Genetics, Agriculture, and Biotechnology

WALTER SUZA AND DONALD LEE

IOWA STATE UNIVERSITY DIGITAL PRESS
AMES, IOWA

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

Contents

About the Authors

Lead Contributors

Walter Suza: Suza is an Adjunct Associate Professor at Iowa State University. He teaches courses on Genetics and Crop Physiology in the Department of Agronomy. In addition to co-developing courses for the ISU Distance MS in Plant Breeding Program, Suza also served as the director of Plant Breeding e-Learning in Africa Program (PBEA) for 8 years. With PBEA, Suza helped provide access to open educational resources on topics related to the genetic improvement of crops. His research is on the metabolism and physiology of plant sterols.

Donald Lee: Lee is a Professor of Plant Breeding and Genetics at the University of Nebraska-Lincoln. For 30 years, he has taught Introductory Genetics, Crop and Weed Genetics, and Crop Genetic Engineering. His research is on detection and assessment of molecular genetic variation in crops, weeds and native plants, and development of Internet resources for teaching genetics to a wide variety of learners.

Contributors

Marjorie Hanneman: Hanneman received her Bachelor's of Science in Agronomy and Genetics from Iowa State University in 2021. She is currently a PhD student in Plant Breeding and Genetics at Cornell University studying biofortification of maize. Marjorie created images and illustrations and assisted with the editing of lesson content.

Philip Becraft: Iowa State University

Patricia Hain: University of Nebraska-Lincoln

Deana Namuth-Covert: University of Nebraska-Lincoln

Abbey Elder: Elder is the Open Access & Scholarly Communication Librarian at Iowa State University. She supports the identification, production, and publication of open educational resources across the university, including open textbooks like *Genetics, Agriculture, and Biotechnology*. Abbey created select images and tables for this text, and reviewed and edited the text for accessibility.

1. Mitosis and Meiosis

DONALD LEE; WALTER SUZA; AND MARJORIE HANNEMAN

Learning Objectives

1. Describe the cell cycle.
2. Explain the events of all stages of mitosis.
3. Track chromosome and chromatid number through all stages of mitosis.
4. Explain the events of all stages of meiosis.
5. Track chromosome and chromatid number through all stages of meiosis.
6. Describe the role of meiosis in gamete formation and how it relates to inheritance.

Introduction

Multicellular organisms such as plants and animals are composed of millions to trillions (1,000,000,000) of cells that work together. The cells that make up different tissues have different shapes and perform different functions for the plant or animal. Even though they have diverse functions, each **somatic cell** in the organism normally has the same chromosomes and therefore the same genetic makeup. Furthermore, the millions of cells that makeup a mature organism originated from a single cell formed when the male and female gametes from the parents of the organism fused. This single cell established the life of the organism. Understanding multicellular organisms requires an understanding of the lifecycle of the cells that make up the organism.

The Cell Cycle

Let us think about the cell cycle from a personal point of view. Your age plus about nine months ago you were a **zygote**, a single cell formed when the sperm and egg from your biological parents fused in a fallopian tube (or possibly in a test tube if in vitro fertilization factored into your birth). You have come a long way since then, progressing one cell cycle at a time. The cell cycle is the life cycle of a single cell. We depict the cell cycle in a circular diagram although the cells in your body do not actually go around in circles. The main idea is that when new cells are made from existing cells, the new cells start their lifecycle, and the old cells end theirs.

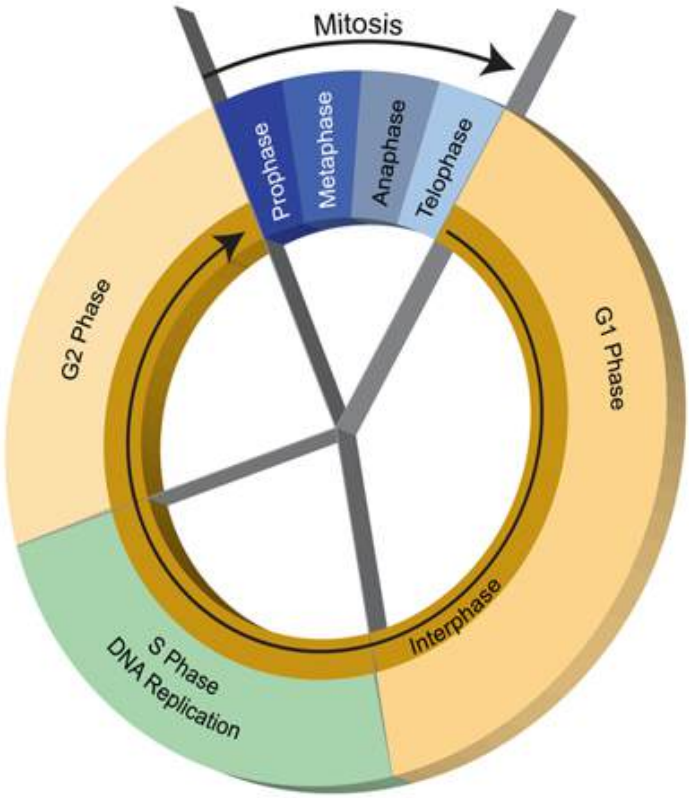


***Figure 1**. The cell cycle depicts the stages in the life of a cell. Image by NIH-NHGRI.*

The first part of the cell cycle is the G1 phase. Cells can go through growth and development in this phase. Some cells differentiate into specialized cells and then never leave this phase. However, a **zygote** does not grow in size, but instead continues the cell cycle so it can quickly give rise to more cells.

The second part of the cell cycle is the S phase (synthesis phase). Here, the cell replicates its chromosomes so that it has a copy of each chromosome to pass on to daughter cells. The third part of the cell cycle is the G2 phase where the cell prepares for division. The G1, S and G2 phases together are called interphase. The M phase completes the cell cycle. 'M' could be mitosis or meiosis depending on the type of cell. For the zygote, the goal is to make more somatic cells. Therefore, it goes through mitosis and gives rise to two daughter cells. This completes the life cycle of the zygote and starts the lifecycle of the new cells. Rounds of the cell cycle continued over and over to form the body you have today. As long as you live, some of your cells must be able to complete the cell cycle.

From a cytogenetics point of view, you have two types of cells in your body. You have cells with 46 **chromosomes** (**somatic** or body cells) and cells with 23 chromosomes (**gamete** or sex cells). Since you started as a single cell with 46 chromosomes, there must be two types of cell division taking place in your body to accommodate both somatic and gamete cells. Mitosis and meiosis are the two types of cell division.

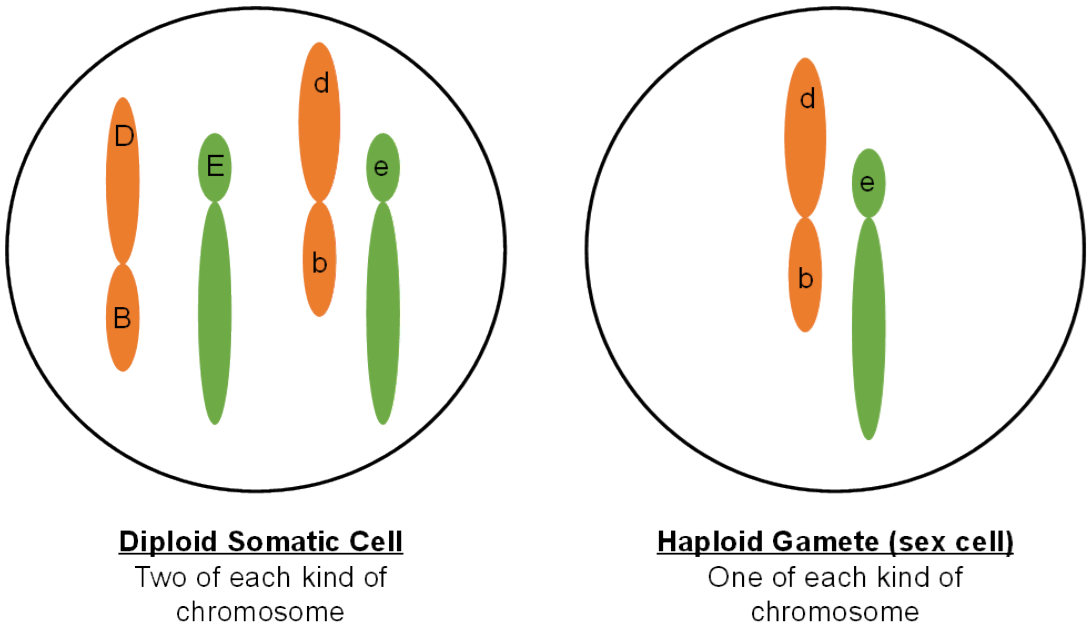


Figure 2. Multicellular organisms that sexually reproduce have diploid and haploid cells. Image by Marjorie Hanneman.

Mitosis: Somatic cell division

The objective of mitosis is to make two genetically identical cells from a single cell. In the cells of our body, we start with 46 chromosomes in a single cell and end up with 46 chromosomes in two cells. Obviously, replicating the chromosomes is a prerequisite to mitosis. Remember, replication takes place during interphase when the chromosomes are dispersed structures in the nucleus. Mitosis is an organized procession of activity in the cell that allows the replicated chromosomes to be properly divided into two identical cells. Chromosomes are important because they contain genes. Therefore, we will include **genes** on our chromosome diagrams and slide show. These pictures depict the four stages of mitosis. We will describe the events at each stage that are important in understanding the distribution of genes during cell division.

Mitosis: Prophase

Prophase is the beginning of mitosis (Figure 3).

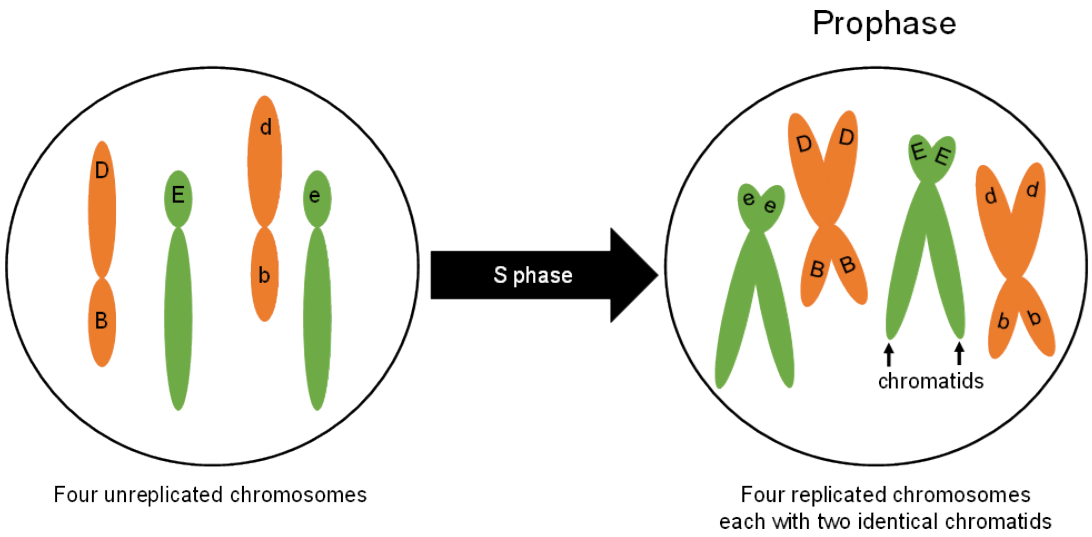


Figure 3. Prophase of mitosis. Image by Marjorie Hanneman.

During interphase, the chromosomes look like a plate of spaghetti in the nucleus. It is difficult to pick out an individual chromosome because they are each so spread out. The chromosomes in the nucleus change from being loosely dispersed to becoming more condensed. This change in chromosome structure makes them easier to move around the cell, an important issue for what is about to happen. As the chromosomes condense, they get shorter and thicker and can be seen through the microscope as individual structures (Figure 4). The chromosomes at prophase will consist of two identical parts called sister chromatids that stay connected at the centromere. It is now clear that the chromosomes have been replicated. Chromosome replication occurs during the S-phase of interphase (Figure 1). Next, the nucleus is dissolved. At the end of prophase the **replicated** chromosomes are moved by the spindle apparatus to the center of the cell.

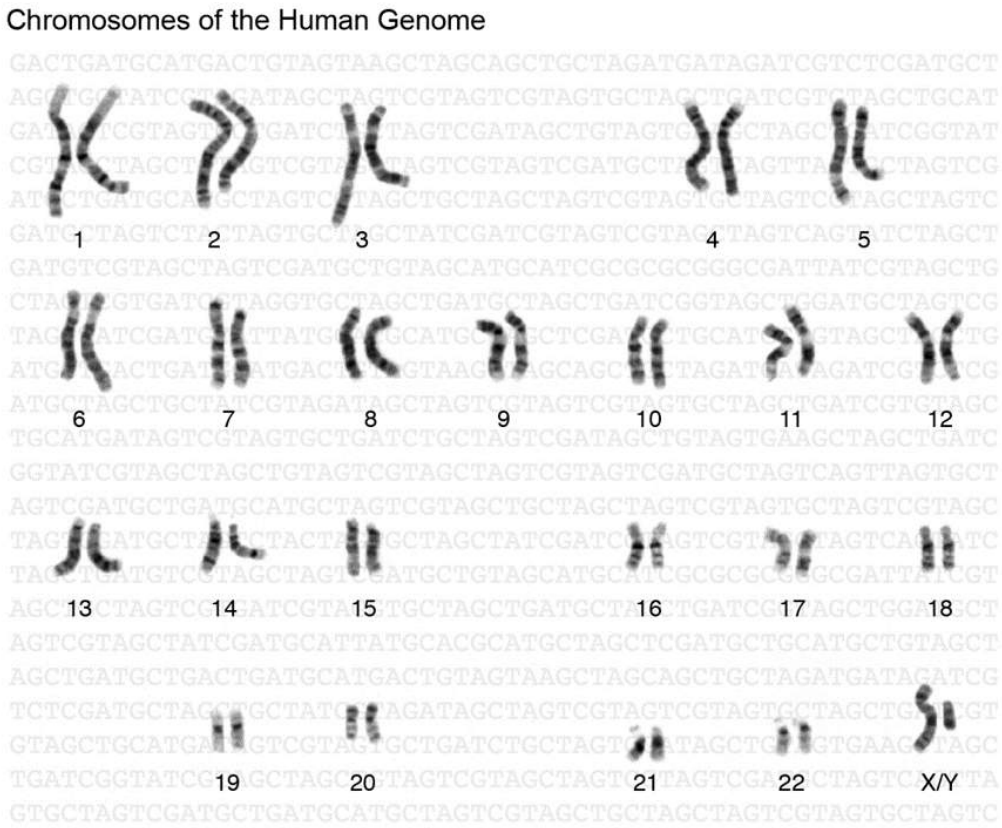


Figure 4. Chromosomes of the human genome. Image by NIH-NHGRI.

