require changing what proteins are produced at certain times. 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.

6.1: Prokaryotic gene regulation

6.1.1: The Use of Mutants to Study the lac Operon

6.1.2: The Lux Operon

6.2: Eukaryotic Gene Regulation

6.2.1: Gene Expression in Evolution

6.2.2: Regulating Transcription at the Chromatin Level

6.2.3: microRNAs

6.3: Epigenetics

6.4: Chapter 6 Review Questions (Draft)

6.5: Chapter 6 Review Questions Draft

Contributors and Attributions

- [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 [6: Regulation of Gene Expression](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [6: Regulation of Gene Expression](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

6.1: Prokaryotic gene regulation

Learning Objectives

- Define operon and describe how it differs from genes organized by single proteins.
- Identify the functions of regulatory proteins, promoters, operators, *cis*- and *trans*-acting factors, and structural genes.
- Distinguish between positive vs negative and inducible vs repressible operons.
- Identify and understand the role of the structure and components of the *lac* operon and predict whether or not gene expression will occur under particular conditions.

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 (though there are some exceptions including *C. elegans* and a few other species).

How can an operon encode multiple proteins?

Think back to the steps involved in transcription and translation. Which steps will be altered if multiple proteins must be translated from a single transcript? How do you think this works?

Types of regulation

Operons can be turned "off" (repressible) or "on" (inducible) and be controlled protein activators or repressors.

	Inducible regulation turns transcription ON.	Repressible regulation turns transcription OFF.
Positive regulation uses activator proteins.	An activator protein turns ON transcription in response to a stimulus / condition.	An activator protein is inactivated in response to a stimulus / condition, turning transcription OFF.
Negative regulation uses repressor proteins.	A repressor protein is inactivated in response to a stimulus / condition, turning transcription turns ON.	A repressor protein turns transcription turns OFF in response to a stimulus / condition.

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*. *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. 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 6.1.1: 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. (Original-Deyholos-CC:AN)

cis- and *trans*- Regulatory Elements

The *lac* operon includes DNA sequences that do not encode proteins, but are instead binding sites for proteins that regulate transcription 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, encoded by a gene distinct from the operon that encodes *lacZ*, *lacY*, and *lacA*, assemble together to form a homotetramer called a repressor. 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. Therefore, the operon will not be transcribed when the operator is occupied by a repressor.

Allosteric regulation occurs when a substrate causes a change to the structure and function of a protein. In this example, in addition to domains that bind operator sequences in DNA, *lacI* protein has sites that bind to allolactose, which is a small molecule produced when lactose is present. When allolactose is bound to *lacI*, the shape of the protein changes and 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.

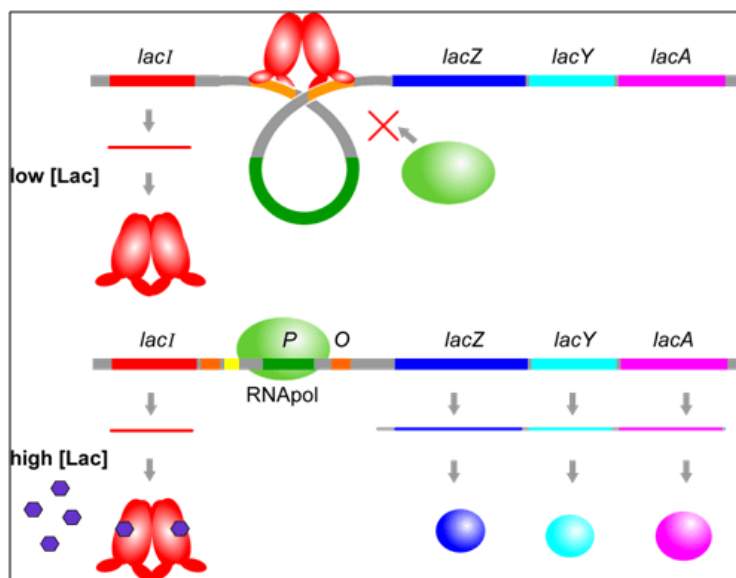


Figure 6.1.2: When the concentration of lactose [Lac] is low, *lacI* tetramers bind to operator sequences (O), thereby blocking binding of RNAPol (green) to the promoter (P). Alternatively, when [Lac] is high, allolactose binds to *lacI*, preventing the repressor from binding to O, and allowing transcription by RNAPol. (Original-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). 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.

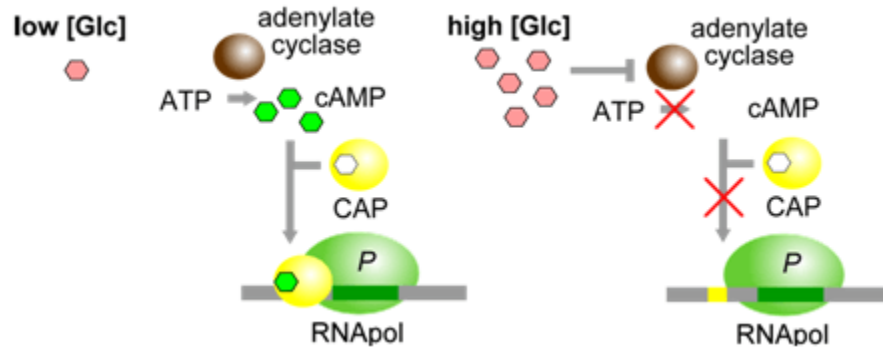


Figure 6.1.3: CAP, when bound to cAMP, helps RNAPol to bind to the lac operon. cAMP is produced only when glucose [Glc] is low. (OrigianI-Deyholos-CC:AN)

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 maximal levels.

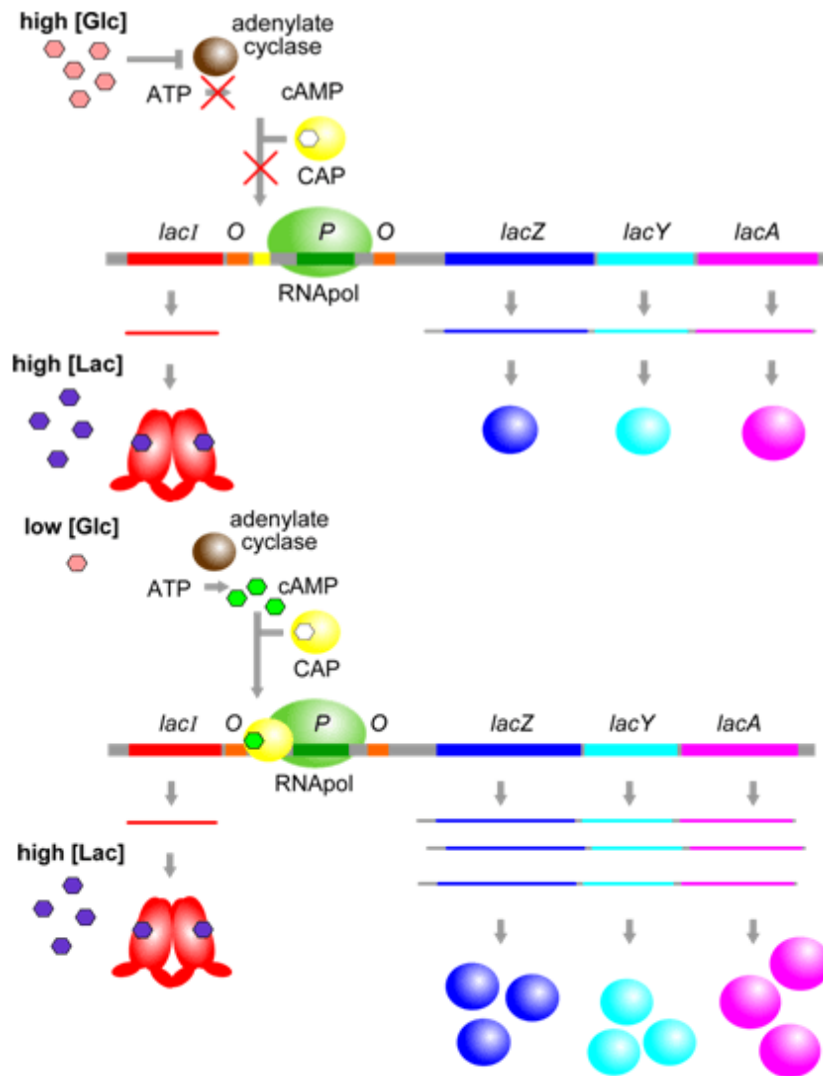


Figure 6.1.4: When glucose [Glc] and lactose [Lac] are both high, the lac operon is transcribed at a moderate level, because CAP (in the absence of cAMP) is unable to bind to its corresponding cis-element (yellow) and therefore cannot help to stabilize binding of RNAPol at the promoter. Alternatively, 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)

Video Summary of the *lac* operon

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 6.1.1: Prokaryotic Gene Regulation: Lac Operon. (AMBOSS: Medical Knowledge Distilled via <https://youtu.be/-4Jr84qyKPo>)

Contributors and Attributions

- [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 [6.1: Prokaryotic gene regulation](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [6.1: Prokaryotic gene regulation](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

6.1.1: The Use of Mutants to Study the *lac* Operon

Learning Objectives

- Describe how partial diploids can be used to examine genetic relationships in haploid organisms.
- Predict whether *lac* operon genes will be expressed when given mutations are present.

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 O^C , for operator constitutive, 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 transcribed whether or not lactose is present.

The *lacI* locus - One type of mutant allele of *lacI* (called I^-) prevents either the production of a repressor polypeptide or produces a polypeptide that cannot bind to the operator sequence. This allele is also a constitutive expresser of the *lac* operon because absence of repressor binding permits transcription.

Another type of mutant of *lacI*, called *lacI^S* for super-repressor, prevents the repressor polypeptide from binding allolactose, and thus will always bind to the operator and be non-inducible. This mutant constitutively represses the *lac* operon whether lactose is present or not.

Two *lac* operons in a single cell create partial diploids in *E.coli*

The regulation of the *lac* operon became further understood by using two copies of the operon sequences in one cell. Although *E. coli* are haploid organisms, scientists use a plasmid called the F-factor to bring an additional copy of gene sequences into the cell, while the other copy is on the genomic *E. coli* chromosome. This results in a partial diploid in *E. coli*.

The F-factor is extra-chromosomal DNA 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 *lacI^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 *lacI^S* is dominant to *lacI⁺*, which in turn is dominant to *lacI⁻*. It can also be used to show the O^C mutation only acts in *cis*- while the *lacI* mutation can act in *trans*-.

Example 6.1.1.1

Consider the question posed above: A cell with a copy of the *lac* operon from an F-factor has the *I^S* mutation, whereas the genomic copy has an O^C mutation. How would this cell respond to the presence/absence of lactose (or glucose)?

Solution

The I^S mutation encodes a form of the protein cannot bind allolactose and therefore continues to recognize the operator, even when lactose is present. However, the genomic copy of the operator O^C is mutated to a sequence that is no longer recognized by the *lac* repressor. For this reason, the *lacZYA* genes in the genomic *lac* operon will be continuously expressed in this cell, because *lacI* repressor will not be able to bind this sequence.

Query 6.1.1.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Contributors and Attributions

- [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 [6.1.1: 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 [Stefanie West Leacock](#).

- [6.1.1: The Use of Mutants to Study the lac Operon](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

6.1.2: The Lux Operon

Learning Objectives

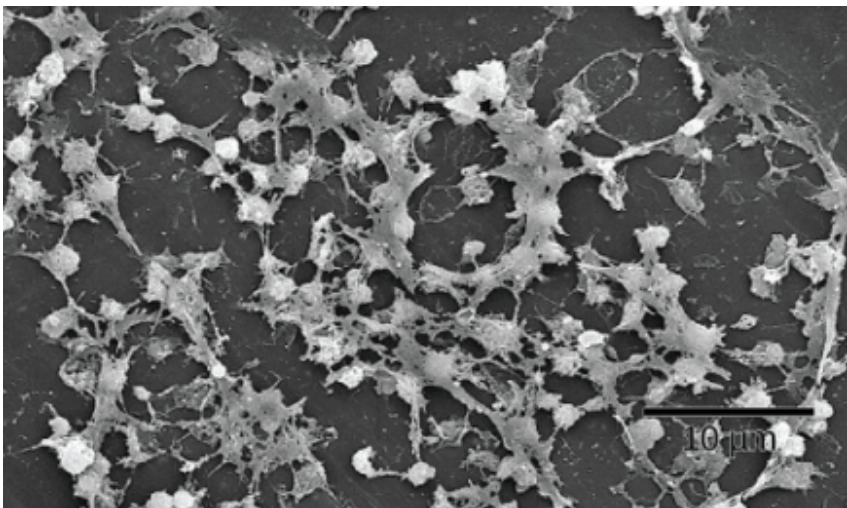
- Define quorum sensing.
- Explain the roles of regulatory proteins and autoinducers in transcriptional regulation by quorum sensing.
- Predict the effects of mutations in quorum sensing genes.

Quorum Sensing

Signaling in bacteria, known as quorum sensing, enables bacteria to monitor extracellular conditions, ensure sufficient amounts of nutrients are present, and avoid hazardous situations. There are circumstances, however, when bacteria communicate with each other.

The first evidence of bacterial communication was observed in a bacterium that has a symbiotic relationship with Hawaiian bobtail squid. When the population density of the bacteria reached a certain level, specific gene expression was initiated: the bacteria produced bioluminescent proteins that emitted light. Because the number of cells present in the environment (the cell density) is the determining factor for signaling, bacterial signaling was named quorum sensing. In politics and business, a quorum is the minimum number of members required to be present to vote on an issue.

Some species of bacteria that use quorum sensing form biofilms, which are complex colonies of bacteria (often containing several species) that exchange chemical signals to coordinate the release of toxins that attack the host. Bacterial biofilms can sometimes be found on medical equipment. When biofilms invade implants, such as hip or knee replacements or heart pacemakers, they can cause life-threatening infections.



(a)



(b)

Figure 6.1.2.1: Bacterial Biofilms: Cell-cell communication enables these (a) *Staphylococcus aureus* bacteria to work together to form a biofilm inside a hospital patient's catheter, seen here via scanning electron microscopy. *S. aureus* is the main cause of hospital-acquired infections. (b) Hawaiian bobtail squid have a symbiotic relationship with the bioluminescent bacteria *Vibrio fischeri*. The luminescence makes it difficult to see the squid from below because it effectively eliminates its shadow. In return for camouflage, the squid provides food for the bacteria. Free-living *V. fischeri* do not produce luciferase, the enzyme responsible for luminescence, but *V. fischeri* living in a symbiotic relationship with the squid do. Quorum sensing determines whether the bacteria should produce the luciferase enzyme. (9.4: Signaling in Single-Celled Organisms)

Autoinducers in Quorum Sensing

Quorum sensing uses autoinducers as signaling molecules. Autoinducers are signaling molecules secreted by bacteria to communicate with other bacteria of the same kind. The secreted autoinducers can be small, hydrophobic molecules, such as acyl-homoserine lactone (AHL), or larger peptide-based molecules. Each type of molecule has a different mode of action. When AHL enters target bacteria, it binds to transcription factors, which then switch gene expression on or off. The peptide autoinducers stimulate more complicated signaling pathways that include bacterial kinases. The changes in bacteria following exposure to autoinducers can be quite extensive. The pathogenic bacterium *Pseudomonas aeruginosa* has 616 different genes that respond to autoinducers.

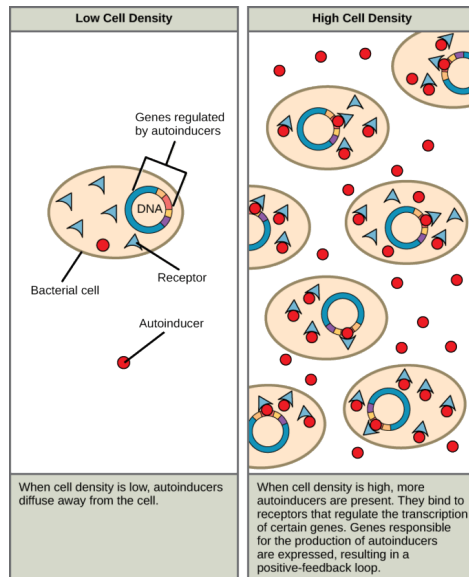


Figure 6.1.2.2: Autoinducers are small molecules or proteins produced by bacteria that regulate gene expression. (9.4: Signaling in Single-Celled Organisms)

The *Lux* operon is regulated by Quorum Sensing

Bioluminescence is produced when enzymes catalyze a reaction that produces light as a by-product. The *Lux* operon in *Vibrio fischeri* includes two proteins that form an enzyme that catalyze this reaction. These bacteria live in specialized light organs in the Hawaiian bobtail squid and produce light to help the squid camouflage itself in the moonlight. However, the transcription of the genes encoding enzymes only occurs when cells are at a high density because a high concentration of autoinducer will be present.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 6.1.2.1: Copy and Paste Caption here. (Copyright; author via <https://www.youtube.com/watch?v=5D7fumyGv-M>)

Two proteins control the amount of autoinducer and transcription of the operon. LuxI is the protein that produces the autoinducer and LuxR is the transcriptional activator that is activated in response to the autoinducer. Thus, this operon is an example of positive inducible regulation in which LuxR recruits RNA polymerase to transcribe the the operon with *LuxICDABE* when the concentration of bacteria (and thus autoinducer) is high. The proteins encoded by *LuxA* and *LuxB* form the luciferase enzyme that

catalyzes the light-producing reaction.

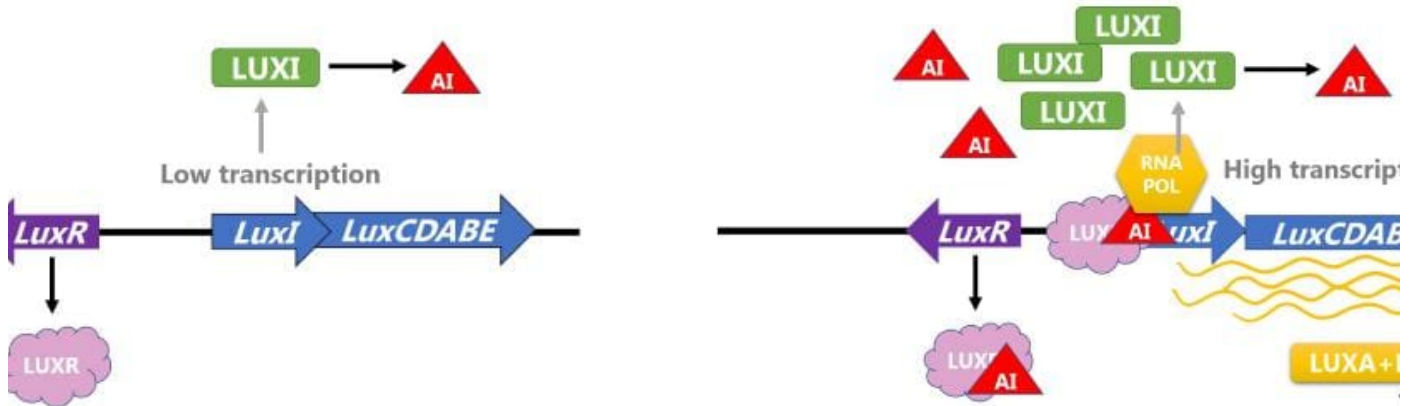


Figure 6.1.2.3: Regulation of the Lux operon of *V. fischeri* by LUXR and bacterial density. (left) When concentration of bacteria is low, the transcription of the operon is low, levels of LUXI are low, and concentration of autoinducer (AI) is low. (right) When concentration of bacteria is high, concentration of autoinducer (AI) is high and AI binds LUXR to increase transcription of the operon. LuxA and LuxB proteins are produced and catalyze the light reaction.

Query 6.1.2.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Additional videos about the *Lux* operon and quorum sensing

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 6.1.2.2: The LUX Operon (HHMI BioInteractive Video via <https://youtu.be/Pf7veBzQ8Gs>)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 6.1.2.3: Tiny Conspiracies - Bonnie Bassler (Princeton/HHMI) (Copyright ibiology.org via <https://youtu.be/Ie4Y4CDDuo>)

Contributors and Attributions

- Source: **9.4: Signaling in Single-Celled Organisms**

References

Engbrecht J, Silverman M. Nucleotide sequence of the regulatory locus controlling expression of bacterial genes for bioluminescence. *Nucleic Acids Res.* 1987 Dec 23;15(24):10455-67. doi: 10.1093/nar/15.24.10455. PMID: 3697093; PMCID: PMC339955.

This page titled **6.1.2: The Lux Operon** is shared under a license and was authored, remixed, and/or curated by Cheneir Neely.

- **6.1.2: The Lux Operon** has no license indicated.

6.2: Eukaryotic Gene Regulation

Learning Objectives

- Describe the *cis*-acting elements that control tissue-specific transcription of eukaryotic protein encoding genes: promoters, enhancers, insulators and how *trans*-acting transcription factors can influence gene expression.
- Predict the effects of mutations in regulatory sequences on gene expression.

Like prokaryotes, transcriptional regulation in eukaryotes involves both *cis*-elements and *trans*-factors; however, there are more factors and they interact in more complex ways, many of which are still be elucidated.

Regulatory sequences

Proximal regulatory sequences

As in prokaryotes the RNA polymerase binds to the gene at its promoter to begin transcription. In eukaryotes, however, RNA polymerase is part of a large protein complex that includes additional proteins (transcription initiation factors) 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 is described relative to the transcription start site (+1), but the distance between any of these elements and the transcription start site can vary. These sites are typically within ~200 base pairs of the start of transcription.

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 DNA sequences called enhancer elements or simply enhancers. Regulatory *trans*-factor proteins called transcription factors bind to specific enhancer sequences, then, while still bound to DNA, these proteins interact with RNA polymerase 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. Enhancers can be located near (~100s of bp) or far (~10K of bp), and either upstream or downstream, from the promoter. The transcription factors and the proteins associated with transcription initiation complex will bind each other, bending the DNA in the process.

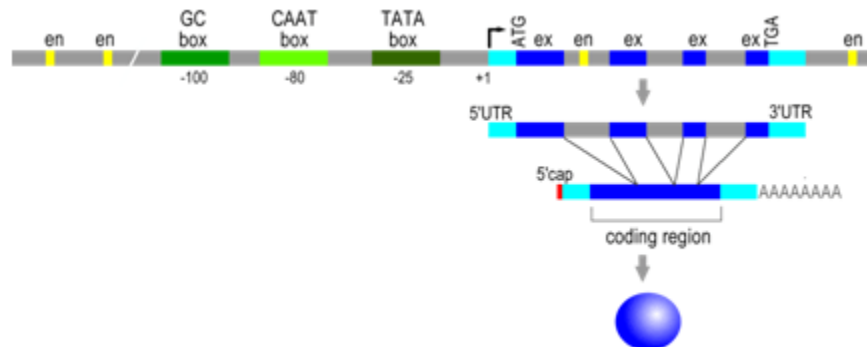


Figure 6.2.1: Structure of a typical eukaryotic gene. RNA polymerase binding may involve one or more *cis*-elements within the proximal region of a promoter (green boxes). Enhancers (en, yellow boxes) may be located any distance upstream or downstream of the promoter and are also involved in regulating gene expression. The processing of a primary transcript to a mature mRNA is also shown. (Original-Deyholos-CC:AN)

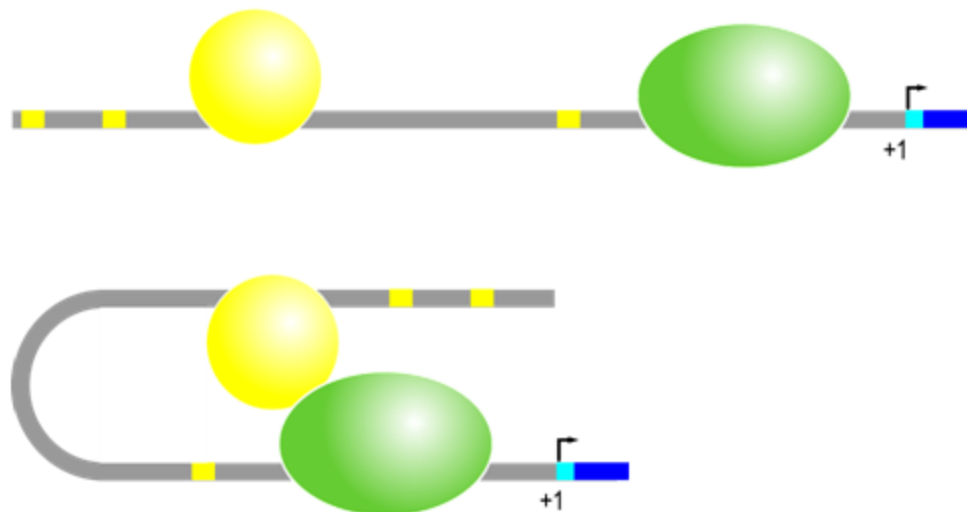


Figure 6.2.2: A transcription factor (yellow) bound to an enhancer that is located far from a promoter. Because of the flexibility of the DNA molecule, the transcription factor and RNA polymerase (green) are able to interact physically, even though the cis-elements to which they are bound are located far apart. In eukaryotic cells, RNA polymerase is actually part of a large complex of proteins (not shown here) that assembles at the promoter. (Original-Deyholos-CC:AN)

Enhancer sequence example

Some genes are transcriptionally regulated in response to hormones, such as estrogen. Estrogen receptors are proteins that bind the hormone estrogen and the protein dimers are transported into the nucleus where they recognize sequences called estrogen response elements (EREs). The consensus ERE sequence is 5'-GGTCAnnnTGACC-3' (where n is any base) at many locations in a genome and recruit RNA polymerase to induce the transcription of the corresponding genes. This consensus sequence was elucidated by studies of a human cell line to which a *Xenopus* (frog) estrogen responsive gene had been added, revealing that the roles of estrogen receptors in binding EREs and activating transcription of target genes is conserved across species.