Mitosis: Metaphase

When the chromosomes are maneuvered to the center of the cell, metaphase begins (Figure 5). The spindle fiber network connects the centromere of the replicated chromosome to the outer part of each cell. The chromosomes now resemble the line of scrimmage of a football team, poised and waiting for some cellular signal to begin the next phase.

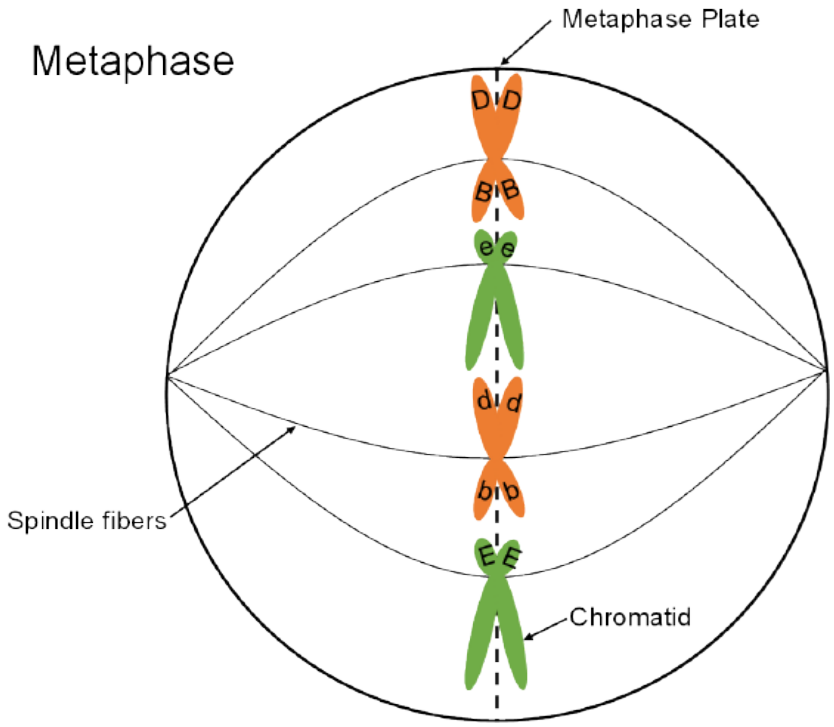


Figure 5. Metaphase of Mitosis. Image by Marjorie Hanneman.

Mitosis: Anaphase

At anaphase (Figure 6) the chromatids making up each chromosome are pulled apart and begin to move away from each other under the control of the spindle fibers. Once they are pulled apart, the chromatids are now considered individual chromosomes. As anaphase progresses, the chromosomes on each end of the cell are pulled into a bundle. When they stop moving, telophase begins.

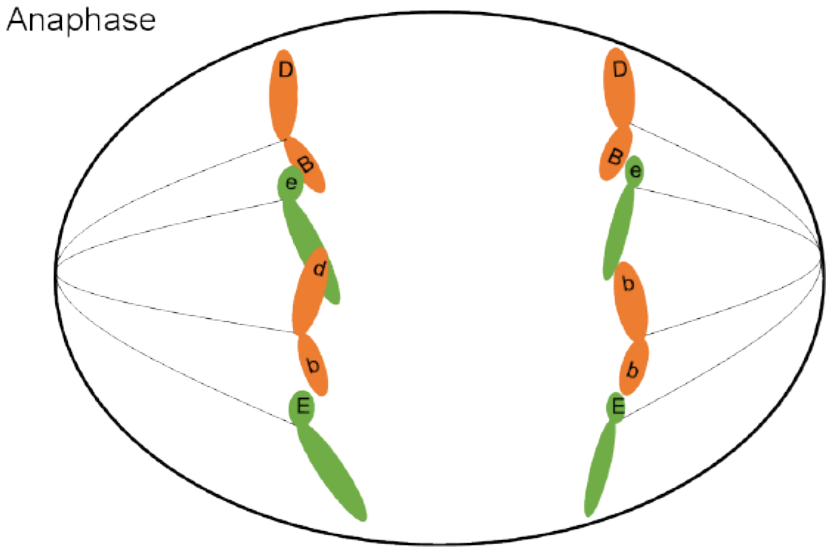


Figure 6. Anaphase of mitosis. Image by Marjorie Hanneman.

Mitosis: Telophase

The final stage of mitosis is telophase (Figure 7). A nuclear envelope will form around each bundle of chromosomes. The cell will undergo cytokinesis and the cytoplasm is split between the two identical daughter cells. Cell membranes (and cell walls in plants) form between the two cells. The chromosomes begin to decondense. They loosen up and spread around the nucleus because they no longer will need to move around. The new cells have now begun their lifecycle.

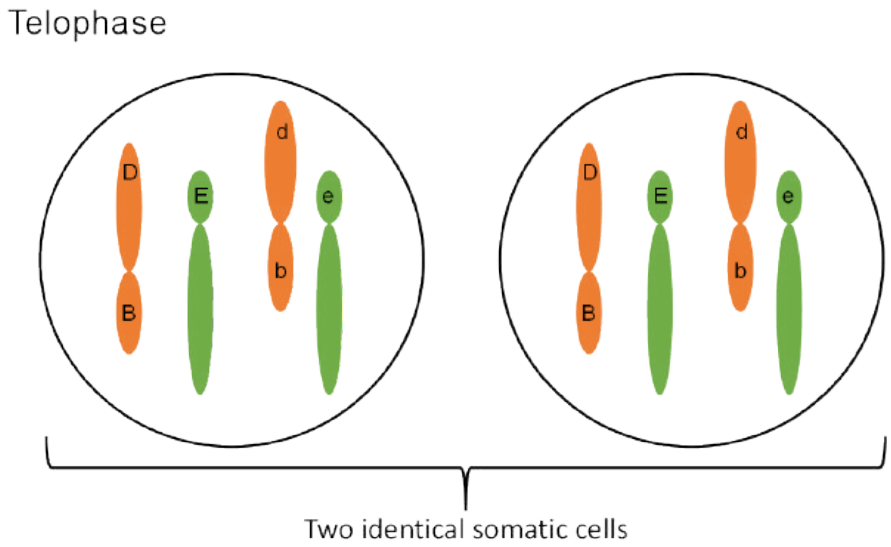


Figure 7: Telophase of Mitosis. Two cells produced from one cell. Image by Marjorie Hanneman.

Meiosis: Gamete formation

The objective of meiosis is to make four cells from a single somatic cell. The four cells each have half the chromosome number found in the somatic cell. In our human bodies, the four gametes will each have 23 chromosomes which means the 46 chromosomes in the somatic cell must replicate during interphase prior to meiosis just as they would before mitosis. Meiosis occurs in specialized cells of the body called germline cells.

To appreciate meiosis and gamete formation it is important to first understand two ideas, chromosome sets and homologous chromosomes.

Chromosome sets: The 46 chromosomes you have consist of two sets. You are a diploid organism ('di' means two and 'ploid' means sets). One set of chromosomes came from each parent when their gametes fused. Therefore, human gametes are haploid (one set).

Homologous chromosomes: The 46 chromosomes in a somatic cell can be arranged into 23 homologous or similar pairs. One chromosome from each pair came from the male parent, the other from the female. Homologous chromosomes have the same genes although in heterozygous people the genes would be different alleles (**A**,**a**). The exception to this would be the sex (X and Y) chromosomes. Passing on a complete set of human genes requires one chromosome from each pair to end up in each gamete.

There are several key differences between meiosis and mitosis that are summarized in the following table:

Table 1. The key events that happen in each of the stages of meiosis are summarized.

Mitosis	Meiosis
Chromosome number stays the same	Chromosome number is halved
One division occurs to make two cells. Four stages of this division.	Two divisions occur to make four cells. Eight stages in these divisions.
Similar or homologous chromosomes do not pair.	Homologous chromosomes pair during prophase l. Pairing is called synapsis.
Crossover exchanges between homologous chromosomes is rare.	Synapsis allows crossing over between homologous chromosomes.
Two cells made are genetically identical.	Four cells made are genetically different.

Meiosis: Prophase I

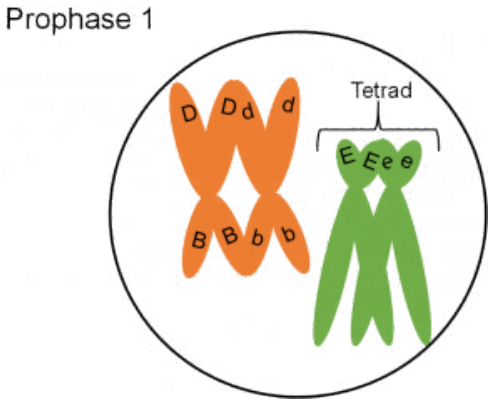


Figure 8. Prophase l of meiosis. Images by Marjorie Hanneman.

This stage starts meiosis and is the same as prophase of mitosis with one important change. As the chromosomes condense, they form groups of four chromatids called tetrads or bivalents. Close inspection reveals that each chromosome is replicated and consists of two sister chromatids. The two chromosomes in each cell that are homologous and have the same genes (but perhaps different alleles if the organism is heterozygous) will pair closely. This close association, or synapsis, allows the homologous chromosomes to crossover and exchange identical parts. The impact of crossing over is that genes that are on the same chromosome (Figure 8) can be recombined so that they are not always inherited together. The tetrad or bivalent formed during synapsis remains assembled as prophase progresses. The tetrad therefore moves as a unit to the center of the cell.

Meiosis: Metaphase I

Metaphase I starts when the tetrads are at the center of the cell (Figure 9). The tetrads have stayed together which insures that during the first division, each cell will get one chromosome from each homologous pair. The chromosomes remain at the center of the cell until the homologous pairs are ready to move away from each other.

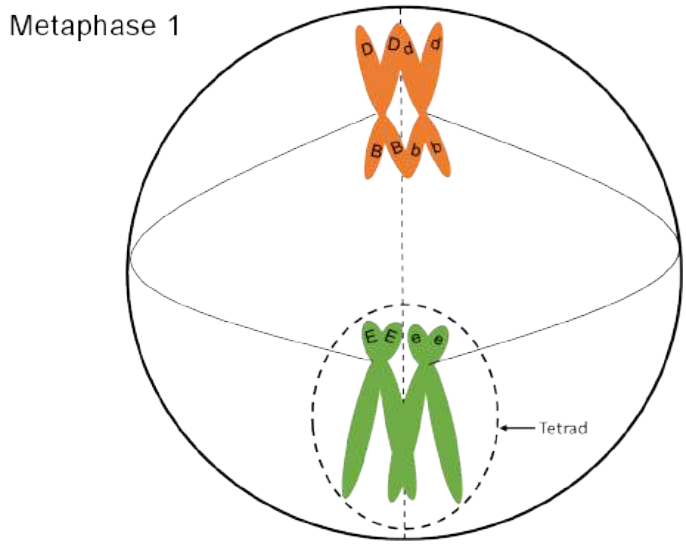


Figure 9. Metaphase I of meiosis. Image by Marjorie Hanneman.

Meiosis: Anaphase I

The chromosomes that make up each tetrad separate during anaphase I (Figure 10). However, the sister chromatids will stay connected at the centromere. Anaphase I proceeds until the chromosomes are pulled into a bundle at opposite ends of the cell.

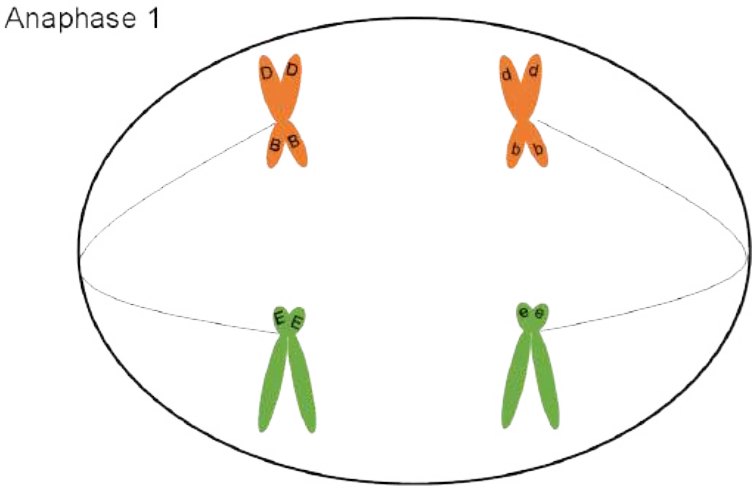


Figure 10. Anaphase l of meiosis. Image by Marjorie Hanneman.

Meiosis: Telophase l

The cell divides into two cells during telophase I (Figure 11). The bundle of chromosomes may have a nuclear envelope develop around them. The germline cells in some organisms such as human females, go through the first four stages of meiosis prior to birth. The **germline cells** remain at telophase I for some time. The second round of division occurs

when the gamete is needed for reproduction. In other situations, telophase I is an abbreviated stage, and the second round of division proceeds without delay.

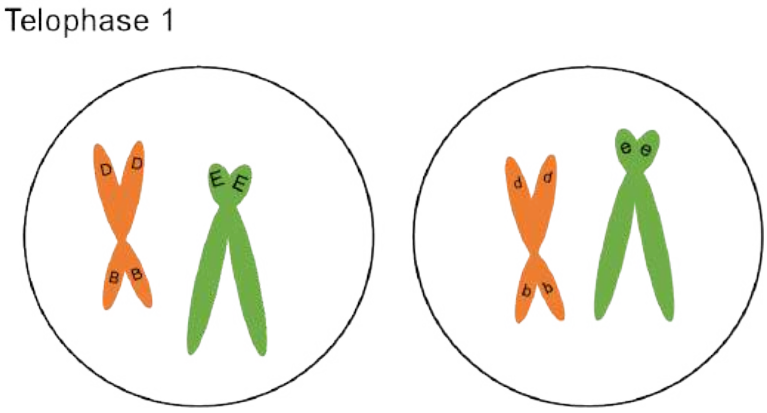


Figure 11: Telophase l of meiosis. Image by Marjorie Hanneman.

Meiosis: Prophase II

If the chromosomes became dispersed in telophase I, they will condense again at prophase II. The spindle apparatus moves the chromosomes to the middle of the cell. Look! The centromeres are still holding the sister chromatids together (Figure 12).