Klein-Hitpass L, Schorpp M, Wagner U, Ryffel GU. An estrogen-responsive element derived from the 5' flanking region of the *Xenopus* vitellogenin A2 gene functions in transfected human cells. *Cell*. 1986 Sep 26;46(7):1053-61. doi: 10.1016/0092-8674(86)90705-1. PMID: 3463433.

Reviewed by Klinge C. *Nucleic Acids Res*. 2001 Jul 15; 29(14): 2905–2919.

Enhancers regulate the *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 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.) Three enhancer elements each bind a different tissue-specific transcription factor in the wing, body, or mouth to enhance transcription of *yellow* in that tissue and make the dark 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.

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 repress or reduce transcription from the adjacent gene. A gene's complete expression profile (transcription level, tissue specific, temporal specific) is a combination of various enhancer and silencer elements



Figure 6.2.3: Tissue-specific cis-regulatory elements within a simplified representation of the yellow gene of *Drosophila*. (Origanl-Deyholos-CC:AN)

Thinking about regulation of transcription factors

Understanding the transcription factors that regulate the expression of the yellow gene can prompt us to consider: why are those transcription factors present in different tissues? These transcription factors themselves are regulated by their own sets of transcription factors and so on and so on. Where does it begin? For *Drosophila*, the answer is in the egg! Specific proteins and mRNAs in the egg define the regions of the future embryo and a cascade of gene expression events lead to each adult cell and tissue type expressing the correct genes for their function.

Exercise 6.2.1

What is the phenotype of the fruit fly when the *yellow* gene and enhancers are fully functional?

What would be the effect of a loss-of-function mutation in the *yellow* coding sequence?

What would be the effect of a mutation in the wing enhancer?

What would be the effect of a mutation in the TATA box region?

Answer available at the end of this section.

Contributors and Attributions

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

Using reporter genes to study transcription

Answers

Exercise 6.2.1

The wild-type fruit fly has a dark black/brown cuticle.

A loss-of-function mutation in the *yellow* gene would lead to flies with yellow appearance, instead of brown/black.

A mutation in the wing enhancer would result in yellow wings, but other parts of cuticle should be unaffected.

A mutation in the TATA box should result in a phenotype similar to loss-of-function in the *yellow* gene.

This page titled 6.2: Eukaryotic Gene Regulation is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stefanie West Leacock.

- 6.2: Eukaryotic Gene Regulation by Stefanie West Leacock is licensed CC BY-SA 4.0.

6.2.1: Gene Expression in Evolution

Learning Objectives

- Using examples, describe how changes in gene expression can be associated with changes in phenotype and 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.

Pitx expression in Stickleback

The three-spined stickleback 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. 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 (the cells that will generate the fin) 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*.

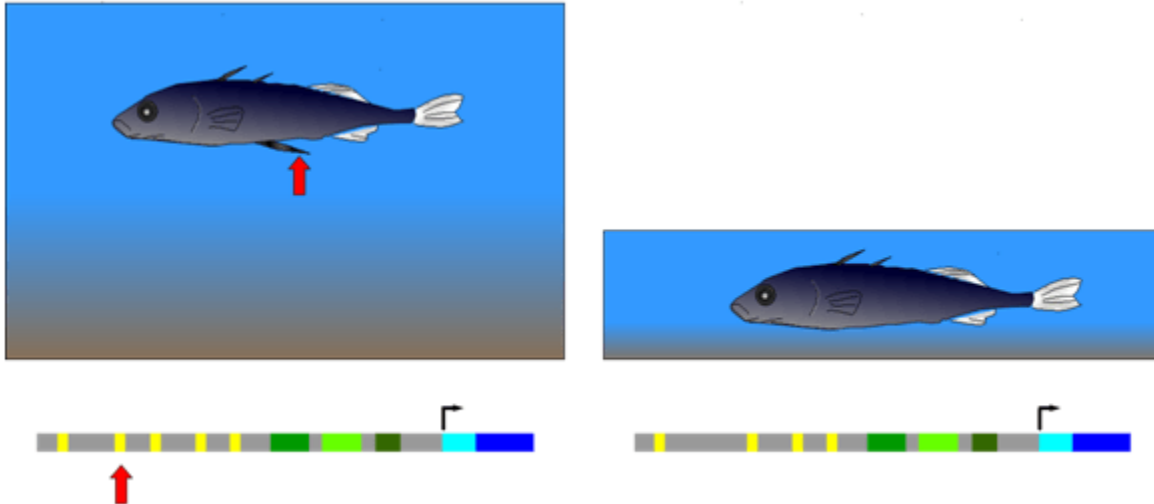


Figure 6.2.1.1: Development of a large, spiny pelvic fin in deep-water stickleback (left) depends on the presence of a particular enhancer element upstream of a gene called *Pitx*. Mutants lacking this element, and therefore the large pelvic fin (right), have been selected for in shallow-water environments. (Wikipedia-Richard Wheeler-GFDL)

Thinking about the mutation

Consider the enhancer element mutated in the shallow-water sticklebacks:

- What is the function of that DNA sequence? What type of protein would bind there?
- Do you think this mutation acts in a dominant or recessive manner?
- Which of these processes are affected by the mutation: DNA replication, transcription, splicing, and/or translation?
- Would it be possible for another mutation to reverse the effects of this mutation and a shallow-water stickleback with a long fin?

Answer available at the end of this section.

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. 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.

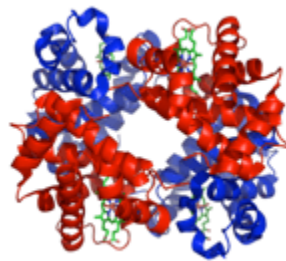


Figure 6.2.1.2: A tetramer of human hemoglobin, type $\alpha_2\beta_2$. The α chains are labeled red, and the β chains are labeled blue. Heme groups are green. (Wikipedia—from Wood, W.G. 1976 Br. Med. Bull. 32, 282-CC:AS)

The composition of the hemoglobin tetramers changes during development. From early childhood onward, most tetramers are of the type $\alpha_2\beta_2$, which means they contain 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.

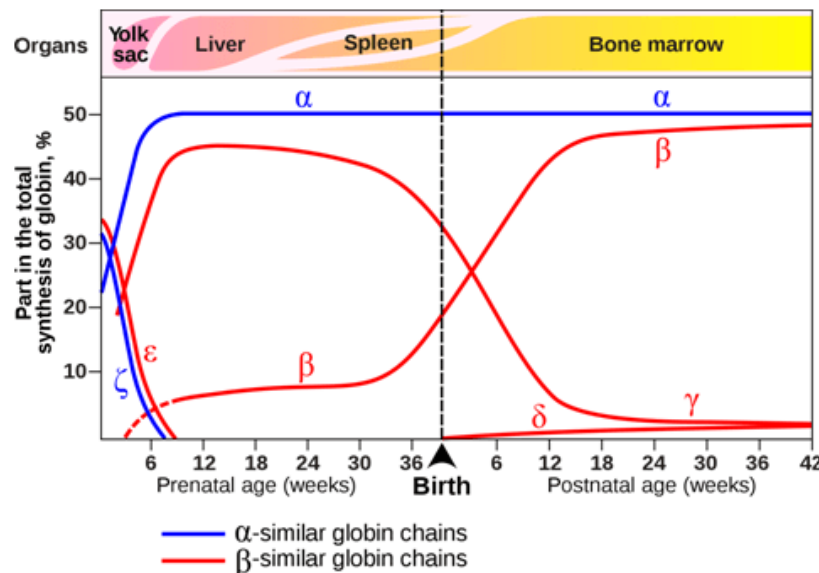


Figure 6.2.1.3: Expression of globin genes during prenatal and postnatal development in humans. The organs in which globin genes are primarily expressed at each developmental stage are also indicated. (Origianl-Deyholos-CC:AN)

Each of these globin polypeptides is encoded by a different gene. In humans, globin genes are located in clusters on two chromosomes. 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 (5.2: Changes in Chromosome Structure). The duplicated genes can then 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.

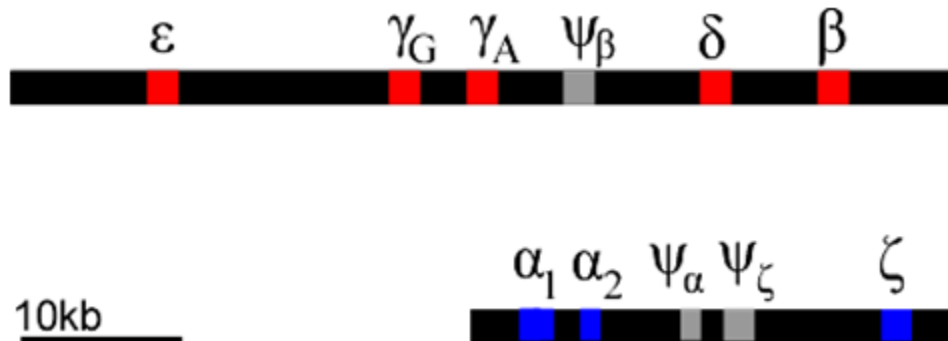


Figure 6.2.1.4: Fragments of human chromosome 11 and human chromosome 16 on which are located clusters of β -like and α -like globin genes, respectively. Additional globin genes (θ , μ) have also been described by some researchers, but are not shown here. (Origianl-Deyholos-CC:AN)

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.

Exercise 6.2.1.1

Individuals with diseases such as sickle cell disease or β -thalassemia have mutations that either cause a malformed protein or the absence of the HBB (beta-globin) protein. How could understanding the normal expression of other hemoglobin proteins help develop new therapies for these patients? What tools and processes might be used?

Answer available at the end of this section.

Contributors and Attributions

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

Answers

Thinking about the mutation

Exercise 6.2.1.1

Recent clinical trials are exploring the possibility of using gene editing to express fetal hemoglobin in sickle cell and beta-thalassemia patients. This process isolates stem cells from patients, edits the DNA *in vitro*, and then transplants the edited cells back into the patient (reviewed in <https://academic.oup.com/hmg/advance-article/doi/10.1093/hmg/ddaa088/5836961>).

A news report about one of these patients is at this link <https://www.npr.org/sections/health-shots/2020/06/23/877543610/a-year-in-1st-patient-to-get-gene-editing-for-sickle-cell-disease-is-thriving>.

Selected references:

Ye L, Wang J, Tan Y, et al. (2016) Genome editing using CRISPR-Cas9 to create the HPFH genotype in HSPCs: An approach for treating sickle cell disease and

β -thalassemia. *Proc Natl Acad Sci U S A*. 2016;113(38):10661-10665. doi:10.1073/pnas.1612075113 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5035856/>)

Weber L, Frait G, Felix T, et al. (2020) Editing a γ -globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. *Science Advances* 12 Feb 2020 (https://advances.sciencemag.org/content/6/7/eaay9392?utm_source=TrendMD&utm_medium=cpc&utm_campaign=TrendMD_1)

Demirci S, Leonard A, Tisdale JF. (2020) Genome editing strategies for fetal hemoglobin induction in beta-hemoglobinopathies. *Human Molecular Genetics* 14 May 2020. <https://doi.org/10.1093/hmg/ddaa088>

This page titled [6.2.1: Gene Expression in Evolution](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [6.2.1: Gene Expression in Evolution](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

6.2.2: Regulating Transcription at the Chromatin Level

Learning Objectives

- Predict regions of the genome that are more likely to have “open” or “closed” chromatin and understand how acetylation affects chromatin.
- Identify the molecular nature of DNA methylation, the types of bases that are methylated, and the effect of DNA methylation on gene expression.

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 remodeling

Despite the simplified way in which we often represent DNA, 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 RNA polymerase and regulatory proteins. So, in regions with highly compact chromatin transcription is unlikely, 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, often the histone proteins in nucleosomes. These post-translational 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.