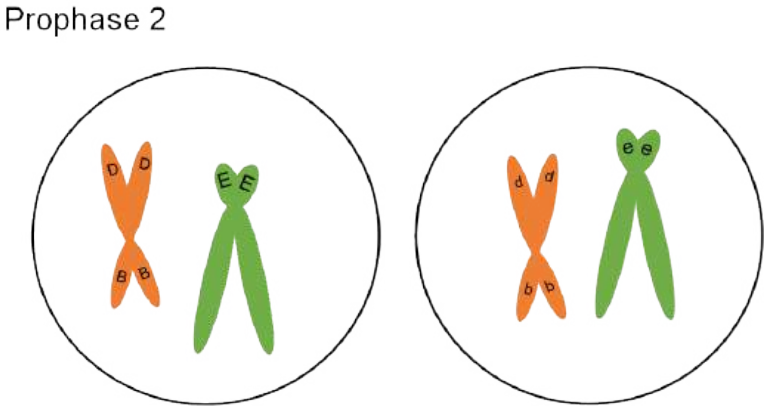


Figure 12. Prophase II of meiosis. Image by Marjorie Hanneman.

Meiosis: Metaphase II

In metaphase II the chromosomes are aligned at the center of the cell (Figure 13). This time there are not homologous chromosomes to be paired with. This metaphase looks similar to metaphase of mitosis but there is a key difference. What is the difference? (Compare Figures 5 and 13).

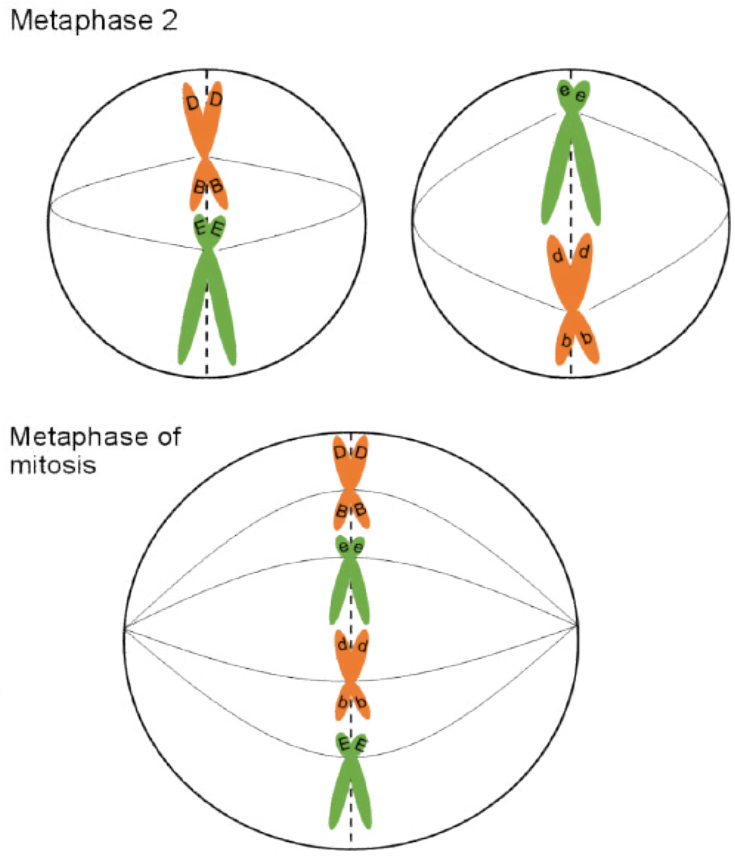


Figure 13. Metaphase II of meiosis and Metaphase of mitosis. Image by Marjorie Hanneman.

Meiosis: Anaphase II

During anaphase II, the chromatids are pulled apart by the spindle fibers. Now they are classified as chromosomes, not chromatids. The chromosomes move apart to opposite ends of the cell (Figure 14).

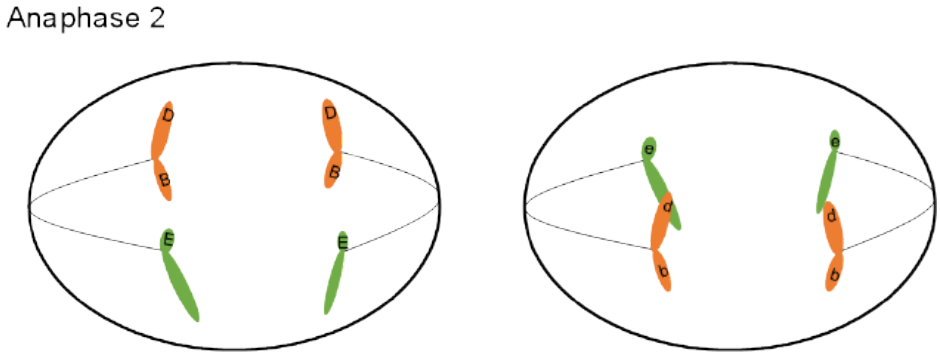


Figure 14. Anaphase II of meiosis. Image by Marjorie Hanneman.

Meiosis: Telophase II

In the final stage of meiosis, telophase II, the nucleus forms around the bundle of chromosomes (Figure 15). The cell divides. Now four cells exist that originated from one germline cell. Each cell is a gamete with half the number of chromosomes and genes as a somatic cell.

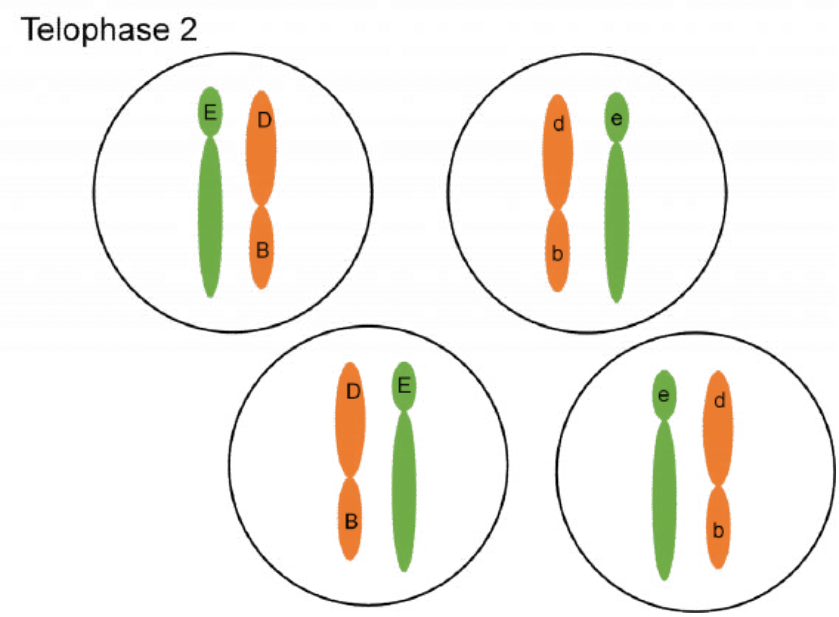


Figure 15. Telophase II of meiosis. Image by Marjorie Hanneman.

Parallel Behavior and the Chromosome Theory

The successful completion of mitosis or meiosis requires the cell to move large objects with precision and control many detailed events. The process surpasses any engineering accomplishments of NASA. The importance of mitosis and meiosis to an organism is obvious when we consider that genes are a part of the chromosome and the genes must be copied and distributed properly to produce viable daughter cells. The mechanisms of these events are far from being completely understood. From our current understanding, we can appreciate how the principles of segregation and independent assortment are controlled by the mechanics of meiosis.

When cell division was first observed and described by cytogeneticists, biologists were just beginning to accept the idea that genes were tiny objects that controlled traits and existed in the cells of living things. Two biologists, a German named Boveri and an American graduate student named Sutton, recognized that chromosome behavior during meiosis matched Mendel's principles of gene behavior. Both scientists proposed the idea that while genes had not yet been directly observed, they must be a part of the chromosome. Sutton and Boveri are both given credit for proposing this chromosome theory; genes are a part of chromosomes.

Segregation predicts gene behavior that matches the chromosome behavior observed by cytogeneticists. Genes are in pairs because chromosomes are in pairs. The gene pairs associate during gamete formation when the homologous chromosomes pair in prophase I. The gene pairs separate when the homologous pairs separate in anaphase I followed by chromatid separation in anaphase II. Thus, chromosome behavior dictates the gene's segregation behavior.

Independent assortment of gene pairs also correlates with chromosome behavior. Let us consider a dihybrid individual

with the genotype **BbEe** (Figure 16). How many kinds of gametes can it make? The four possible combinations (**BE**, **bE**, **Be** and **be**) are made at equal frequencies. We can understand why this occurs if we think about what happens to chromosomes at metaphase I of meiosis. The chromosome pair that carries the '**E**' and '**e**' genes will be moved independently from the chromosome pair with the '**B**' and '**b**' chromosomes. When the chromosomes arrive at the center of the cell at metaphase I, the two tetrads may be aligned so that the '**E**' and '**B**' genes move to one cell and the '**e**' and '**b**' genes move to the other during the first division. In the other half of the cells that go through meiosis, the tetrads will line up so that the '**E**' genes will be passed on with the '**b**' genes and the '**e**' genes will go with the '**B**' genes. Keep in mind that organisms that make gametes make thousands of them. The chance alignment of the tetrads at metaphase I therefore dictates the overall frequency of gametes with different combinations of genes when the genes are on separate chromosomes.

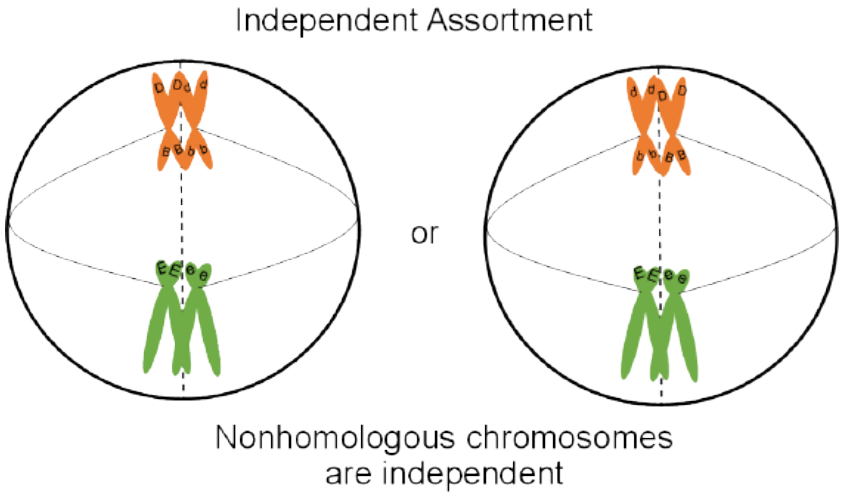


Figure 16. Alignment of chromosomes during metaphase I will determine the gene combinations that segregate into gametes. Image by Marjorie Hanneman.

When more gene pairs are considered, the same scenarios described above will be true as long as the genes are on separate chromosomes. When genes are on the same chromosome, the role of crossing over on gene inheritance needs to be considered in more detail. We will cover the inheritance of genes on the same chromosome in a later lesson.

Learning Activities

1. Watch this video about Mitosis from Daily Med Ed
2. Watch this video about Meiosis from Daily Med Ed

2. DNA: The Genetic Material

WALTER SUZA; DONALD LEE; PHILIP BECRAFT; MARJORIE HANNEMAN; AND PATRICIA HAIN

Learning Objectives

1. Detail the molecular nature of a gene.
2. Recognize when and where replication of the DNA that is packaged in an organism's chromosome is happening for growth, maintenance, and reproduction.
3. Describe how chromosome replication speed can be increased with multiple origins of replication and bidirectional replication.
4. Explain the role of each of the primary DNA replicating enzymes in conducting replication of a chromosome. Include the role of these enzyme in ensuring the accuracy of the replication process.
5. Predict how the semiconservative replication of a double stranded DNA will happen given the structure of the molecule and the specific function of the DNA replicating enzymes.

Introduction

The genetic (hereditary) material for all living things is composed of DNA (deoxyribonucleic acid). The structure of DNA must enable this substance to store coded information that control the biological function of cells. The genetic material transmits this hereditary information in a stable form for the cell and organism through accurate **replication** of DNA. Although DNA is capable of change (we will discuss how this happens when we talk about DNA mutations), the replication process ensures high accuracy in copying the genetic information so that all progeny cells receive the same information. The DNA from all chromosomes in a human cell would be more than 6.5 feet long! Therefore, to fit inside the nucleus of a cell, DNA is packaged into chromosomes.

Read this article about Watson and Crick for more insight on the experimental facts discovered by chemists and biologist which contributed to the determination of the structure of DNA.

Chemical Structure of DNA Subunits

DNA is a polymer made of nucleotide subunits. A **nucleotide** consists of 3 chemical groups; a sugar, a phosphate and a nitrogenous base (Figure 1). In the case of DNA, the sugar is deoxyribose.

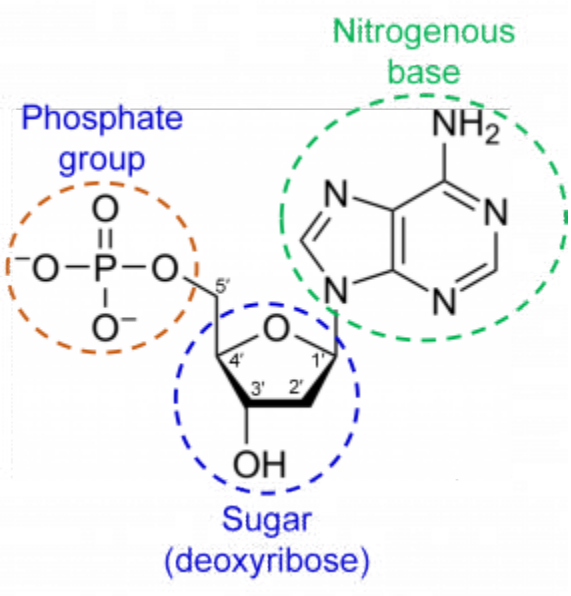


Figure 1. A nucleotide consists of three chemical groups. Image adapted from Wikipedia.

There are 4 different nucleotides in DNA that contain 4 different **bases** (Figure 2). These bases are named adenosine (A), cytosine (C), guanine (G) and thymine (T).