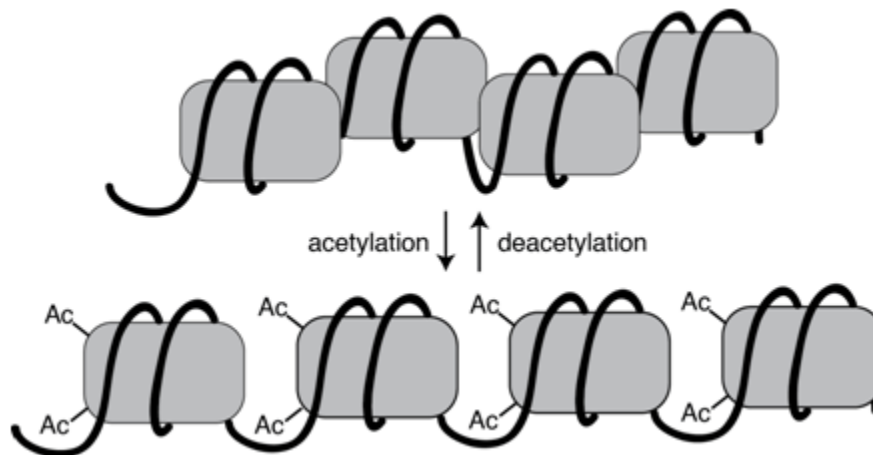


Figure 6.2.2.1: Acetylation of histone proteins is associated with more a more open chromatin configuration. Acetylation is a reversible process. (Origanl-Deyholos-CC:AN)

Likewise, **methylation** of DNA itself is also associated with transcription regulation. Cytosine bases, particularly when followed by a guanine are important targets for DNA methylation. Using a p to represent the phosphodiester bond between the nucleotides, these are often called **CpG sites**, and regions in which many CpG sites are nearby are known as CpG islands. Methylated cytosine within clusters of CpG sites is often associated with transcriptionally inactive DNA.

Thinking about CpG

Why is C followed by G an important site for DNA methylation? Think about the sequence of both strands of DNA. What is the complementary sequence of 5' ...CG... 3' ? Does this allow both strands to be methylated?

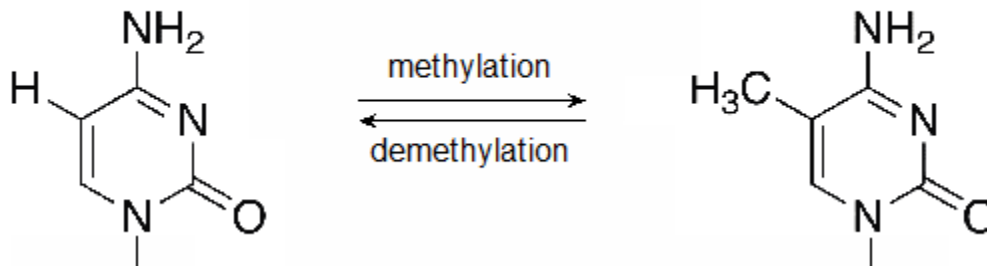


Figure 6.2.2.2: A methylation reaction produces 5-methylcytosine (5mC). Methyl groups may also be removed by DNA demethylases. (flickr-Beardy Git-CC:AND)

Common post-translational modifications to histones

These post-translational modifications are the result of enzymes that covalently link functional groups to amino acids at specific positions in histones.

- Acetylation
- Methylation
- Phosphorylation
- Ubiquitination

Combinations of these modifications was first postulated to be a "code" by which the regulation of genes is interpreted by David Allis and colleagues in 2001.

Heritability of Chromatin

The modification of DNA and its associated proteins is enzymatically reversible (acetylases/deacetylases; methyltransferases/demethylases) and thus a cyclical activity. However, when DNA is replicated, existing modifications on the parental strand will be added to the newly synthesized strand as well, maintaining the pattern in daughter cells until the modification is enzymatically removed. Regulation of these chemical modifications provides another layer through which eukaryotic cells control the transcription of specific genes.

Code readers

In addition to influencing an overall level of chromatin accessibility, modified histones may interact specifically with proteins that recognize their modifications to promote recruitment of additional proteins for functions in repair, replication, or transcription.

Contributors and Attributions

- [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 [6.2.2: Regulating Transcription at the Chromatin Level](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [6.2.2: Regulating Transcription at the Chromatin Level](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

6.2.3: microRNAs

Learning Objectives

- Define microRNA and siRNA.
- Briefly describe the mechanisms by which RNAi interference affects gene expression.
- Predict the effect of a miRNA or siRNA that is fully or partially complementary to an mRNA.

Introduction

MicroRNAs (miRNAs) are small (20-25 nucleotide) RNAs that are encoded by genomic loci. MicroRNAs were first discovered when scientists tried to identify the gene responsible for a heterochronic mutant (*lin-4*) *C. elegans* in the early 1990s. The mutant had been identified nearly two decades prior as one in which the normal pattern of larval molting did not stop, instead adult worms continued to repeat larval patterns, hence the name meaning cell lineage-defective. The mutation turned out to be in a non-protein-coding region that produced a small RNA with partial complementarity to seven sites in the 3' UTR of the *lin-14* RNA (Lee et al, 1993). Previous studies had shown that *lin-4* was a negative regulator of *lin-14*, so this result revealed that the mechanism involved an RNA acting by repressing the another RNA! Further work in *C. elegans* revealed that another gene, *let-7*, also encodes a microRNA that regulates multiple target RNAs via RNA-RNA interactions.



Figure 6.2.3.1: Potential RNA/RNA duplexes between *let-7* and *lin-41*. The position of the *let-7*(n2853) mutation is shown by an arrow below the duplexes. (Copyright CC BY Vella and Slack via http://www.wormbook.org/chapters/www_microRNA/microRNAs.html#bib28)

This work has led to the discovery of many more microRNAs in many different organisms. The current estimate of number of miRNAs in the human genome is about 2,300 (Alles et al., 2019).

Why are small RNAs important?

The discovery of two non-coding RNAs in *C. elegans* turned out to be just the beginning of our understanding of the prevalence and activity of small RNAs. Two other *C. elegans* biologists, Andrew Fire and Craig Mello, were studying another aspect of RNA activity at the same time. They were following up on some research that had used antisense RNAs to inhibit the expression of certain genes.

Antisense RNA: the theory

Remember that only one strand of a gene is transcribed into RNA.

The mRNA contains the codons needed to make the correct protein so it makes **SENSE**, but it was transcribed using the **ANTISENSE** strand as a template.

The **ANTISENSE** RNA does not usually exist in cells, but if you provide it, it will be complementary to the mRNA.

The theory was that this would prevent the translation of the RNA.

The experiments of Fire and Mello revealed that injecting double-stranded RNA (dsRNA = sense RNA + antisense RNA) cause a more potent response than injecting antisense alone. They also demonstrated that the injected dsRNA caused a decrease in mRNA levels, not just protein levels that was predicted by the antisense RNA idea. The phenomenon of a small RNA that targets a cellular mRNA for degradation is known as RNA interference, and these small RNAs are known as siRNAs (small interfering RNAs). These typically differ from microRNAs in that they have perfect complementarity to their target genes and cause their destruction.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 6.2.3.1: RNA interference (RNAi) is an important process, used by many different organisms to regulate the activity of genes. This animation explains how RNAi works and introduces the two main players: small interfering RNAs (siRNAs) and microRNAs (miRNAs). We take you on an audio-visual journey, diving into a cell to show how genes are transcribed to make messenger RNA (mRNA) and how RNAi can silence specific mRNAs to stop them from making proteins. The animation is based on the latest research, to give you an up-to-date view.. (RNA interference (RNAi): by Nature Video via https://youtu.be/cK-OGb1_ELE)

Mechanism Overview

The timely discovery of microRNAs and RNA interference by dsRNA revealed some shared mechanisms of action. Both microRNAs produced by transcription and siRNAs provided experimentally are cleaved into smaller (20–25 nt) fragments by the enzyme Dicer. The short pieces then are loaded to a RISC (RNA induced silencing complex), which contains a single-stranded small RNA while surveilling the cell for complementary sequences. When perfect (100%) complementarity is found, the target RNAs are cleaved; when imperfect complementarity is found, the outcome is often translation repression, although other effects are possible.

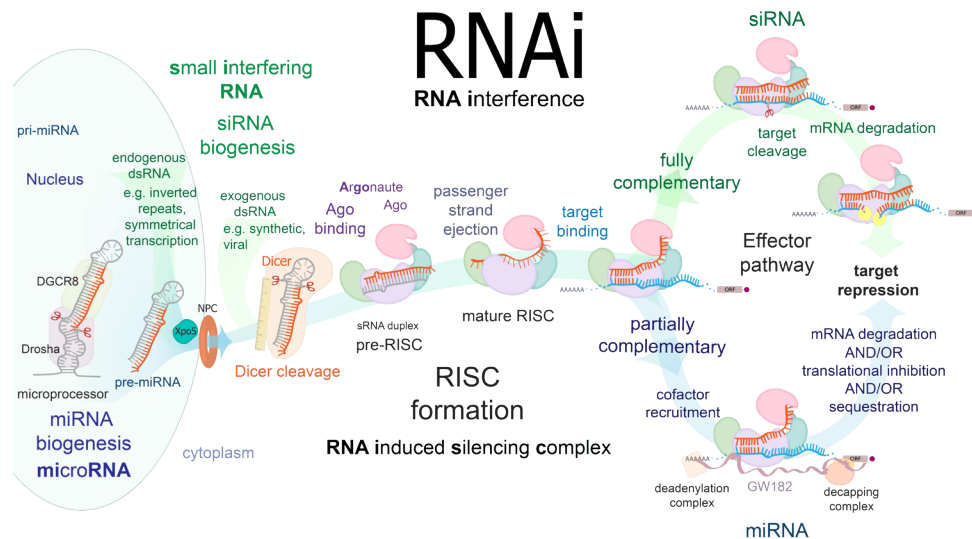


Figure 6.2.3.2: An infographic overview of RNA interference (RNAi). (Copyright CC BY SA; Biochemlife via An infographic overview of RNA interference (RNAi))

References

- Fire, A., Xu, S., Montgomery, M. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998). <https://doi.org/10.1038/35888>
- Julia Alles, Tobias Fehlmann, Ulrike Fischer, Christina Backes, Valentina Galata, Marie Minet, Martin Hart, Masood Abu-Halima, Friedrich A Grässer, Hans-Peter Lenhof, Andreas Keller, Eckart Meese, An estimate of the total number of true human miRNAs, *Nucleic Acids Research*, Volume 47, Issue 7, 23 April 2019, Pages 3353–3364, <https://doi.org/10.1093/nar/gkz097>
- Vella, M.C. and Slack, F.J. C. *elegans* microRNAs (September 21, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.26.1, <http://www.wormbook.org>.

6.3: Epigenetics

Learning Objectives

- Describe examples of gene regulation through DNA methylation and mechanisms of genomic imprinting. Explain how methyltransferases allow methylation to be heritable.
- Explain how DNA methylation can be associated with monoallelic expression.

The term **epigenetics** describes any heritable change in phenotype that is not associated with a change in 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 histone 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 in the chromosomal DNA sequence.

Heritability of epigenetic modifications

During DNA replication, parental strands are used as a template to create complementary strands. However, modified bases or histones are not copied directly. Instead, cells have enzymes that recognize partially modified DNA and add modifications to the newly synthesized DNA. One example of an enzyme that plays this role is DNMT1 (DNA methyltransferase 1). When chromosome with methylated DNA is replicated, the parental strand retains the methylation, but the new strand does not, resulting in hemi-methylated DNA. Recall that DNA methylation usually occurs at CpG sequences, such that a CpG also exists on the complementary strand at the same position. DNMT1 recognizes hemi-methylated DNA and adds the corresponding methyl group to the new strand.

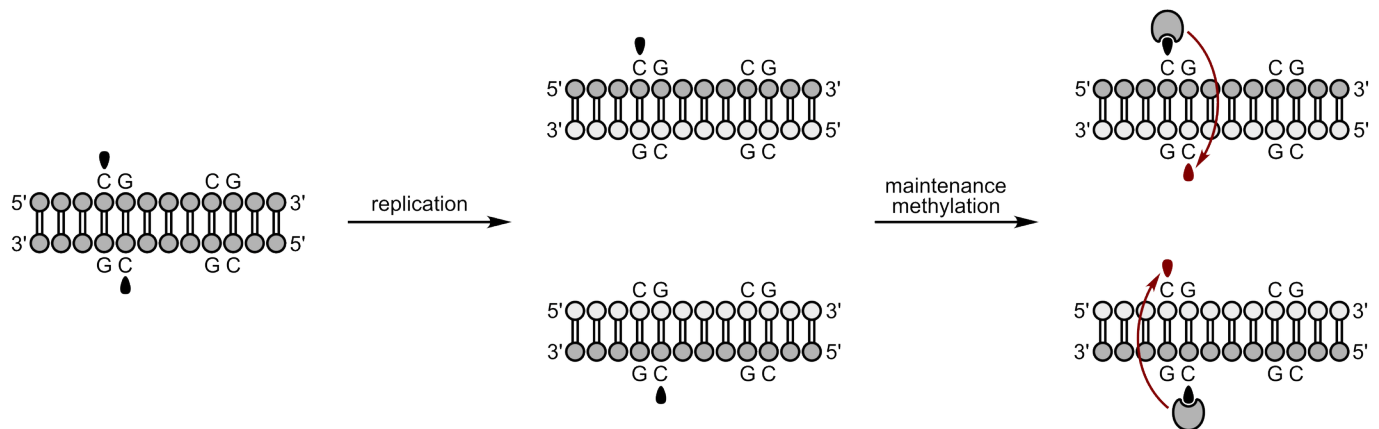


Figure 6.3.1: Scheme of **DNA maintenance methylation** by a maintenance methyltransferase such as DNMT1. The black pins represent methylation. CpGs that were methylated on both strands of the original DNA (left) will be methylated on both strands of the daughter DNA (right), whereas CpGs that were unmethylated in the original DNA will not be methylated in the daughter DNA. Maintenance methyltransferases recognize single-side methylated CpGs and methylate the complementary strand of those CpGs (red arrow). Reference: Latchman D (2010) *Gene Control* ISBN: 9780203833568. (Figure By Hbf878 - Own work, CC0, <https://commons.wikimedia.org/w/index.php?curid=73257041>)

Research Examples

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 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.

Epigenetics and maternal care

Watch this introductory video about experiments that tested the relationship between maternal care and offspring in rats.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



In experiments, Meaney and colleagues examined the influence of maternal licking on epigenetic modifications of genes involved in stress. Previous studies had shown that maternal care was associated with reduced fear and stress responses in adult rats. To determine whether this effect was due to changes in gene expression, epigenetic modifications near the glucocorticoid receptor (GR) gene were examined. In rats raised by mothers exhibiting high licking-grooming behaviors, methylation of a CpG in the GR was absent and the region was associated with histone acetylation (Weaver et al, 2008). Both of these modifications are typically associated with active chromatin, or gene expression. This GR acts in a feedback loop to reduce some stress responses. Therefore, these findings are consistent with the rat behaviors with high levels of maternal care associated with higher expression of GR, reducing the stress response of adult rats. This example illustrates how the DNA nucleotide sequence of the offspring does

not change, but their experiences and environments alters their gene expression profile and therefore their phenotype.

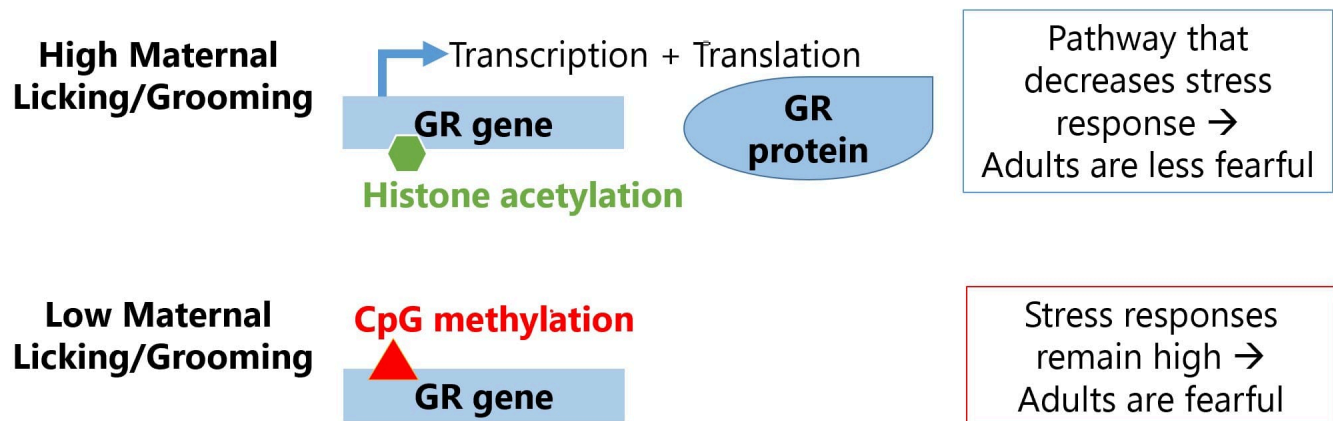


Figure 6.3.1: Summary of relationship between maternal behavior and glucocorticoid (GR) gene epigenetic modifications (based on Weaver et al., 2008)

Imprinting

For some genes, the allele inherited from the oocyte is expressed differently than the allele that is inherited from the sperm. This pattern 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 or 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. For example, a sperm-producing individual inherits epigenetic information from both and oocyte and sperm, but the epigenetic marks will be erased before sperm development and only one the sperm pattern of imprinting will be passed to his offspring, either male or female.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Video 6.3.2: Mammalian Imprinting Cycle. (CC BY Leacock via <https://youtu.be/HfBQ9hU2dsA>)

Because imprinting often produces monoallelic expression of a gene, if one copy is mutated or lost, phenotypic changes may be observed. This phenomenon produces an exception to Mendelian rules about recessive alleles not producing a phenotype when heterozygous.

Imprinted genes, of which there are currently about 100 genes in humans, appear to explain many 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. Adults who were conceived during the Dutch famine of 1944-1945 have *IGF2* genes that are less methylated than their siblings born before or after the famine, even decades after birth (Heijmans et al, 2008). The IGF2 protein is a signaling molecule that promotes growth and cell division. Further research is ongoing to determine how much impact this change in methylation has on phenotypes.

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 (Pembrey et al, 2006). At least one study has recently replicated this finding in an independent population from Uppsala, Sweden (Vågerö et al, 2021)

In mouse models, the *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 (for example) or by mutating a gene that controls methylation, some mice become yellow and overweight. Although the nucleotide sequence of the gene remains unchanged, the methylation at the locus decreases, causing overexpression of the protein (review 4.4: [Exceptions to simple dominance](#)). 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.

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. The requirement to experience a long period of cold temperatures prior to flowering is called vernalization.

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.

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 (Michaels and Amasino, 1999). 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 remove methyl and acetyl groups from the histones associated with the *FLC* locus. The longer the cold temperatures persist, the more acetyl groups are removed, 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.

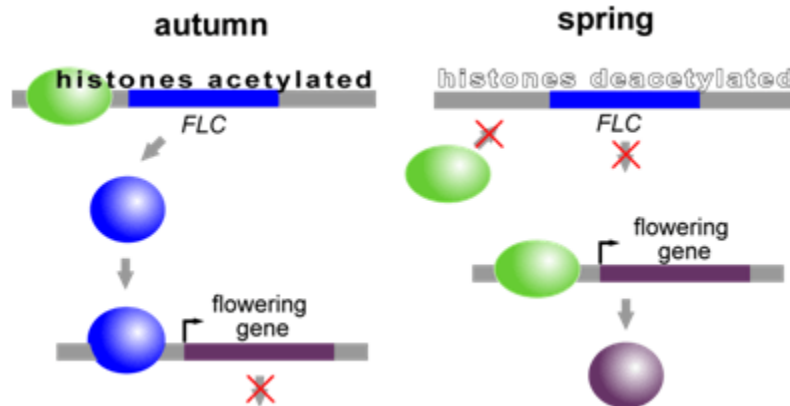


Figure 6.3.2: In the autumn, histones associated with FLC are acetylated, allowing this repressor of flowering genes to be expressed. During winter, enzymes progressively deacetylate FLC, preventing it from being expressed, and therefore allowing flowering genes to respond to other signals that induce flowering. (OriganI-Deyholos-CC:AN)

Query 6.3.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



References

- Bygren LO, Tinghög P, Carstensen J, Edvinsson S, Kaati G, Pembrey ME, Sjöström M. Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. *BMC Genet.* 2014 Feb 20;15:12. doi: 10.1186/1471-2156-15-12. PMID: 24552514; PMCID: PMC3929550.
- He Y, Michaels SD, Amasino RM. Regulation of flowering time by histone acetylation in Arabidopsis. *Science.* 2003 Dec 5;302(5651):1751-4. doi: 10.1126/science.1091109. Epub 2003 Oct 30. PMID: 14593187.
- Heijmans BT, Tobi EW, Stein AD, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 2008;105(44):17046-17049. doi:10.1073/pnas.0806560105 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2579375/>)
- Michaels SD, Amasino RM, *FLOWERING LOCUS C* Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering, *The Plant Cell*, Volume 11, Issue 5, May 1999, Pages 949–956, <https://doi.org/10.1105/tpc.11.5.949>.
- Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjöström M, Golding J; ALSPAC Study Team. Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet.* 2006 Feb;14(2):159-66. doi: 10.1038/sj.ejhg.5201538. PMID: 16391557.
- Vågerö D, Pinger PR, Aronsson V, van den Berg GJ. Paternal grandfather's access to food predicts all-cause and cancer mortality in grandsons. *Nat Commun.* 2018 Dec 11;9(1):5124. doi: 10.1038/s41467-018-07617-9. Erratum in: *Nat Commun.* 2021 Mar 23;12(1):1954. PMID: 30538239; PMCID: PMC6290014.

Weaver, I., Cervoni, N., Champagne, F. *et al.* Epigenetic programming by maternal behavior. *Nat Neurosci* **7**, 847–854 (2004).
<https://doi.org/10.1038/nn1276>

Contributors and Attributions

- [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 [6.3: Epigenetics](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [6.3: Epigenetics](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

6.4: Chapter 6 Review Questions (Draft)

6.1: Prokaryotic gene regulation

Query 6.4.6.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



6.5: Chapter 6 Review Questions Draft

6.1: Prokaryotic gene regulation

Query 6.5.6.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



6.2: Eukaryotic gene regulation

Query 6.5.6.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 6.5.6.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 6.5.6.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 6.5.6.5

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 6.5.6.6

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



6.3: Epigenetics

Query 6.5.6.7

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



7: MANIPULATING AND ANALYZING GENES AND GENOMES

The study of some genetic technologies have been integrated in previous chapters.

PCR and Sequencing were covered in [1.4: In vitro applications of DNA replication](#).

CRISPR gene editing was introduced in [4.1.2: CRISPR](#).

This chapter connects these tools to a few applications that are relevant to current science and society.

[7.1: Analysis of STRs](#)

[7.2: Transgenic Organisms](#)

7.1: Analysis of STRs

Polymorphisms

The difference in nucleotide sequences between humans lies between 0.1-0.4%. This means that people are greater than 99% similar. But when you look at your classmates around the room, you can see that that small difference amounts to quite a bit of variation within our species. The bulk of these differences aren't even within the coding sequences of genes but lie outside in regulatory regions that change the expression of those genes. Imagine if there were mutations to the coding sequences, this could be very deleterious to the well-being of the organism. We say that the coding sequences of genes that ultimately lead to proteins have **selective pressure** to remain the same. The areas outside of the coding sequences have a reduced and sometimes non-existent selection pressure. These areas are allowed to mutate in sequence and even expand or contract. Areas of changes or differences are called **polymorphic** (many forms). If you were to read a repetitive set of sequences and count the repetition, you' would make mistakes and lose count. Likewise, DNA polymerase will make errors or stutter in areas of repetitiveness and produce polymorphic regions.

Tandem Repeats

A type of polymorphism occurs due to these repeats expanding and contracting in non-coding regions. These regions are called variable number tandem repeats (**VNTRs**) or sometimes short tandem repeats (**STRs**). Any region or location on a chromosome is referred to as **locus** (loci for plural). Scientists use polymorphic loci that are known to contain VNTRs/STRs in order to differentiate people based on their DNA. This is often used in forensic science or in maternity/paternity cases. Any variation of a locus is referred to as an **allele**. In standard genetics, we often think of an allele as a variation of a gene that would result in a difference in a physical manifestation of that gene. In the case of STRs, these alleles are simply a difference in the number of repeats. This means the length of DNA within this locus is either longer or shorter and gives rise to many different alleles. VNTRs are referred to as minisatellites while STRs are called **microsatellites**.