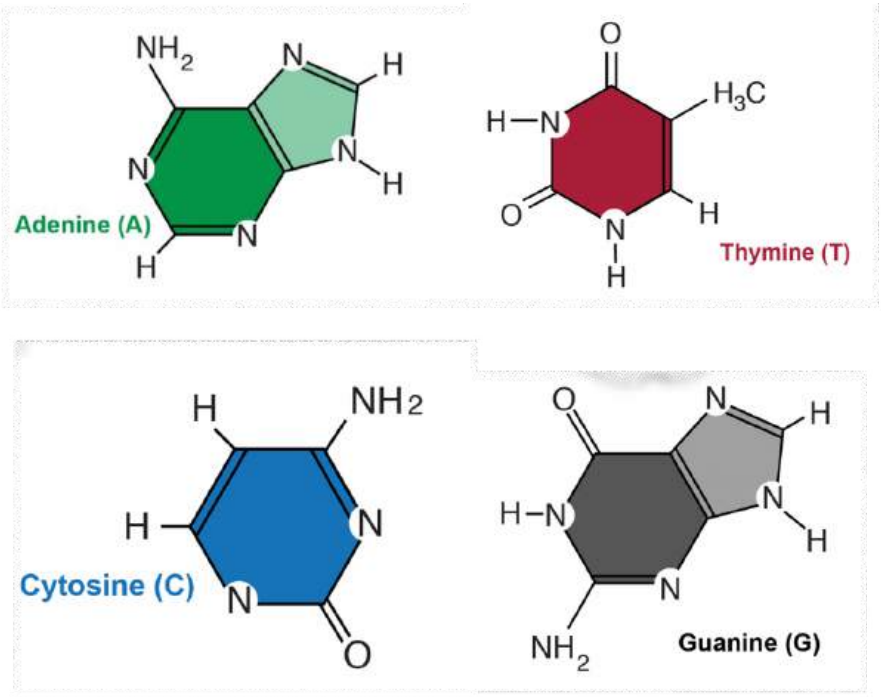


Figure 2. Adenine, Thymine, Cytosine, and Guanine are the nitrogenous bases of DNA. Adapted from NIH-NHGRI.

Complementary, antiparallel DNA strands

DNA codes information by ordering the sequence of the A,T,G and C nucleotides in a long polymer. The nucleotides are connected by phosphate bonds to form a strand. Two strands form a double stranded molecule (Figure 3). Note that in Figure 1, the carbon atoms on the deoxyribose sugar are numbered. The phosphate bond connects carbon #3 (the 3 prime or 3′ carbon) of one sugar to the 5′ carbon of the next. This gives a DNA strand directionality because the 5′ carbon faces one way and the 3′ carbon faces the other.

The bases on one strand form hydrogen bonds with the bases on the other. This is called "**base pairing**" and it is very specific, A only pairs with T and C only pairs with G. Thus, the sequences on the two strands of DNA are "**complementary**"; whenever there is an "A" on one strand, there is a "T" in the corresponding position of the complementary strand.

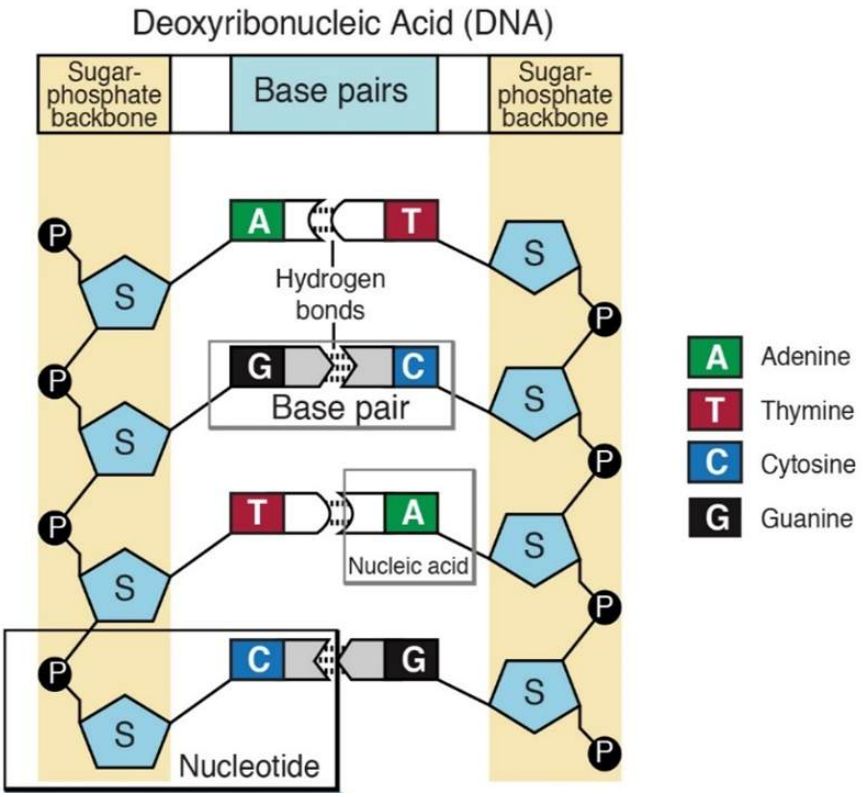


Figure 3. DNA is a double-stranded molecule. Illustration by NIH-NHGRI.

In addition to containing complementary nucleotide sequences, the strands are in opposite orientations. The 5′ end of one strand is oriented toward the 3′ end of the other. For this reason, the strands are referred to as "**antiparallel**".

Double helix

The final feature of the molecular structure is that DNA assumes a helical conformation (Figure 4). This is the most stable configuration that accommodates all the molecular structures and chemical bonds that make up DNA.

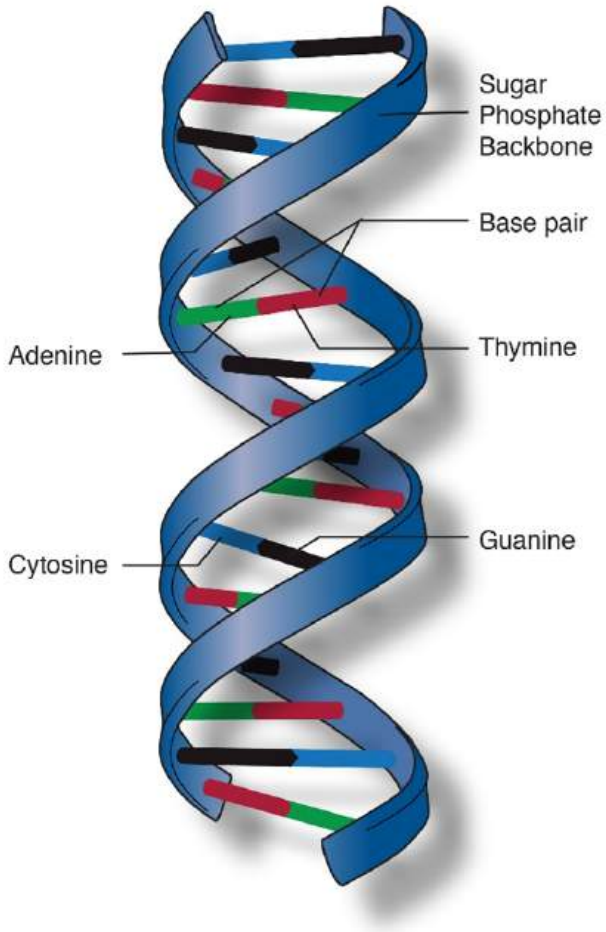


Figure 4. DNA is a double helix. Illustration by NIH-NHGRI.

Activity

Given the following sequence of a DNA strand, predict the sequence and orientation of the complementary strand.

5′-G-A-C-C-G-T-A-A-T-C-G-C-3′

Show Answer
Answer: 3′-C-T-G-G-C-A-T-T-A-G-C-G-5′

Packaging of DNA into chromatin and chromosomes

Chromosomes are long double stranded DNA molecules. Over 150,000,000 nucleotide pairs make up the human X chromosome. The complete replication and orderly transfer of something this large requires the chromosome to be packaged for stability and organization which creates genetic stability in the cells of a species. Every cell in a diploid

organism contains two copies (2n) of every chromosome present in that organism (Figure 5). For example, humans have 46 chromosomes in their body, 23 were inherited from the father and 23 from the mother. **Gametes**, the reproductive cells of an organism, (egg or sperm), have only one set (1n) of chromosomes. When the two gametes unite, they form a living embryo with two sets of genetic information. Therefore, we actually have two copies of the genetic information for each trait. Sometimes one copy controls **trait expression**, and other times both copies influence a trait. As a result, the offspring will have characteristics of both the mother and the father.

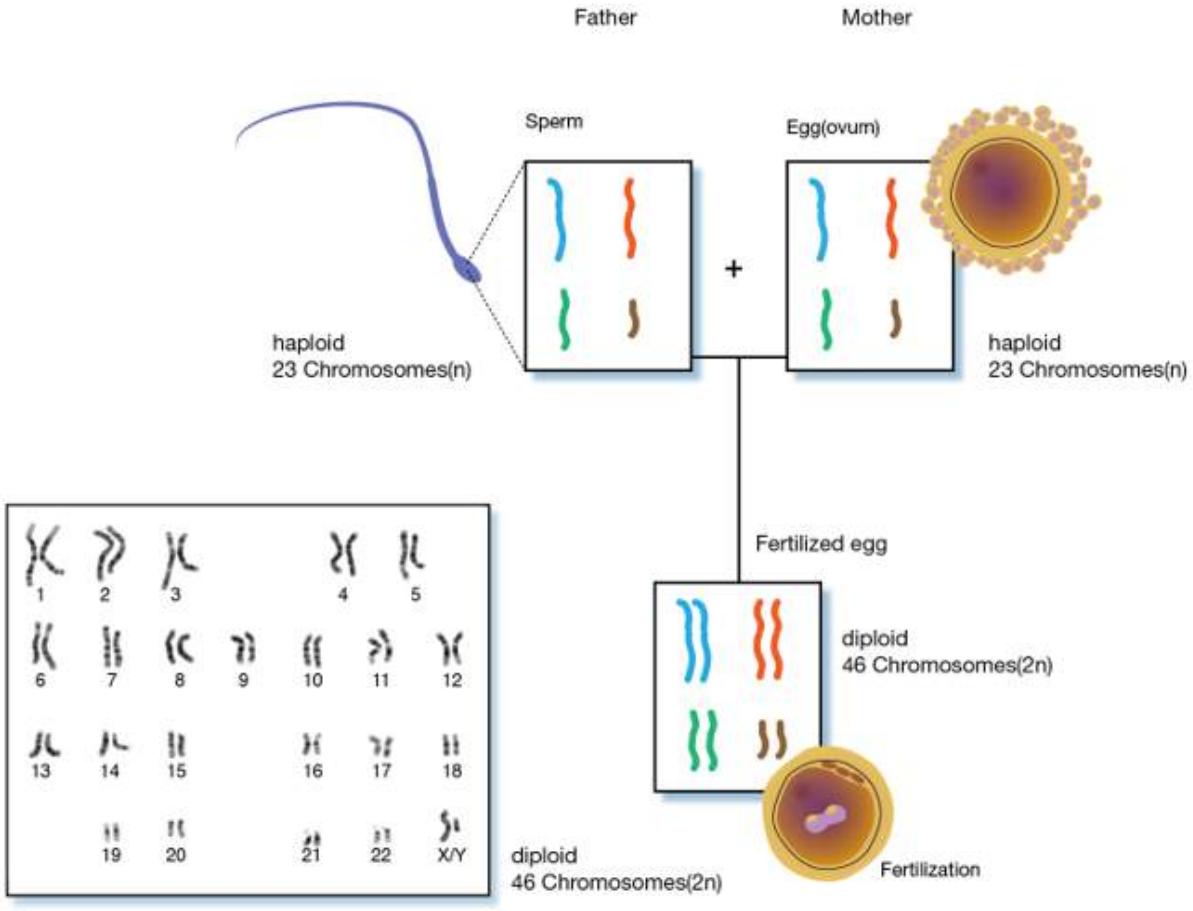


***Figure 5**. Human non-sex cells are diploid (2n) and contain 46 chromosomes. On the other hand, human sex cells or gametes (sperm and egg) are haploid (n) and contain single set of 23 chromosomes. Image from NIH-NHGRI.*

Inside cell nuclei, DNA is assembled, or **packaged**, into material called chromatin, which is composed of both DNA and proteins (Figure 6). Chromatin functions to protect and regulate DNA, as well as to efficiently store the very long DNA molecules that make up chromosomes within the limited space of the nucleus.

DNA from higher eukaryotic organisms is about 109 to 1010 bp. Human and maize chromosomal DNA is about 109 bp, which is equivalent to 1.8 m if all chromosomal DNA were stretched end to end in a linear manner. The diameter of the nucleus is only about 4-6 μm making it a challenge to fit chromosomal DNA inside the cell nucleus. This is accomplished by the ability of DNA to assume a very condensed structure to fit inside the nucleus. Therefore, the organization of eukaryotic DNA into chromatin is an important aspect of DNA packaging.

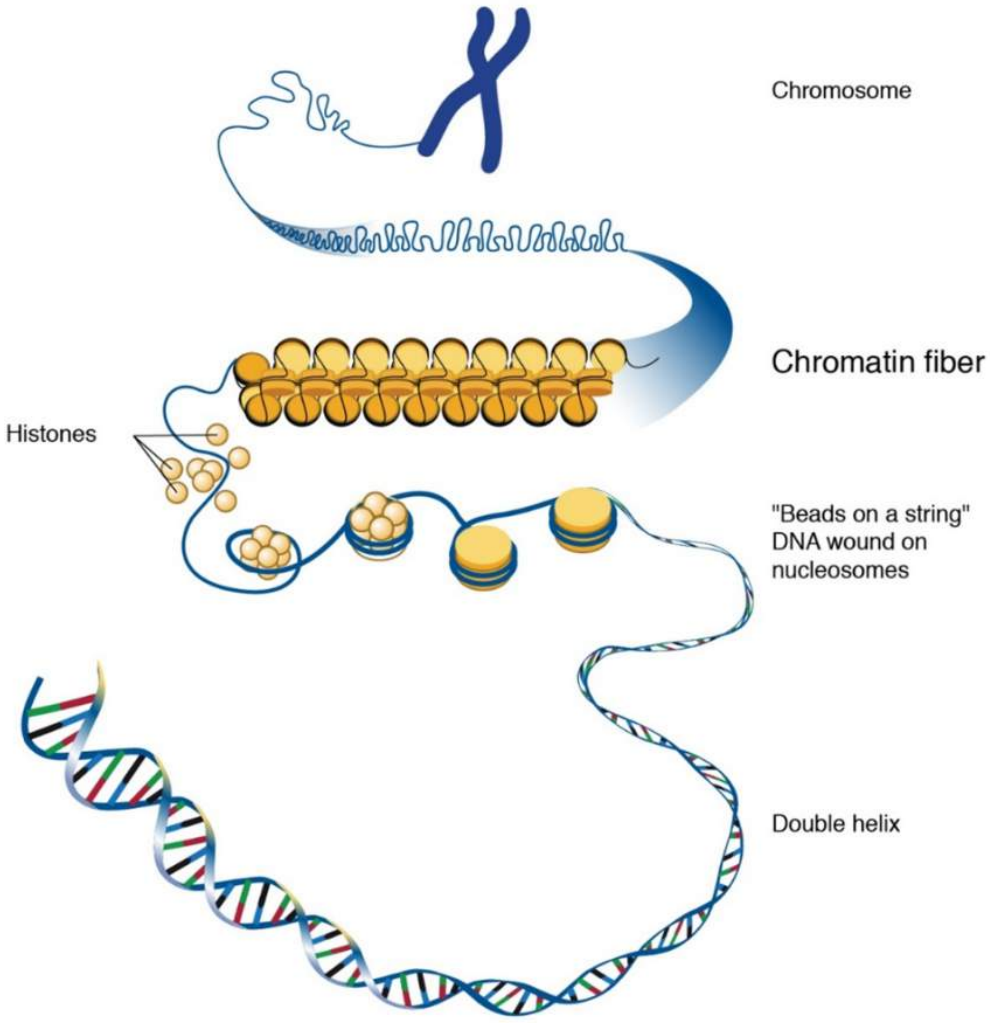


Figure 6. DNA is packed into chromatin and chromosomes. Illustration by NIH-NHGRI.

Nucleosomes and histones

The ordered coiling of chromosomal DNA around a histone protein core forms the **chromatin**. Chromatin is made up of **nucleosomes** (Figure 6) which represent the association of chromosomal DNA with histone proteins. A nucleosome is made up of about 145-147 base pairs of DNA coiled around each histone octamer, for about two complete turns. A histone octamer consists of two copies of each core histone. The total mass of the histones in the nucleus approaches that of the DNA – 2 molecules of each core histone to approximately 200 bp of DNA.

Higher order structures

If the higher order chromatin structure is disrupted, electron microscopy reveals the appearance of "beads on a string" with a diameter of about 10 nm. The "beads" represent the DNA wrapped around histones. Nucleosome formation results in a DNA fiber that is about 10 nm and a packaging ratio of about 7. The higher order chromatin structure results when the 10 nm fiber is coiled into a solenoid. The result of nucleosome coiling is a chromatin fiber of 30 nm that is observed by electron microscopy.

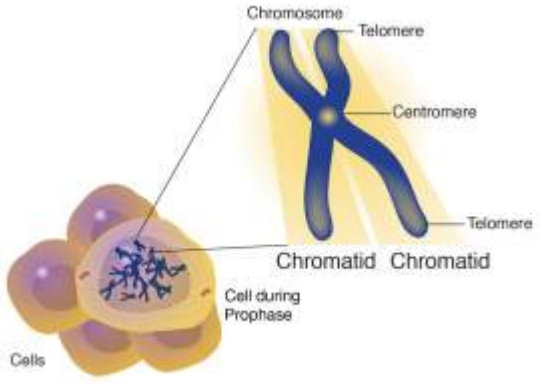


Figure 7. Chromosomes have centromeres and telomeres. Illustration by NIH-NHGRI.

Centromeres

The centromere is a chromosomal region that controls chromosome segregation at mitosis and meiosis (Figure 7). Centromeres connect to microtubules of the spindle apparatus, which directs their movement to opposite poles (daughter cell nuclei) during cell division. Scientists were successful in isolating these centromeric sequences from yeast and engineered brewer's yeast (*Saccharomyces cerevisiae*) plasmids that were able to replicate like the chromosomes. Through these efforts scientists were able to point centromeric function to a DNA stretch of about 120 bp that was resistant to DNAse and bound to a single microtubule.

Telomeres

The telomere lies at the end of the chromosome and confers stability by "sealing" the end of a chromosome (Figure 7). Telomeres consist of long series of short repeated DNA sequence that occurs in tandem arrays and are added to the end of the chromosome during DNA replication by an enzyme called telomerase. In most plant species the sequence **TTTAGGG** constitutes a conserved telomere motif.

The knowledge about the structure and function of centromeres and telomeres has had a profound impact in molecular genetics. For example, yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) have proven useful in the physical mapping of plant genomes requiring the cloning and multiplying large (>100 kb) DNA fragments in yeast or bacterial cells.

DNA Replication

In 1953, J.D. Watson and F.H.C. Crick published a note in NATURE; one of the world's most widely read research journals. The note cited just six references.

Watson and Crick's writing had limited chemistry details given the title of their note, "A Structure for Deoxyribonucleic Acid". A later article would share more of the chemistry behind their proposed double helix structure for Deoxyribonucleic Acid (DNA). This note in Nature was written for a broader audience, including biologists who recognized that the structure of DNA was a prerequisite for understanding the function of the genetic material.

They included a diagram to allow readers to visualize this unique double helix structure. To emphasize the structure determines function importance, Watson and Crick included this statement in their note…

> "It has not escaped our notice that the specific pairing we have postulate immediately suggests a possible copying mechanism for the genetic material."

Accurate replication prior to cell division is one of the three key functions of the genetic material. While later experiments would reveal the details of how cells replicate DNA, Watson and Crick were obviously impressed that the proposed double helix structure would fit with the concept that living cells are needed to produce more living cells and the transmission of a complete set of biological instructions. The genetic material must have a structure that lends itself to being copied. Using old to help build new leads to a proposed semiconservative model for replicating the double stranded DNA molecule (Figure 8).

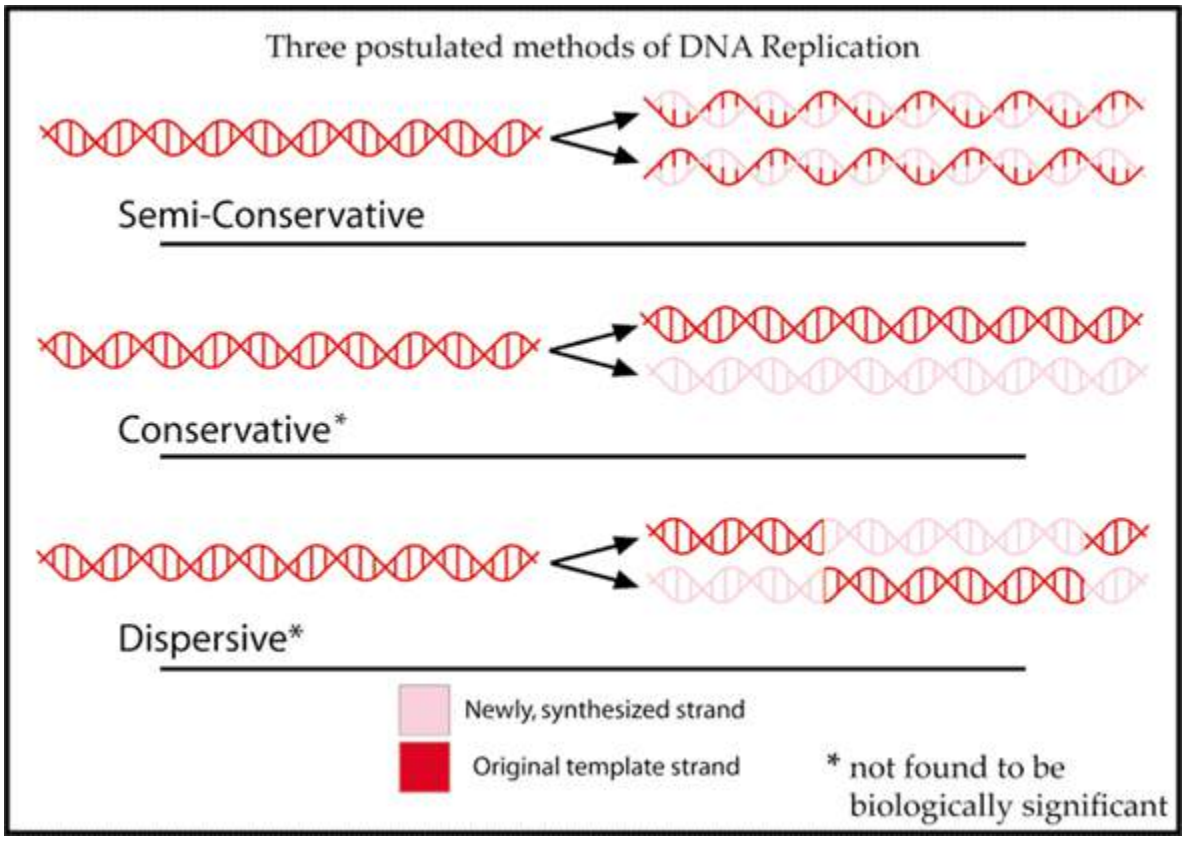


Figure 8. Watson and Crick predicted that the double helix molecule could be replicated in a living cell by the semi-conservative method. Each strand of the old double stranded molecule is read to build a new strand. Early experiments on replication confirmed this prediction. Image Source: Adenosine, CC BY-SA 2.5, via Wikimedia Commons.

Prior to both mitosis and meiosis cell division in multicellular organisms, and cell division in prokaryotes, the DNA inside the cell must be replicated. (see lesson Mitosis and Meiosis). This replication process generates the genetic information needed for either two cells, genetically identical with the original cell or in sexually reproducing organisms, four gamete cells with half of the original cell's genetic information.

When and Where:

DNA replication will happen before cells divide. In hours old embryos that are developing into multicellular organisms, all the cells are replicating their DNA and dividing into new cells. Hours or days later, multicellular organisms have developed specific meristem regions where new cell production occurs. A visual example of meristem cells where DNA replication is needed is the tip of a plant root (Figure 9). New cells made at the root tip allow the root to grow. The root must also replace the cells that are damaged as the growing root moves through soil. In Figure 8, the meristem region (1) contains meristem cells that are replicating their chromosomes, then dividing into two identical cells by mitosis. Cells on the tip part of the meristem specialize in root cap cells (2,3) with the job of protecting the meristem. The fate of these cells is to die (4), and they need to be replaced by new cells made in the meristem. Cells made on the other half of the meristem (5) can elongate and eventually specialize. These root cells will then function for the plant, all season for annual plants, years for perennial plants. Replication of DNA will only happen in the meristem cells.

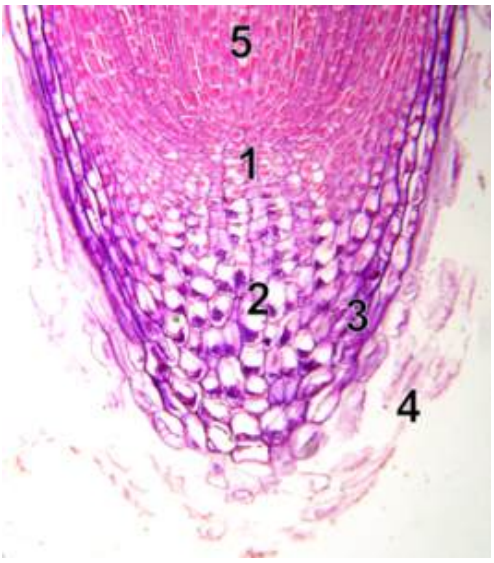


Figure 9. Plant root tip. Image Source: SuperManu, CC BY-SA 2.5, via Wikimedia Commons.

Speed of chromosome replication

Fifty-two years after Watson and Crick's Nature note, a group of about 50 molecular geneticists published the complete DNA sequence of the human X chromosome (Figure 5). This discovery revealed that X chromosome is one double-stranded DNA molecule that is about 155 million nucleotide pairs long. Every time a human cell divides, one or two of these 155,000,000 nucleotides must be replicated. This must be done with speed and accuracy.

The main DNA replication enzyme (DNA polII, see below) works fast. Estimates of its replicating speed are around 750 nucleotides per second.

Activity

Let us do the math.

155,000,000 nucleotides / chromosome X 1 second / 750 nucleotides

= 206,666 seconds per X chromosome

Let us convert these seconds to hours.

1 hour / 360 seconds X 206,666 seconds/chromosome = 547 hours per chromosome

That means one replication enzyme would take 547 hours or about 24 days to replicate an X chromosome. What are the biological ramifications of this 24 days? If you cut yourself, it would take 24 days to replicate the chromosomes before the skin cells could divide. The formation of new cells needed for healing cuts in our skin would keep us in a wounded state.

To speed up the chromosome replication process, two tactics are used by living cells. Both tactics are shown in Figure 10. First, replication will not start on just one end of the chromosome. Instead, there are hundreds of origins of replication (ori) along the chromosome. In addition, the replication process moves bidirectionally from the ori. This bidirectional replication creates forks of replication that move toward each other as the chromosome is replicated.

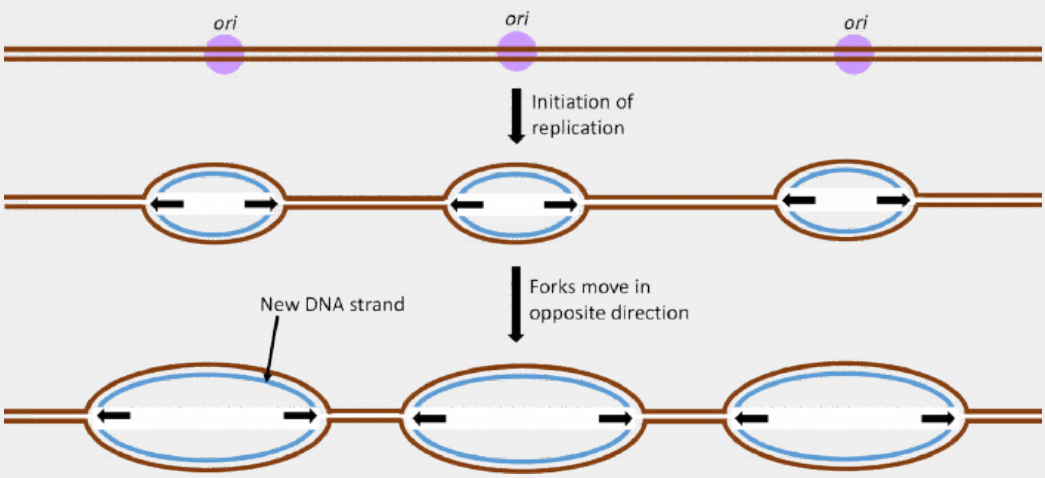


Figure 10. Origin of replication and replication forks. Image by Walter Suza.

As a result of these two tactics, a large chromosome can be replicated in hours rather than weeks and cell division can happen quickly when it is needed for growth or cell replacement.

The Chromosome Replicating Team

Like all processes in living cells, a collection of specific proteins works together to perform functions which control and complete chromosome replication. All the proteins involved in DNA replication make and break bonds, so they are enzymes. The function of the main DNA replicating enzymes is described below.

Five of the main DNA replicating enzymes

- **Helicase**: This is one of the enzymes that unwinds the double stranded DNA molecule. The unwound single strands no longer hydrogen bond to their complementary strand but the sugar-phosphate bonds remain intact.

- **Topoisomerase**: The unwinding work of Helicase creates tension in the double stranded DNA ahead of the replication fork. Topoisomerase relieves this tension by catalyzing a series of sugar phosphate bond breaking and making steps. Without this tension relief, the Double stranded DNA could break and the intact chromosome cannot finish replicating.