CoDIS

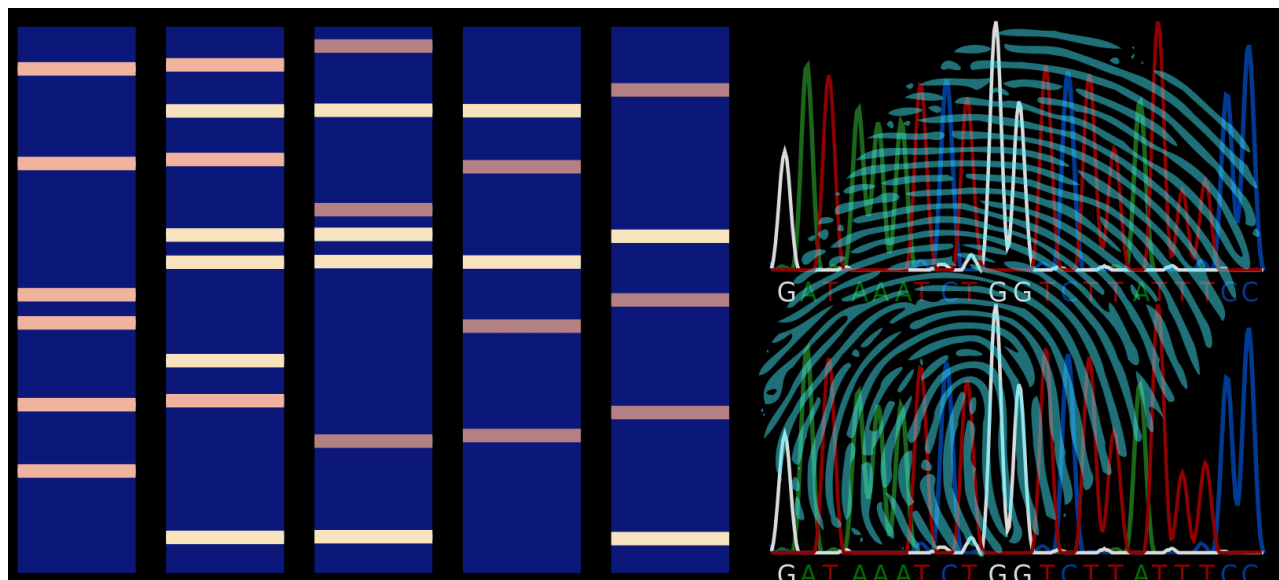


Figure 7.1.1: DNA fingerprinting. Credit: Helixitta, The Photographer and Jeremy Seto (CC-BY-SA 3.0)

The FBI and local law enforcement agencies have developed a database called the Combined DNA Index System (**CoDIS**) that gathers data on a number of STRs. By establishing the number of repeats of a given locus, law enforcement officials can differentiate individuals based on the repeat length of these alleles. CoDIS uses a set of 20 loci that are tested together. As you would imagine, people are bound to have the same alleles of certain loci, especially if they were related. The use of 20 different loci makes it statistically improbable that 2 different people could be confused with each other. Think about this in terms of

physical traits. As you increase the number of physical traits used to describe someone, you are less likely to confuse that person with someone else based on those combinations of traits. Using the CoDIS loci increases the stringency since there are many alleles for each locus. The twentieth locus in CoDIS (called AMEL) discriminates between male and female.

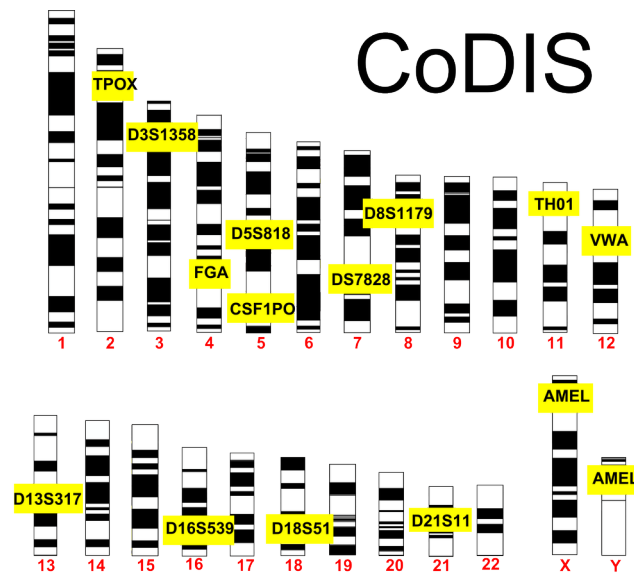


Figure 7.1.2: **CoDIS STRs:** The FBI utilizes 20 different loci to discriminate between people. AMEL discriminates by gender and is located on the X & Y. Credit: Jeremy Seto (CC-BY-NC-SA)

Crime Scene Investigation

This lab uses a CoDIS locus called TH01. TH01 is a locus on chromosome 11 that has a repeating sequence of TCAT. There are reported to be between 3-14 repeats in this locus. With the exception of X and Y in a male, all chromosomes have a homologous partner. Therefore, each individual will have 2 alleles for each CoDIS locus.

The TH01 locus contains repeats of TCAT.

CCC **TCAT** **TCAT** **TCAT** **TCAT** **TCAT** **TCAT** AAA

This example has 6 TCAT repeats.

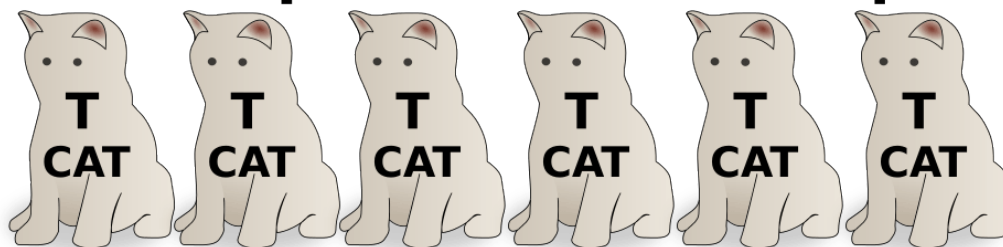


Figure 7.1.2: **TH01 STR:** Outside of the STR, there are flanking areas of known sequence. The primers that amplify TH01 in PCR recognize these flanking sequences to amplify the TCAT repeats. Credit: Jeremy Seto (CC-BY-NC-SA)

At a crime scene, criminals don't often leave massive amounts of tissue behind. Scant evidence in the form of a few cells found within bodily fluids or stray hairs can be enough to use as DNA evidence. DNA is extracted from these few cells and amplified by PCR using the specific primers that flank the STRs used in CoDIS.

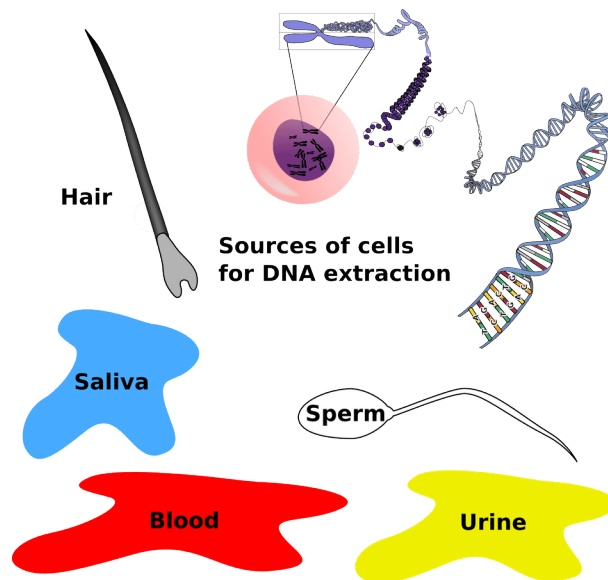


Figure 7.1.4: **DNA evidence from a crime scene:** DNA can be extracted from cells found from various sources at a crime scene. PCR can amplify this small amount of DNA. Credit: Jeremy Seto (CC-BY-NC-SA)

Amplified DNA will be separated by gel electrophoresis and analyzed. Size reference standards and samples from the crime scene and the putative suspects would be analyzed together. In a paternity test, samples from the mother, the child, and the suspected father would be analyzed in the same manner. A simple cheek swab will supply enough cells for this test.

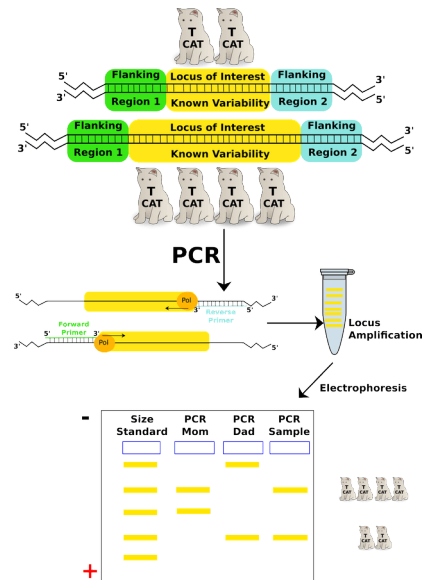


Figure 7.1.5: **TH01 locus used in a Paternity/Maternity test:** Individual PCR reactions are run for each sample (mom, dad, and child). The TH01 primer pair specifically amplifies the locus. Each amplified sample is run on the same gel to resolve the different alleles of TH01 from each individual. From this test, the sample could be the offspring from these 2 parents but the use of more STRs would make it more definitive. Count the TCATs. Credit: Jeremy Seto (CC-BY-NC-SA)

External Resources

- Flash animation walking through what an STR is: <http://www.dnalc.org/view/15981-DNA-variations.html>

This page titled 7.1: Analysis of STRs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Cheneir Neely.

- 8.3: Variable Number Tandem Repeats by Bio-OER is licensed CC BY-NC-SA 4.0.

7.2: Transgenic Organisms

Learning Objectives

- Define transgenic organism.
- Understand the roles of plasmids and vectors in creating transgenic organisms.
- Explain why stem cells are used to create transgenic mice.
- Distinguish between germ line and somatic gene therapy.
- Discuss the uses, benefits, risks, and challenges associated with 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 become 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 7.2.1). *Agrobacterium tumefaciens* 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.

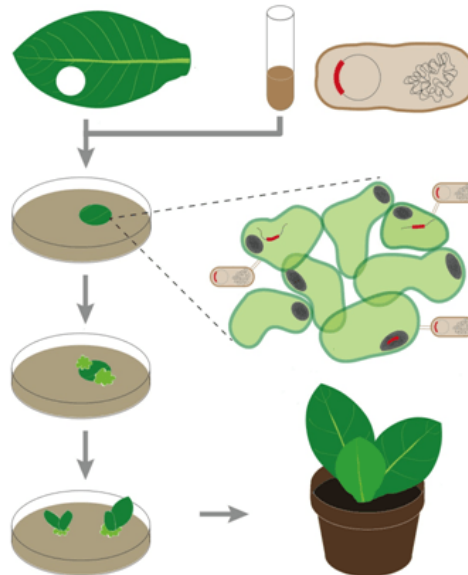


Figure 7.2.1: Production of a transgenic plant using Agrobacterium-mediated transformation. The bacterium has been transformed with a T-DNA plasmid that contains the transgene and a selectable marker that confers resistance to a herbicide or antibiotic. A bacterial culture and plant tissue (e.g. a leaf 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 7.2.2). 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 7.2.2: 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 7.2.3). 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, *neo^r*, 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.

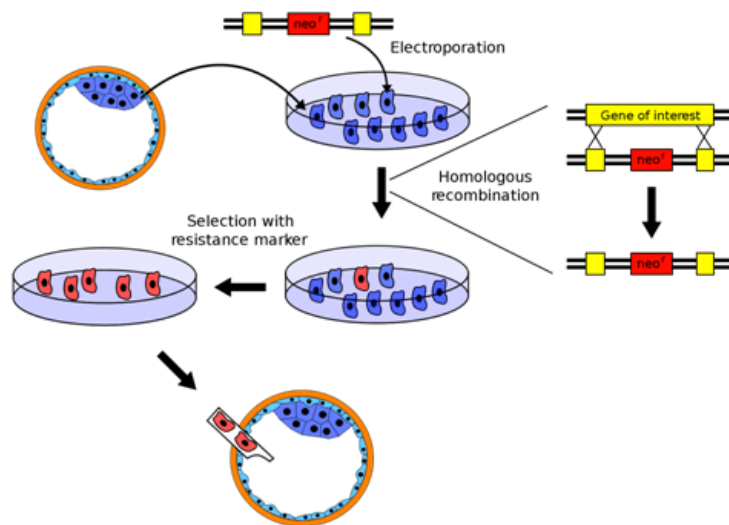


Figure 7.2.3: Production of a transgenic mouse. Stem cells are removed from an embryo, and are transfected (using electroporation) with a transgenic construct that bears a neomycin resistance gene (*neo^r*) flanked by two segments of DNA homologous to a gene of interest. In the nucleus of a transgenic 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 [7.2: Transgenic Organisms](#) is shared under a license and was authored, remixed, and/or curated by Cheneir Neely.

- [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.

8: 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.

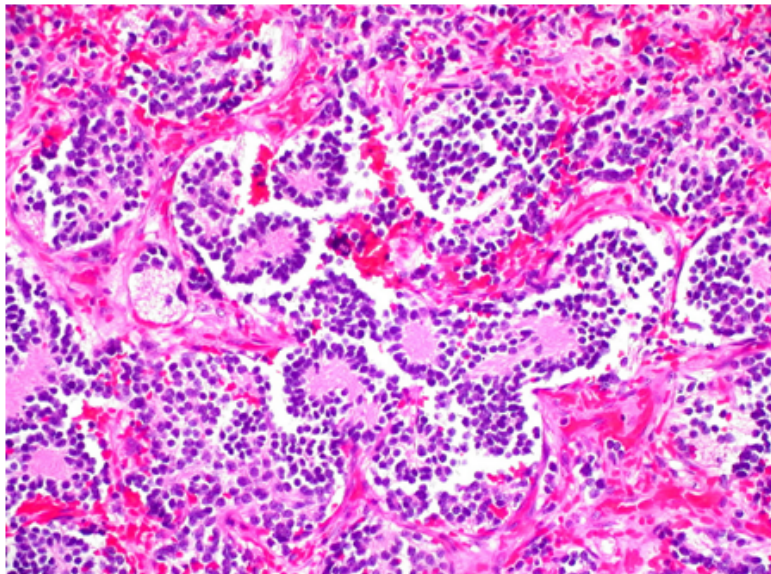


Figure 8.1: Stained histological section of a neuroblastoma in an adrenal gland. (Flickr-Ed Uthman-CC:AS)

[8.1: Classification of Cancers](#)

[8.2: Hallmarks of Cancer](#)

[8.3: Mutagens and Carcinogens](#)

[8.4: Oncogenes](#)

[8.5: Tumor Suppressor Genes](#)

[8.6: Genetics and Targeted Cancer Treatment – The Story of Gleevec™ \(Imatinib\)](#)

[8.7: Cancer Genetics \(Exercises\)](#)

[8.8: Chapter 8 Review Questions](#)

Contributors and Attributions

- [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: Cancer Genetics](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [8: Cancer Genetics](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

8.1: 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 [8.1: Classification of Cancers](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [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.