- **DNA Polymerase III (DNA pol III)**: This is the main DNA synthesizing enzyme. The enzyme reads the single strand as a template and places in the complementary deoxyribonucleotide. DNA pol III reads the template in the 3′ to 5′ direction and builds the new strand in the 5′ to 3′ direction, adding the next nucleotide to the 3′ end by catalyzing sugar-phosphate bonding. DNA pol III illustrates the specificity of enzymes several ways; it reads and proofreads the placement of new nucleotides to insure accurate replication. It can only read and build in one direction. DNA pol III can only add nucleotides to a free 3′ end. This last specification means that DNA pol III cannot start the replication process on the single stranded template. Another enzyme needs to be part of the in vivo replication team to prime the work of DNA pol III.

- **Primase**: Biochemists named this enzyme to describe its role in starting or priming the replication process. DNA (or RNA) primase is a special RNA polymerase. The enzyme reads the single stranded DNA template 3′ to 5′ and adds ribonucleic acid (RNA) nucleotides in the 5′ to 3′ direction. Once a few hundred RNA nucleotides are added, primase falls off the template strand and leaves the 3′ end that DNA Pol III needs to continue the process.

- **DNA Polymerase I (DNA pol I)**: The DNA polymerase specializes in the removal of the RNA primers and replacing them with DNA nucleotides. The enzyme works with the same 3′ to 5′ reading and 5′ to 3′ building.

- **DNA Ligase**: After the five enzymes described above have completed their work, some sugar phosphate bonds need to be made to complete the double stranded molecule. DNA ligase has this backbone bond sealing assignment.

Now that we have introduced the replication enzymes, we can describe the step-by-step action of this team. We will focus on the action that happens at one origin of replication.

Step 1: Initiating replication at the Ori

Helicase enzymes bind to the ori sequences and start to unwind the double stranded DNA. This establishes two replication forks (one shown in Figure 10a) and a helicase enzyme will be working to unwind the DNA at each fork as the forks move away from each other. The unwinding exposes single strands of DNA that are the template for building a new strand. Topoisomerase can be seen working to relieve tension ahead of the replication fork in Figure 10a.

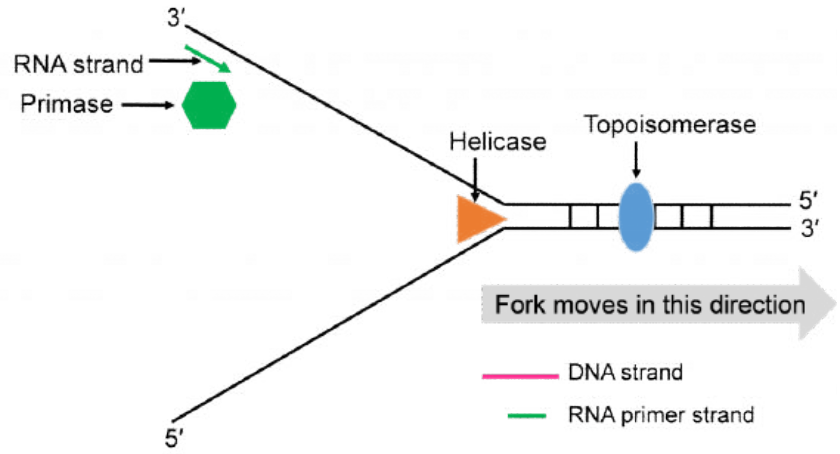


Figure 10a. Helicase and Topoisomerase work together to expose single strand DNA templates. Primase adds RNA nucleotides, reading the old DNA single strand template 3′ to 5′. Image by Walter Suza.

Step 2: Priming DNA with RNA

The process of priming with primase and then replicating with DNA pol III happens at both replication forks (Figure 10a and b). DNA pol III enzymes can replicate new strands as fast as the DNA is unwound at each replication fork. The unwinding creates two template strands at each fork so primase enzyme must work to prime both old strands.

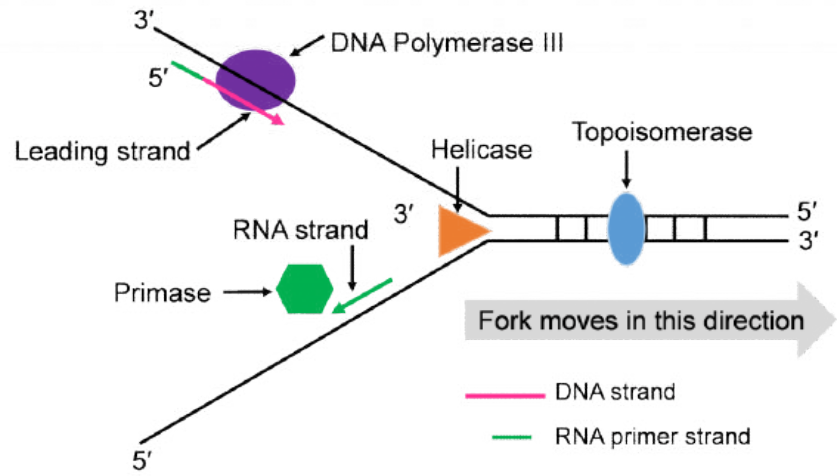


Figure 10b. DNA Polymerase III can add to the 3′ end of the RNA primer and start to read the template 3′ to 5′. Primase can prime the other template strand at this replication fork. Image by Walter Suza.

Step 3: Synthesis of leading and lagging strands

The double strands of DNA are anti-parallel and 'run' in opposite directions and DNA pol III can only work in one direction (Figure 10c). Thus, there is continuous and discontinuous replication happening at each replication fork. The strand that can be replicated as helicase unwinds is the leading strand. The DNA pol III must replicate the other strand in the opposite direction so this strand will lag behind. (Figure 10c).

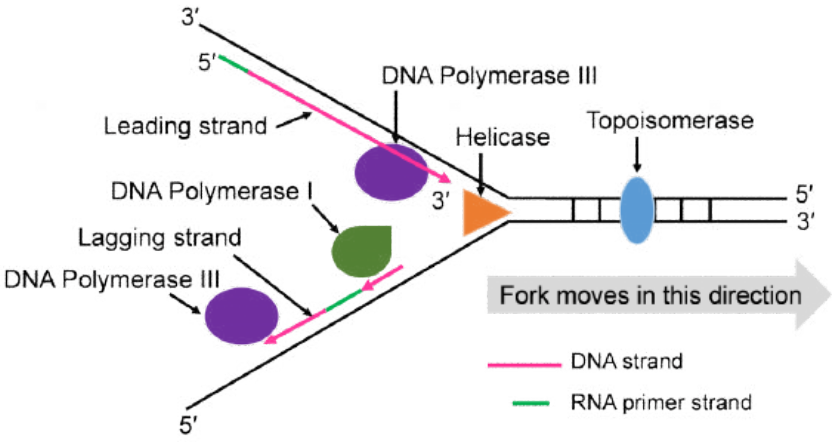


Figure 10c. Both leading and lagging strands are being made at this replication fork. DNA polymerase I removes RNA primer nucleotides and replaces them with DNA. Image by Walter Suza.

There is more priming needed to replicate the lagging strand and the priming leaves RNA-DNA hybrid stretches in the new double stranded molecule. This means there is work for the enzyme DNA pol I. This enzyme will be part of the removal of the RNA nucleotide primers, and replacement with DNA nucleotides (Figure 10c and d).

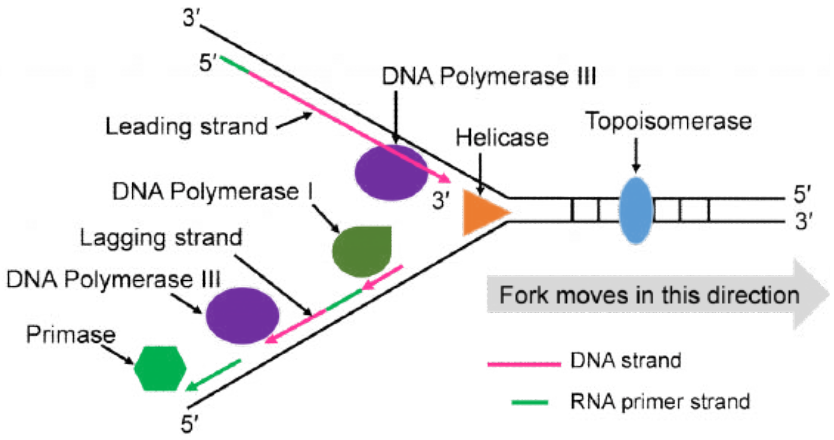


Figure 10d. Lagging strand synthesis must be primed multiple times with primase enzyme creating discontinuous replication. Leading strand just needs to be primed one time. Image by Walter Suza.

Both replication forks have a leading and a lagging strand of new DNA. The DNA nucleotide stretches between the red RNA primers would be the Okazaki fragments.

One of the first scientists to provide evidence to support this model of DNA replication was the team of Reiji and Tsuneko Okazaki in the 1960s. They used radioactive nucleotides to track the newly replicating fragments and found that many of the fragments were short. This fits with the discontinuous model.

As the four enzymes just described attend to their specific roles in replication, two double stranded DNA molecules with an old strand and a new strand are built. If the DNA pol III makes a mistake and adds the wrong, non-complementing nucleotide, the enzyme can proof read its work and replace these replication mistakes. When replication is perfect, the two double stranded DNA molecules will have identical sequences.

Step 4: Ligation of Okazaki fragments

The fifth enzyme that has a final role in completing replication is DNA ligase (not shown in Figure 10d and e). This enzyme will seal the sugar phosphate bond between the last replacement nucleotide added by DNA pol I and the first nucleotide that had been added after priming by DNA pol III. With this bond sealing work, the replication process is complete.

The two double stranded DNA molecules that get made will have identical sequences unless rare mistakes happen as the enzymes conduct their work.

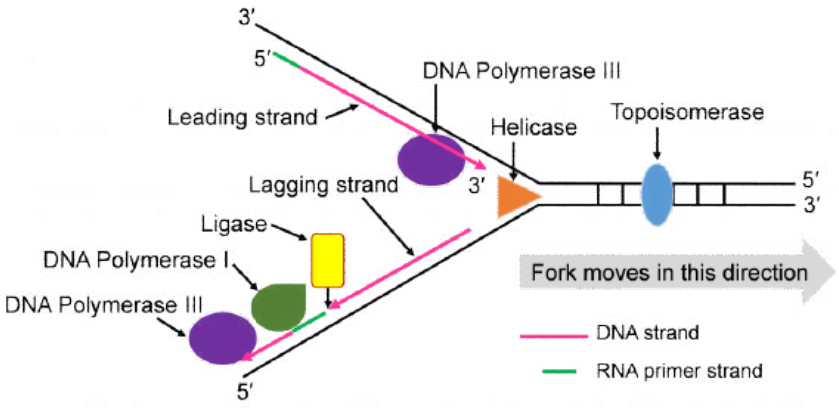


Figure 10e. Ligase must form a sugar-phosphate bond between the last nucleotide placed by DNA pol I and the first nucleotide placed by DNA pol III. Image by Walter Suza.

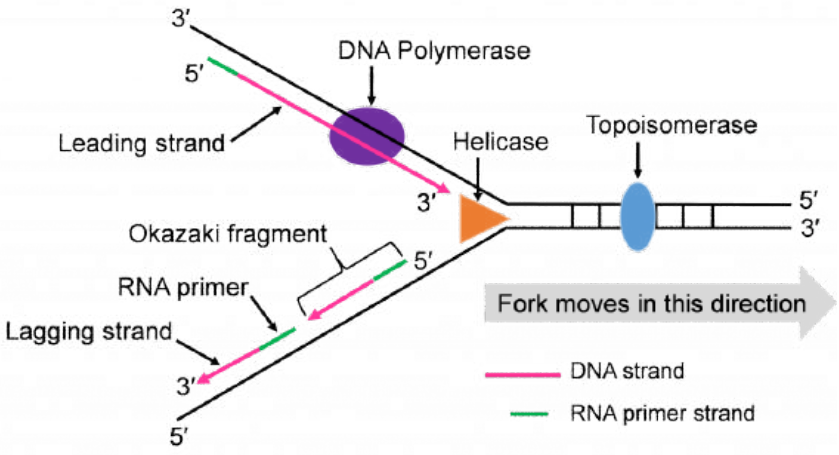


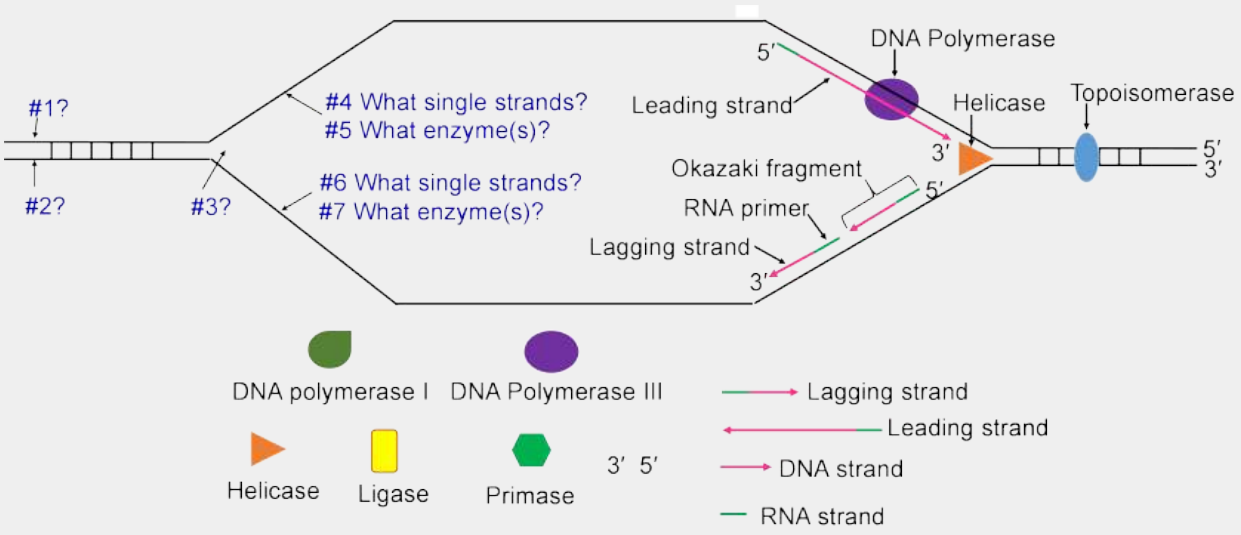
Figure 10f. Short fragments made by discontinuous replication were discovered by the Okazaki team and are called the Okazaki fragments. Image by Walter Suza.