8.2: Hallmarks of Cancer

Researchers have identified molecular and cellular traits that characterize most cancers. An original six hallmarks of cancer were identified and later four additional hallmarks were added (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011 [https://www.cell.com/fulltext/S0092-8674\(11\)00127-9](https://www.cell.com/fulltext/S0092-8674(11)00127-9)).

Hallmarks of Cancer

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.

Cancer hallmarks reflect genetic changes

Previous chapters have covered many topics that are related to these hallmarks.

For each hallmark, try to think of a gene or process that could be mutated to produce the effect.

Contributors and Attributions

- 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 [8.2: Hallmarks of Cancer](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [8.2: Hallmarks of Cancer](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

8.3: 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 8.3.4) 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 8.3.5). 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.

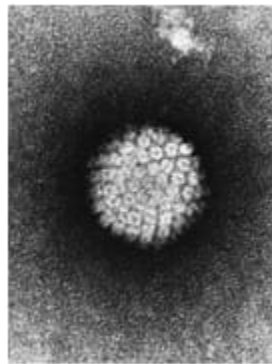


Figure 8.3.4: Electron micrograph of HPV. (Wikipedia-Unknown-PD)

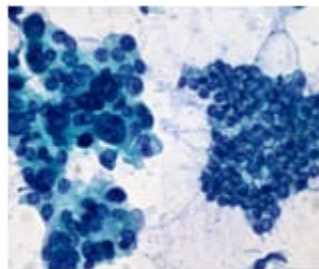


Figure 8.3.5: Dysplastic (left) and normal (right) cells from a Pap smear. (Flickr-Ed Uthman-CC:AS)

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 8.3.2) 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 8.3.2: *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

Testing for carcinogens

The Ames Test ([4.1.3 Genetic screens](#)) can be used as one screen for carcinogenic (mutagenic) potential. Because all living things share the same genetic material (DNA), a chemical that can damage DNA in one cell type may also damage DNA in another cell type.

Such tests do not perfectly predict carcinogenic activity; for example, a chemical that is not carcinogenic in one organism might be metabolized to a carcinogenic form in another organism.

Contributors and Attributions

- [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 [8.3: Mutagens and Carcinogens](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [8.3: Mutagens and Carcinogens](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

8.4: 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.

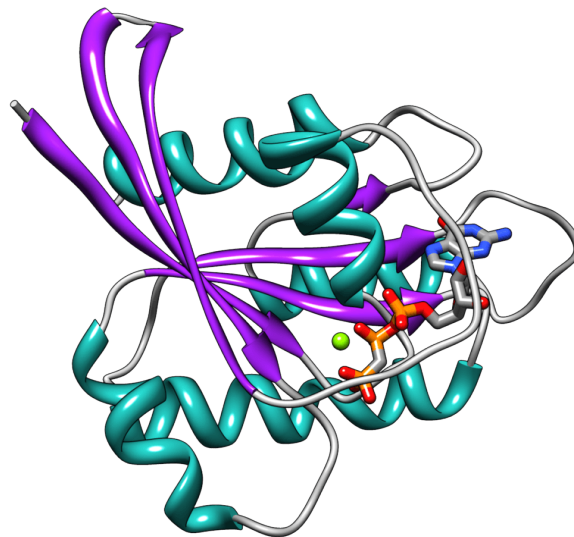


Figure 8.4.1: HRas structure PDB 121p, ribbon showing strands in purple, helices in aqua, loops in gray. Also shown are the bound GTP analog and magnesium ion. (CC-BY-SA ElaineMeng)

The *ras* oncogene was originally identified from cancer-causing viruses. *ras* is an example of a proto-oncogene; this protein acts as a transducer in 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 8.4.2.

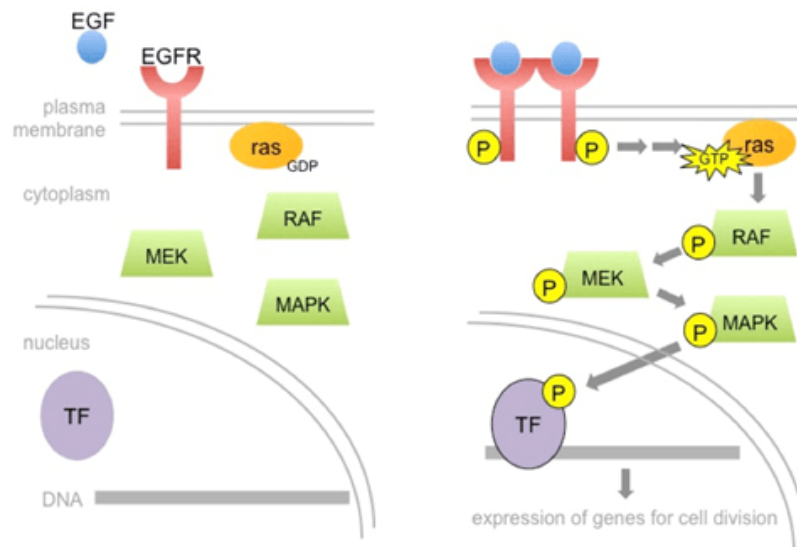


Figure 8.4.2: Simplified representation of the epidermal growth factor (EGF) signaling pathway. In the panel on the left, the components are shown in their inactive forms, prior to stimulation of the pathway. The components include the soluble ligand, EGF, its receptor (EGFR, a tyrosine kinase), ras (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 stimulates 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)

Ras mutations in cancer

There are three *ras* homologs in the human genome (HRas, NRas, and KRas). As many cancer genomes are sequenced, databases such as COSMIC (<http://www.sanger.ac.uk/cosmic>) can be used to search mutations across cancer types.

Visit the site and search for HRAS, NRAS, or KRAS. Can you find answers to these questions?

- What amino acid positions are frequently mutated?
- In what cancer types is the gene most often mutated?
- Are the mutations most often missense, nonsense, frameshift, or silent? Is this what you would have predicted -- why or why not?

Exercise 8.4.1

All human cells have proto-oncogenes. True or False?

Answer available at the end of this section.

Exercise 8.4.2

Are the mutations that convert proto-oncogenes to oncogenes like to be loss-of-function or gain-of-function mutations? Explain your reasoning.

Answer available at the end of this section.

Contributors and Attributions

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

licensing guidelines.

Answers

Exercise 8.4.1

True.

Exercise 8.4.2

Oncogenic mutations are typically gain-of-function, resulting in increased protein activity and therefore increased cell proliferation. A loss-of-function mutation in a proto-oncogene, would result in reduced cell proliferation, not cancer.

This page titled [8.4: Oncogenes](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [8.4: Oncogenes](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

8.5: 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).

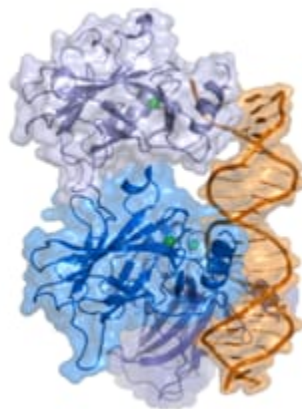


Figure 8.5.1: p53 bound to its target site on a DNA molecule. (Wikipedia-Thomas Spettstoesser from Cho et al, Science 265PP346, 1994-CC:AS)

Inherited Cancer Predisposition

While cancer itself is not inherited, mutations that can drive cancer can be passed from parent to offspring. Many of the commonly inherited cancer-associated mutations are in tumor suppressor genes. In most cases, individuals inherit only one mutant allele of the tumor suppressor; they are heterozygous for the recessive mutation. However, if the second copy (due to random chance) becomes mutated or deleted in a somatic cell, that cell (and all its descendants) now lack the tumor suppressor entirely. This process is known as loss-of-heterozygosity.

Contributors and Attributions

- [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 [8.5: Tumor Suppressor Genes](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [8.5: Tumor Suppressor Genes](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

8.6: Genetics and Targeted Cancer Treatment – The Story of 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), 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-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 8.6.9). 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.

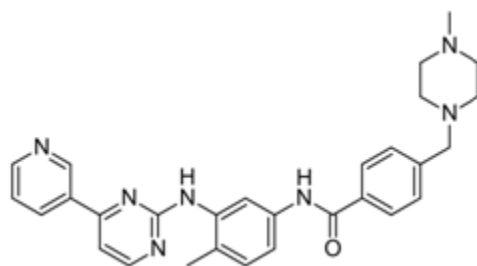


Figure 8.6.1: Biochemical structure of Gleevec™ or Imatinib. . (Wikipedia-Fuse809-CC:AN)

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 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.

Contributors and Attributions

- [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 [8.6: Genetics and Targeted Cancer Treatment – The Story of Gleevec™ \(Imatinib\)](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Stefanie West Leacock](#).

- [8.6: Genetics and Targeted Cancer Treatment – The Story of Gleevec™ \(Imatinib\)](#) by [Stefanie West Leacock](#) is licensed [CC BY-SA 4.0](#).

8.7: 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?

8.8: Chapter 8 Review Questions

Query 8.8.8.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 8.8.8.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 8.8.8.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



Query 8.8.8.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



9: HOMEWORK

If a Homework assignment is not on a single page here, then it can be accessed via ADAPT: <https://adapt.libretexts.org/>

9.1: 1. Instructions

9.2: Homework 2

9.1: 1. Instructions

Creating an ADAPT Libretexts Account

You will complete homework by logging into your LibreTexts account. **You must use your school (@ualr.edu) email** to register your LibreTexts account.

1. Make sure you have our access code.
2. Go to <https://adapt.libretexts.org>
3. Click Register in the top right corner
4. Select "Student" from the drop-down menu
5. Click on "Campus Registration". **(BOX AT THE TOP) -- NOT THE BOXES UNDER REGISTER WITH ADAPT.**
6. In the overlay window, select "Google Workspace"
7. Enter your **school email (@ualr.edu)** and click "Next" or select your **school email (@ualr.edu)** from the list if you are already logged into Google
8. If you are prompted to log in with your UALR Net ID and Password, do so
9. Select your Time Zone (Central Time) and set registration type to "Student"
10. Type in your T-number in the Student ID field and Click "Submit"
11. After your account is created you will be brought to My Courses page
12. Click "Enroll in Course"
13. Paste the Access Code
14. Click "Submit"
15. Next, Click on your name in the top right corner and go to Settings.
16. Go to Notifications and choose the time option for your reminders. Click "Update"

Creating Accounts. (2021, August 23). Retrieved August 23, 2021, from <https://chem.libretexts.org/@go/page/347140> " Creating Accounts" by LibreTexts is licensed under notset .