Lesson Summary:

DNA is composed of nucleotides. Nucleotides are connected together by phosphate bonds to form a strand. The bases on one strand form hydrogen bonds with the bases on the other to form a double stranded molecule. The final feature of the molecular structure is that DNA assumes a helical conformation. To fit inside the nucleus, DNA assumes a very condensed structure. Therefore, chromosomal DNA is coiled around a histone protein core to form chromatin. The tight packaging of DNA in chromatin must be modified to allow DNA replication and transcription. In the process of replication, the two DNA strands separate and act as templates for the synthesis of complementary daughter strands.

Learning Activities

Place the single strands, replication enzymes, and 3′ / 5′ labels in the appropriate places (#1-#7)



3. DNA Mutations

WALTER SUZA; DONALD LEE; PHILIP BECRAFT; AND MARJORIE HANNEMAN

Learning Objectives

- List the types of mutation.
- Predict the effects of specific mutations in gene coding regions on the protein encoded by the gene.
- Describe ways in which mutations arise, naturally and experimentally.

Introduction

Crop improvement relies on genetic variation, which is synonymous with DNA variation. Mutations are the source of genetic variation that is the driver for trait improvement. In general, mutations are rare events, and most are deleterious and get selected against during the course of evolution. However, occasionally mutations create genetic variation that is beneficial. As such, it is often of interest for plant improvement or for genetic studies to induce genetic mutations by exposing seed or plant tissue to certain agents, called mutagens.

By definition, a mutation is a heritable change in DNA sequence. This can happen in several ways: substitution of a DNA base, insertion or deletion of one or more DNA bases, or by large-scale chromosomal rearrangements, the latter of which will not be considered here. It has also become clear in recent years that chromatin carries additional information to the DNA sequence in the form of base modifications such as cytosine methylation. This is known as **epigenetic** modification and can impact gene function and plant traits.

A. DNA base substitutions

The simplest type of mutation is a substitution of one base for another in the DNA sequence. Substitutions most often arise as errors during DNA replication or repair. The most common type is the **transition**, where one pyrimidine may be substituted by the other, or one purine by the other. The less common type is the **transversion**, in which a purine may be substituted by a pyrimidine or vice versa. As described in the next section, these can have various effects on gene function depending on where they occur in a gene.

B. Insertions & deletions

Insertions or deletions of DNA are other types of mutations that occur quite frequently. They can vary in size from one to thousands or more nucleotide bases. Their effects on gene function depend on the size of the insertion or deletion, and the location relative to the gene. As described below, insertions and deletions that are in multiples of 3 are referred to as in-frame insertions and deletions and result in the addition or deletion of amino acids from the protein sequence. When not in multiples of three, insertions and deletions lead to **frameshift mutations**, altering the amino acid sequence of the protein encoded by the mutated gene.

Figure 1. Example of DNA base substitution. In this example, the base G is substituted for T. Image by NIH-NHGRI.

Figure 2. Example of deletions. In this example, the base A is deleted. Image by NIH-NHGRI.

A common source of inserted DNA is from **transposons**, sequences of DNA that are able to move from one genomic site and insert into another. Transposons were discovered in maize by Barbara McClintock. For this discovery, she was awarded the Nobel Prize in Physiology and Medicine in 1983.

Some transposons copy themselves and multiply in number within a genome. Transposons can cause wide scale chromosome rearrangements and gene mutations. A feature of mutations caused by many plant transposons is that they are unstable. Excision of a transposon from the mutated gene can often restore it to wild type or create a new stable mutant allele. Transposons are found in every living organism studied and are a major driving force in genome evolution.

Figure 3. Example of a DNA insertion. In this example, the base A is inserted between T and C. Image by NIH-NHGRI.

Effects of mutations on gene function

A. Coding region mutations

Silent mutation

A silent mutation is a mutation that results in the change of a codon without a change in the amino acid represented by

the codon. Because of the redundancy of the genetic code, DNA base substitutions may not lead to the incorporation of an incorrect amino acid in the protein.

Missense mutation (amino acid substitution)

A missense mutation is a single nucleotide base substitution that alters a codon such that it codes for a different amino acid that is incorporated into the encoded protein. The amino acid substitution may affect the function of the protein if it occurs in a critical portion of the protein.

Frameshift mutation

Due to the triplet nature of the genetic code, an insertion or deletion can change the **reading frame** for the entire subsequent sequence. For example, if a particular sequence is read sequentially (Figure 4).

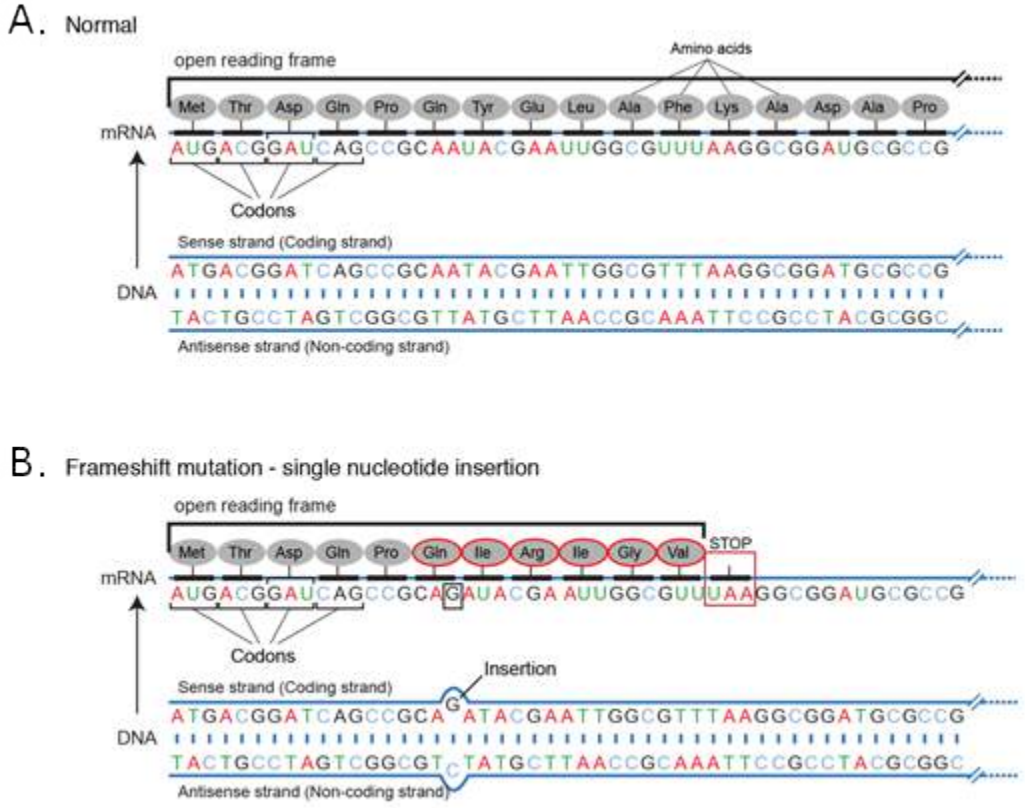


Figure 4. Impact of frameshift mutation on gene function. The normal gene sequence is shown in part A. The insertion of the base G shown in part B, results in a new protein. Image by NIH-NHGRI.

Because the new sequence of codons is different from the original, the entire amino acid sequence is changed from the point of insertion. A similar effect is seen for deletions. This effect is seen whenever the number of nucleotide bases inserted or deleted is not a multiple of 3 and it usually will result in the loss of the function of the protein.

Amino acid additions or deletions

On the other hand, if insertions or deletions are in multiples of 3, amino acid additions or deletions can arise. Consider the following 3-base insertion (XXX):

UUU CCC XXX AAA GGG (codons)

aa1 aa2 aa5 aa3 aa4 (amino acids)

If such insertions or deletions occur in important regions of a protein, they might impair function.

Nonsense mutation

A nonsense mutation arises when a functional codon is changed to a stop codon. Nonsense mutations may cause premature termination of translation.

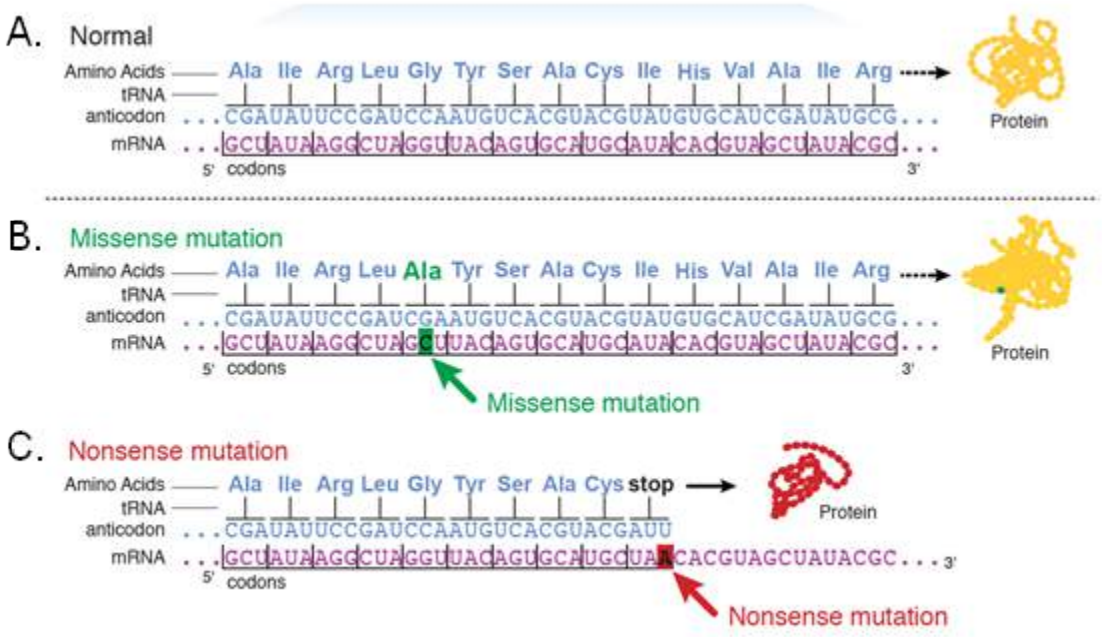


Figure 5. Impact of missense and nonsense mutations on gene function. The normal gene sequence is shown in part A. In part B, a missense mutation leads to replacement of the Gly by Ala. In part C, a nonsense mutation results in premature stop codon and shortening of the protein. Adapted from NIH-NHGRI.

B. Regulatory region mutations

Promoters and **enhancers** control temporal, spatial and quantitative aspects of gene expression. As such, mutations in these regulatory regions of a gene can alter the regulation of the gene's expression. Genes can become expressed at inappropriate times, places or levels, which can affect a plant's phenotype.

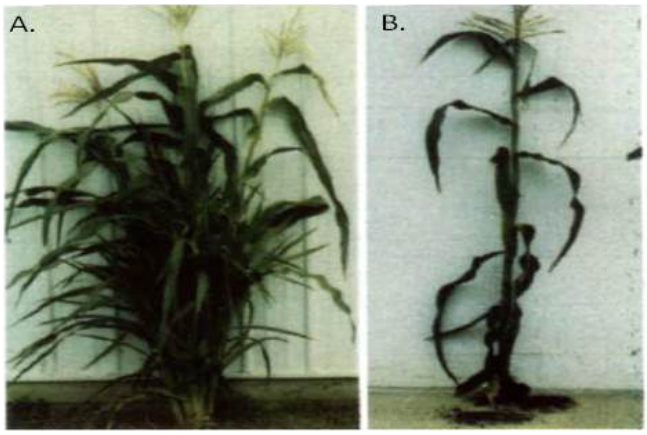


Figure 6. Regulatory region mutation of the tb1 gene converted teosinte (A) to modern maize (B). Image courtesy of John Doebley.

Teosinte, the progenitor of modern maize, produces multiple branches (Figure 6A) due to reduced apical dominance. In teosinte and modern maize, apical dominance is under the control of a gene called *teosinte branched1*, ***tb1*** (Doebley et al. 1997). An insertion of a transposon in the regulatory region of the ***tb1*** gene in modern maize causes the gene to be strongly expressed (Doebley et al. 1997; Studer et al. 2011), resulting in enhanced apical dominance, and suppression of excessive branching (Figure 6B).

C. Recessive vs. dominant mutations

Recessive mutations

Recessive mutations most commonly result from a loss of function. A loss of function mutation causes a decrease in the function of a gene. A complete loss of function is known as a **null mutation** while partial loss of function mutants are often referred to as **hypomorphs** or "leaky". Loss of function mutations can affect gene expression or the function of the resultant protein. Most loss of function mutations are recessive because in a heterozygote, the normal allele can provide the necessary function.

Dominant mutations

Dominant mutations can affect genes in several different ways. One general class is gain of function mutations. These encompass several different types.

1. **Overexpression mutants**. As the name implies, the affected gene is expressed at levels that are inappropriately high, causing too much of a gene product to be produced.
2. **Neomorphic mutations** arise when a mutation causes a qualitatively new effect not seen with normal alleles. This can occur when a mutation affects the regulation of a gene such that it becomes expressed in a new time or place in the plant. It can also happen when a mutation alters the properties of the encoded protein. For example, a transcription factor could be mutated to recognize a different promoter sequence, or an enzyme could recognize a new substrate
3. **Dominant negative mutations** alter a gene product such that the mutant gene product interferes with the function of the normal one. This can occur at the protein level. For example, a protein might function as a

multimeric complex and the presence of one mutant subunit could compromise the function of the entire complex. Dominant negatives can also occur at the RNA level for example if an insertion, deletion or rearrangement causes the wrong strand to be transcribed, creating an antisense or RNAi suppression effect.

Effects of mutations on traits

Example of a loss of function mutation – gene function is required for a trait

Sweetness is an important quality trait in maize (corn). One of the loss of function mutants used to improve sweetness in corn is ***sugary1*** (***su1***). The Su1 gene is expressed in the endosperm (Lertrat and Pulam, 2007), and it encodes and enzyme that is required for normal starch biosynthesis (Rahman et al. 1998). In the ***su1*** mutant, with defective starch biosynthesis, the concentration of sugar is 3 times higher than wild type making this mutant valuable for sweet corn (Lertrat and Pulam, 2007). In addition, ***su1*** kernels accumulate a highly branched polysaccharide called phytoglycogen, which gives the kernels a creamy texture. New commercial sweet corn hybrids are developed using a combination of starch biosynthesis mutants (Nelson and Pan, 1995; Lertrat and Pulam, 2007).

Example of a gain of function mutation – gene function is sufficient for a trait

The plant growth hormone, gibberellin (GA) regulates plant height, thus, a plant that fails to produce GA, or does not respond to GA will have a dwarfed phenotype (Figure 7). At the molecular level, the cell response to GA involves several processes, including the destruction of repressor proteins, referred to as DELLAs.