9.2: Homework 2

ADAPT 9.2.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



ADAPT 9.2.2

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



ADAPT 9.2.3

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



ADAPT 9.2.4

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



ADAPT 9.2.5

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



ADAPT 9.2.6

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



INDEX

A

- allele
 - [4.2: Mendelian Genetics](#)
- aneuploid
 - [5.1: Changes in Chromosome Number](#)
- annotation:hypothesis
 - [1.6: Chapter 1 Review Questions](#)

C

- cAMP binding protein (CAP)
 - [6.1: Prokaryotic gene regulation](#)
- Chargaff's Rules
 - [1.1: The Structure of DNA](#)
- chimera
 - [4.4: Exceptions to autosomal inheritance](#)
- chromatin
 - [1.2: Chromosomes and chromatin](#)
- chromosome number
 - [5.1: Changes in Chromosome Number](#)
- cloning
 - [4.2.1: Monohybrid Crosses and Segregation](#)
- constitutive mutants
 - [6.1.1: The Use of Mutants to Study the lac Operon](#)
- CRISPR
 - [4.1.2: CRISPR](#)

D

- dihybrids
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)

E

- epigenetics
 - [6.3: Epigenetics](#)
- episome
 - [6.1.1: The Use of Mutants to Study the lac Operon](#)
- euchromatin
 - [1.2: Chromosomes and chromatin](#)
- euploid
 - [5.1: Changes in Chromosome Number](#)
- Expressivity
 - [4.6: Exceptions to simple dominance](#)

G

- glnol:yes
 - [2.3: Genetic Code and Translation](#)
- gsmol:yes
 - [1.1: The Structure of DNA](#)
- gynandromorph
 - [4.4: Exceptions to autosomal inheritance](#)

H

- Hayflick limit
 - [1.2: Chromosomes and chromatin](#)
- hemizygous
 - [4.2: Mendelian Genetics](#)

heterochromatin

- [1.2: Chromosomes and chromatin](#)

heterozygous

- [4.2: Mendelian Genetics](#)

histone

- [1.2: Chromosomes and chromatin](#)

holocentric

- [1.2: Chromosomes and chromatin](#)

L

- lac operon
 - [6.1: Prokaryotic gene regulation](#)
- lacI
 - [6.1: Prokaryotic gene regulation](#)
- Law of Equal Segregation
 - [4.2: Mendelian Genetics](#)
- Law of Independent Assortment
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)
- linkage
 - [4.5: Linkage](#)
- linked genes
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)

M

- meiocytes
 - [4.1: Meiosis](#)
- meiotic crossovers
 - [5.2: Changes in Chromosome Structure](#)
- Mendel's First Law
 - [4.2: Mendelian Genetics](#)
- Mendel's Second Law
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)
- monohybrid crosses
 - [4.2.1: Monohybrid Crosses and Segregation](#)
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)
- mosaic
 - [4.4: Exceptions to autosomal inheritance](#)

N

- nondisjunction
 - [5.1: Changes in Chromosome Number](#)
- nucleosome
 - [1.2: Chromosomes and chromatin](#)

O

- OGG
 - [1.1: The Structure of DNA](#)
 - [1.2: Chromosomes and chromatin](#)
 - [4.1: Meiosis](#)
 - [4.2: Mendelian Genetics](#)
 - [4.2.1: Monohybrid Crosses and Segregation](#)
 - [4.4: Exceptions to autosomal inheritance](#)
 - [4.4.2: Organellar Inheritance](#)
 - [4.6: Exceptions to simple dominance](#)
- oncogene
 - [8.4: Oncogenes](#)

OOG

- [1.1: The Structure of DNA](#)
- [1.2: Chromosomes and chromatin](#)
- [4.1: Meiosis](#)
- [4.2: Mendelian Genetics](#)
- [4.2.1: Monohybrid Crosses and Segregation](#)
- [4.4: Exceptions to autosomal inheritance](#)
- [4.4.2: Organellar Inheritance](#)
- [4.6: Exceptions to simple dominance](#)

origin of replication

- [1.2: Chromosomes and chromatin](#)

P

- penetrance
 - [4.6: Exceptions to simple dominance](#)
- polymorphisms
 - [7.1: Analysis of STRs](#)
- Punnett square
 - [4.2.1: Monohybrid Crosses and Segregation](#)

R

- reciprocal crosses
 - [4.4: Exceptions to autosomal inheritance](#)
- RNA processing
 - [2.2: RNA processing](#)

S

- scaffold proteins
 - [1.2: Chromosomes and chromatin](#)

T

- transgenic organisms
 - [7.2: Transgenic Organisms](#)

V

- vernalization
 - [6.3: Epigenetics](#)

INDEX

A

- allele
 - [4.2: Mendelian Genetics](#)
- aneuploid
 - [5.1: Changes in Chromosome Number](#)
- annotation:hypothesis
 - [1.6: Chapter 1 Review Questions](#)

C

- cAMP binding protein (CAP)
 - [6.1: Prokaryotic gene regulation](#)
- Chargaff's Rules
 - [1.1: The Structure of DNA](#)
- chimera
 - [4.4: Exceptions to autosomal inheritance](#)
- chromatin
 - [1.2: Chromosomes and chromatin](#)
- chromosome number
 - [5.1: Changes in Chromosome Number](#)
- cloning
 - [4.2.1: Monohybrid Crosses and Segregation](#)
- constitutive mutants
 - [6.1.1: The Use of Mutants to Study the lac Operon](#)
- CRISPR
 - [4.1.2: CRISPR](#)

D

- dihybrids
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)

E

- epigenetics
 - [6.3: Epigenetics](#)
- episome
 - [6.1.1: The Use of Mutants to Study the lac Operon](#)
- euchromatin
 - [1.2: Chromosomes and chromatin](#)
- euploid
 - [5.1: Changes in Chromosome Number](#)
- Expressivity
 - [4.6: Exceptions to simple dominance](#)

G

- glmol:yes
 - [2.3: Genetic Code and Translation](#)
- gsmol:yes
 - [1.1: The Structure of DNA](#)
- gynandromorph
 - [4.4: Exceptions to autosomal inheritance](#)

H

- Hayflick limit
 - [1.2: Chromosomes and chromatin](#)
- hemizygous
 - [4.2: Mendelian Genetics](#)

heterochromatin

- [1.2: Chromosomes and chromatin](#)

heterozygous

- [4.2: Mendelian Genetics](#)

histone

- [1.2: Chromosomes and chromatin](#)

holocentric

- [1.2: Chromosomes and chromatin](#)

L

- lac operon
 - [6.1: Prokaryotic gene regulation](#)
- lacI
 - [6.1: Prokaryotic gene regulation](#)
- Law of Equal Segregation
 - [4.2: Mendelian Genetics](#)
- Law of Independent Assortment
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)
- linkage
 - [4.5: Linkage](#)
- linked genes
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)

M

- meiocytes
 - [4.1: Meiosis](#)
- meiotic crossovers
 - [5.2: Changes in Chromosome Structure](#)
- Mendel's First Law
 - [4.2: Mendelian Genetics](#)
- Mendel's Second Law
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)
- monohybrid crosses
 - [4.2.1: Monohybrid Crosses and Segregation](#)
 - [4.2.2: Dihybrid Crosses and Independent Assortment](#)
- mosaic
 - [4.4: Exceptions to autosomal inheritance](#)

N

- nondisjunction
 - [5.1: Changes in Chromosome Number](#)
- nucleosome
 - [1.2: Chromosomes and chromatin](#)

O

- OGG
 - [1.1: The Structure of DNA](#)
 - [1.2: Chromosomes and chromatin](#)
 - [4.1: Meiosis](#)
 - [4.2: Mendelian Genetics](#)
 - [4.2.1: Monohybrid Crosses and Segregation](#)
 - [4.4: Exceptions to autosomal inheritance](#)
 - [4.4.2: Organellar Inheritance](#)
 - [4.6: Exceptions to simple dominance](#)
- oncogene
 - [8.4: Oncogenes](#)

OOG

- [1.1: The Structure of DNA](#)
- [1.2: Chromosomes and chromatin](#)
- [4.1: Meiosis](#)
- [4.2: Mendelian Genetics](#)
- [4.2.1: Monohybrid Crosses and Segregation](#)
- [4.4: Exceptions to autosomal inheritance](#)
- [4.4.2: Organellar Inheritance](#)
- [4.6: Exceptions to simple dominance](#)

origin of replication

- [1.2: Chromosomes and chromatin](#)

P

- penetrance
 - [4.6: Exceptions to simple dominance](#)
- polymorphisms
 - [7.1: Analysis of STRs](#)
- Punnett square
 - [4.2.1: Monohybrid Crosses and Segregation](#)

R

- reciprocal crosses
 - [4.4: Exceptions to autosomal inheritance](#)
- RNA processing
 - [2.2: RNA processing](#)

S

- scaffold proteins
 - [1.2: Chromosomes and chromatin](#)

T

- transgenic organisms
 - [7.2: Transgenic Organisms](#)

V

- vernalization
 - [6.3: Epigenetics](#)

GLOSSARY

S

Sample Word 1 | Sample Definition 1

DETAILED LICENSING

Overview

Title: BIO 300: Introduction to Genetics (Neely)

Webpages: 79

Applicable Restrictions: Noncommercial

All licenses found:

- Undeclared: 59.5% (47 pages)
- CC BY-SA 4.0: 38% (30 pages)
- CC BY 4.0: 1.3% (1 page)
- CC BY-NC-SA 4.0: 1.3% (1 page)

By Page

- BIO 300: Introduction to Genetics (Neely) - *Undeclared*
 - Front Matter - *Undeclared*
 - TitlePage - *Undeclared*
 - InfoPage - *Undeclared*
 - Table of Contents - *Undeclared*
 - Licensing - *Undeclared*
 - 1: Chemistry to Chromosomes - *Undeclared*
 - 1.1: The Structure of DNA - CC BY-SA 4.0
 - 1.2: Chromosomes and chromatin - CC BY-SA 4.0
 - 1.3: Mitosis - *Undeclared*
 - 1.3.1: DNA replication - *Undeclared*
 - 1.4: In vitro applications of DNA replication - *Undeclared*
 - 1.5: Genomes and variation - *Undeclared*
 - 1.6: Chapter 1 Review Questions - *Undeclared*
 - 2: Central Dogma - *Undeclared*
 - 2.1: Overview of Transcription - CC BY 4.0
 - 2.2: RNA processing - *Undeclared*
 - 2.3: Genetic Code and Translation - *Undeclared*
 - 2.4: Chapter 2 Review Questions - *Undeclared*
 - 3: Mutation - *Undeclared*
 - 3.1: Origins of Mutations - CC BY-SA 4.0
 - 3.2: Consequences of mutations - *Undeclared*
 - 3.3: Example - Mutations and Cystic Fibrosis - CC BY-SA 4.0
 - 3.4: Ames Test - *Undeclared*
 - 3.5: DNA Repair - *Undeclared*
 - 3.6: Chapter 3 Review Questions - *Undeclared*
 - 4: Inheritance - *Undeclared*
 - 4.1: Meiosis - CC BY-SA 4.0
 - 4.1.1: Homologous recombination - *Undeclared*
 - 4.1.2: CRISPR - *Undeclared*
 - 4.2: Mendelian Genetics - CC BY-SA 4.0
 - 4.2.1: Monohybrid Crosses and Segregation - CC BY-SA 4.0
 - 4.2.2: Dihybrid Crosses and Independent Assortment - CC BY-SA 4.0
 - 4.3: Pedigrees - *Undeclared*
 - 4.4: Exceptions to autosomal inheritance - CC BY-SA 4.0
 - 4.4.1: Inheritance patterns for X-linked and Y-linked genes - *Undeclared*
 - 4.4.2: Organellar Inheritance - CC BY-SA 4.0
 - 4.5: Linkage - CC BY-SA 4.0
 - 4.5.1: Linkage and Mapping - *Undeclared*
 - 4.5.2: GWAS - *Undeclared*
 - 4.6: Exceptions to simple dominance - CC BY-SA 4.0
 - 4.7: Gene Interactions - CC BY-SA 4.0
 - 4.7.1: Complementation Test - *Undeclared*
 - 4.8: Chapter 4 Review Questions (draft) - *Undeclared*
 - 5: Chromosome variation - *Undeclared*
 - 5.1: Changes in Chromosome Number - CC BY-SA 4.0
 - 5.2: Changes in Chromosome Structure - CC BY-SA 4.0
 - 5.3: Chapter 5 Review Questions (Draft) - *Undeclared*
 - 6: Regulation of Gene Expression - CC BY-SA 4.0
 - 6.1: Prokaryotic gene regulation - CC BY-SA 4.0
 - 6.1.1: The Use of Mutants to Study the lac Operon - CC BY-SA 4.0
 - 6.1.2: The Lux Operon - *Undeclared*
 - 6.2: Eukaryotic Gene Regulation - CC BY-SA 4.0
 - 6.2.1: Gene Expression in Evolution - CC BY-SA 4.0
 - 6.2.2: Regulating Transcription at the Chromatin Level - CC BY-SA 4.0

- 6.2.3: microRNAs - *Undeclared*
- 6.3: Epigenetics - *CC BY-SA 4.0*
- 6.4: Chapter 6 Review Questions (Draft) - *Undeclared*
- 6.5: Chapter 6 Review Questions Draft - *Undeclared*
- 7: Manipulating and Analyzing Genes and Genomes - *Undeclared*
 - 7.1: Analysis of STRs - *CC BY-NC-SA 4.0*
 - 7.2: Transgenic Organisms - *Undeclared*
- 8: Cancer Genetics - *CC BY-SA 4.0*
 - 8.1: Classification of Cancers - *CC BY-SA 4.0*
 - 8.2: Hallmarks of Cancer - *CC BY-SA 4.0*
 - 8.3: Mutagens and Carcinogens - *CC BY-SA 4.0*
 - 8.4: Oncogenes - *CC BY-SA 4.0*
 - 8.5: Tumor Suppressor Genes - *CC BY-SA 4.0*
 - 8.6: Genetics and Targeted Cancer Treatment – The Story of Gleevec™ (Imatinib) - *CC BY-SA 4.0*
 - 8.7: Cancer Genetics (Exercises) - *CC BY-SA 4.0*
 - 8.8: Chapter 8 Review Questions - *Undeclared*
- 9: Homework - *Undeclared*
 - 9.1: 1. Instructions - *Undeclared*
 - 9.2: Homework 2 - *Undeclared*
- Back Matter - *Undeclared*
 - Index - *Undeclared*
 - Index - *Undeclared*
 - Glossary - *Undeclared*
 - Detailed Licensing - *Undeclared*