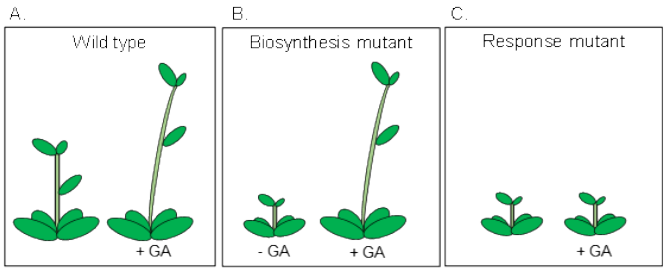


Figure 7. GA promotes plant growth. (A) When applied to wild type plants, GA induces height. (B) A mutation that blocks GA biosynthesis will result in short plants, and exogenous addition of GA to a biosynthetic mutant restores growth. (C) However, a mutation that affects GA signaling leads to failure to sense GA. Thus, exogenous addition of GA does not restore growth of a GA response mutant (C). (Image informed by Peng et al. 1999).

The "green revolution gene" ***Reduced height-1*** (***Rht-1***) from wheat encodes a DELLA repressor of GA signaling (Peng et al. 1999). A stretch of five amino acids referred to as the DELLA domain is required for the destruction of the repressor in response to GA. Mutations in the DELLA domain of ***Rht-1*** genes result in a protein product that is resistant to degradation, leading to failure to signal a GA response. The failure to signal the presence of GA in the ***rht-1*** mutants is responsible for their short stature. When the mutant ***rht-1*** gene is transformed into rice, the result is also short plants, which is proof that the new function of the DELLA protein is sufficient to induce dwarfism (Peng et al. 1999).

Generation of mutations

Because mutations are the ultimate source of genetic variation for breeding and genetic studies, it is sometimes of interest to generate new mutations. New DNA mutations (Figure 8) can arise spontaneously or they can be induced by experimental methods.

Spontaneous mutations arise “naturally”

That is, no particular effort was made on the part of the scientist to increase the mutation rate. Spontaneous mutations generally occur at very low rates, on the order of 1 per gene per million gametes.

Mistakes in DNA replication

Very rarely, incorrect bases are incorporated or bases omitted from a DNA strand during synthesis. Such mistakes can lead to spontaneous substitutions, insertions or deletions. For example, strand slippage due to the formation of a loop.

Environmental mutagens

There are several chemical agents present in the environment with potential to damage DNA and create mutations. These include naturally occurring compounds as well as man-made pollutants. Some modify DNA nucleotide bases through deamination (e.g., nitrous acid), while others promote oxidative reactions that may damage DNA (e.g., ozone).

Radiation is another type of environmental mutagen. Ionizing radiation (e.g., X-rays) can shatter DNA sequences and promote chromosome rearrangements. The less powerful ultraviolet rays can penetrate the cell and promote the formation of pyrimidine dimers which may inhibit replication and transcription.

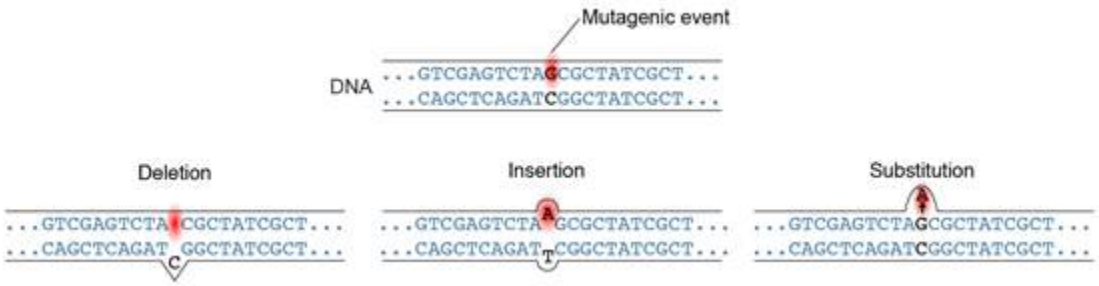


Figure 8. Mutations can arise spontaneously or they can be induced by experimental methods. Illustration by NIH-NHGRI.

Experimental mutagenesis involves applying a mutagenic agent to a plant or plant part to generate a mutation. For the mutation to be useful, it must be heritable, which imposes limitations on what tissues can be treated. Following the mutagenic treatment, a subsequent breeding scheme must be implemented to generate pure breeding lines of appropriate genotypes to display mutant phenotypes. These will depend on the species and the design of the mutagenesis.

Chemical mutagenesis

A commonly used chemical mutagen for experimental plant biology and mutation breeding is ethyl methanesulfonate (EMS). The compound is known as an alkylating agent and induces a CG to TA transition.

Seed Mutagenesis

This is the most common method of chemical mutagenesis in most species. Seeds are soaked in EMS solution, planted and allowed to flower and set seed. During EMS mutagenesis, every cell in the embryo will be independently mutagenized and the resultant seed and subsequent plant will be chimeric. These seeds can be sown and plants screened for traits of interest. For self-pollinating species, the required breeding is very simple, whereas for dioecious species it would be much more elaborate and labor-intensive.

Pollen Mutagenesis

For a few species such as maize, pollen mutagenesis is the method of choice. Pollen are treated with EMS and applied to silks of an acceptor plant. The benefit is that the resultant seed is not chimeric but rather all the cells of the embryo carry the same mutagenized paternal genome inherited from the pollen grain.

Insertional mutagenesis

The idea behind insertional mutagenesis is that a mobile or introduced piece of DNA can sometimes insert into a gene, thereby inactivating or modifying the function of the gene. The approach is useful in gene cloning because the inserted DNA serves as a "tag" that can be used to locate the gene of interest, and subsequently isolate that gene.

Another strategy used by researchers is mutagenesis using T-DNA from the ***Agrobacterium tumefaciens*** Ti plasmid. A larger number of plants (up to 10,000 for Arabidopsis) are transformed with T-DNA to generate mutants at random.

Similar to the example of transposon mutagenesis above, a gene can be "tagged" by a T-DNA and the tag used to isolate the gene. However, a T-DNA insertions are stable and thus disrupt the gene permanently.

In recent years, insertional mutagens have been engineered in many innovative ways. For example, strong promoters have been engineered into transposons such that if the transposon inserts near a gene, it can cause strong transcriptional expression of that gene. Such "activation tagging" approaches have been successful in identifying many genes that were not amenable to other mutational approaches.

Summary

Mutations alter A-T and G-C base pairs in DNA. A mutation in a coding sequence may alter the sequence and function of the protein product. A frameshift mutation changes the reading frame through insertions or deletions to produce an entirely novel product. A point mutation on the other hand alters only the amino acid represented by the codon in which the mutation exists. Mutations in regulatory regions might alter the expression of a gene. Mutations have been powerful tools for genetic studies as well as to study biological functions, for example, growth and development. The natural occurrence of mutations can be enhanced experimentally by applying agents referred to as mutagens. Researchers have used certain mutagens to induce mutations to study gene function and apply new traits for crop improvement.

Activity 1

For this activity, it may be helpful to refer to the genetic code chart.

For the following DNA sequence:

5′ ATG TTG GAG AAG GTT GAA ACT TTC 3′

Write the coding strand

3′ TAC AAG CTC TTC CAA CTT TGA AAG 5′

Write the mRNA sequence from the given DNA sequence

AUG UUC GAG AAG GUU GAA ACU UUG

Write the resulting peptide sequence

Met-Phe-Glu-Lys-Val-Glu-Thr-Leu

If the **third** triplet of the original DNA sequence is mutated to GA<u>A</u> to result in the following sequence

5′ ATG TTG GAA AAG GTT GAA ACT TTC 3′

What will be the mRNA sequence from the mutated DNA sequence?

AUG UUC GAA AAG GUU GAA ACU UUG

What will be the peptide sequence from the mutated DNA sequence?

Met-Phe-Glu-Lys-Val-Glu-Thr-Leu

What type of coding region mutation is this? Give a reason for your answer.

1. Silent
2. Amino acid substitution
3. Frameshift
4. Nonsense

If the **sixth** triplet of the original DNA sequence is mutated to <u>T</u>AA to result in the following sequence

5′ ATG TTG GAG AAG GTT TAA ACT TTC 3′

What will be the mRNA sequence from the mutated DNA sequence?

AUG UUC GAG AAG GUU UAA ACU UUG

What will be the peptide sequence from the mutated DNA sequence?

Met-Phe-Glu-Lys-Val

What type of coding region mutation is this? Give reason for your answer.

1. Silent
2. Amino acid substitution
3. Frameshift
4. Nonsense

The original DNA sequence is mutated to result in the following sequence

5′ ATG TTG GAG AAG GTT CGA AAC TTT C 3′

What will be the mRNA sequence from the mutated DNA sequence?

AUG UUC GAG AAG GUU CGA AAC UUU C

What will be the peptide sequence from the mutated DNA sequence?

Met-Phe-Glu-Lys-Val-Arg-Asn-Phe

What type of coding region mutation is this? Give a reason for your answer.

1. Silent
2. Amino acid substitution
3. Frameshift
4. Nonsense

Activity 2

1. A hypothetical plant gene encodes a protein with the following amino acid sequence:

Met-Leu-Lys-Thr-Phe-Val-Glu

A mutation of a single nucleotide alters the amino acid sequence to:

Met-Tyr-Asp-Pro-Gly-Ala-Lys-Ile-Arg

Describe the type of the mutation and the position of the codon that is affected.

2. For the following DNA sequence

3′ **A**TG GCC GGC AAT CAA CTA TAT TG**A** 5′

1 24 (nucleotide number)

a. Write the coding strand

b. Write the mRNA strand

c. Write the protein strand

d. Give the altered amino acid sequence of the protein for each of the following mutations:

i. A transition at nucleotide 11

ii. A transition at nucleotide 13

iii. A single nucleotide deletion at nucleotide 7

iv. A **T** to **A** change at nucleotide 15

v. An addition of ACC after nucleotide 6

vi. A transition at nucleotide 9

References

Doebley et al. 1997. The evolution of apical dominance in maize. Nature 386: 485-488.

Lertrat, K., Pulam, T. 2007. Breeding for increased sweetness in sweet corn. Int. J. Plant. Breed. 1: 27-30.

Nelson, O., Pan, D. 1995. Starch synthesis in maize endosperms. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46: 475-496.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintman, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D., Harberd, N.P. 1999. “Green revolution” genes encode mutant gibberellin response modulators. Nature 400: 256-261.

4. PCR and Gel Electrophoresis

WALTER SUZA; DONALD LEE; MARJORIE HANNEMAN; AND DEANA M. NAMUTH

Learning Objectives

At the completion of this PCR lesson, learners will be able to:

- List the 5 chemical components of a PCR reaction and describe their roles.
- List the functions of the 3 temperature cycles which are repeated during a PCR reaction.
- Describe the process of observing results and interpreting results of a PCR experiment.
- List possible uses of PCR in genetic testing and in research.

Overview

The polymerase chain reaction laboratory technique is used in a variety of applications to make copies of a specific DNA sequence. This lesson describes how a PCR reaction works, what it accomplishes, and its basic requirements for success. Examples of interpreting results are given. PCR's strengths, weaknesses, and applications to plant biotechnology are explained.

The Discovery of PCR

In 1983, Kary Mullis was driving along a Californian mountain road late one night. As a molecular biologist, Dr. Mullis was imagining a better way to study **DNA**. This late-night thinking led to a revolutionary way to make laboratory copies of DNA molecules (Saiki et al. 1985, Mullis 1990). In the decades since, the polymerase chain reaction or **PCR**, has become the standard method used for detecting specific DNA or RNA sequences. Selling the equipment and reagent kits for PCR is a multi-billion-dollar business because DNA and RNA detection is critical information in many applications.

In Vitro vs. *In Vivo* Replication

PCR is an *In Vitro* process; a series of chemical reactions that happen outside of a living cell. This laboratory technique is modeled after an In vivo process, the living cell's natural ability to replicate **DNA** during normal cell cycles (see Lesson on DNA: The Genetic Material). Every living cell makes a duplicate copy of each **chromosome** before the cell is ready to divide. Figure 1 below illustrates the key parts of *In vivo* DNA replication that are the basis for PCR success.

There are other enzymes that play an important role in in vivo replication. However, PCR works as an *in vitro* DNA replication process by using just one of these enzymes. Mullis imagined a chemical reagent and a temperature change

step in the method that could perform the work of the other two enzymes. It should be noted that because Dr. Kerry Mullis had learned about the details of *in vivo* DNA replication, he could create this science changing *in vitro* method.

Before reading the description of PCR components and processes, watching this video can help you visualize the importance of each step,

Name and Chemical Components of PCR

The name 'Polymerase Chain Reaction' represents the nature of the process. 'Polymerase' because DNA Polymerase III is required for constructing new DNA strands, just like in a living cell. 'Chain Reaction' describes repeating cycles of replication which target a specific segment of a chromosome and use a "copy the copies" progression each cycle that doubles the amount of **DNA copies of a specific segment of DNA** present each cycle. In just 20 cycles of the chain reaction, over one million (2^{20}) copies of that specific segment of DNA can be produced. This is enough DNA to see with your naked eye. The goal of PCR is to make millions of copies of a specific segment of DNA that all originate from a single DNA sample.

The five chemical components that must be added to a test tube for the PCR reaction to work, include a **DNA template, DNA polymerase III enzyme**, single stranded DNA primers, nucleotides, and reaction buffer.

1. The DNA template is a sample of DNA that contains the target sequence of DNA for copying.
2. DNA pol III. There are two requirements for a suitable DNA polymerase enzyme for PCR. First, the enzyme must have a good activity rate around 75°C. Second, the enzyme should be able to withstand temperatures of 95-100°C without denaturing and losing activity.
3. Two primers. Primers are short oligonucleotides of DNA, usually around 20 base pairs in length. Because the purpose of PCR is to amplify a specific section of DNA in the genome, such as a known gene, then primers of specific sequences must be used. The geneticist planning the PCR reaction will design a forward primer to bind to one strand and a reverse primer that complements and binds to the other strand. The **primer design** process to select forward and reverse primers is requiring appropriate genetics thinking and is describe later in this reading.
4. The four different deoxyribonucleotide triphosphates (dNTPs). Adenine (A), guanine (G), cytosine (C), and thymine (T) are needed to provide the building blocks for DNA replication. DNA polymerase will add each **complementary** base to the new growing DNA strand according to the original strand's sequence following normal A-T and C-G pairings.
5. Finally, a reaction buffer. This creates a stable pH and provides the Mg^{2+} cofactor needed for DNA pol III activity.

Three Temperature Cycles

A key insight to the success of PCR as an in vitro DNA replication process which generates millions of specific sequence copies was a three-temperature cycle which accomplish three parts of DNA replication: **denaturation** of the double stranded template, **annealing** of the primers to the single strands and **extension** of new strand synthesis by DNA pol III.

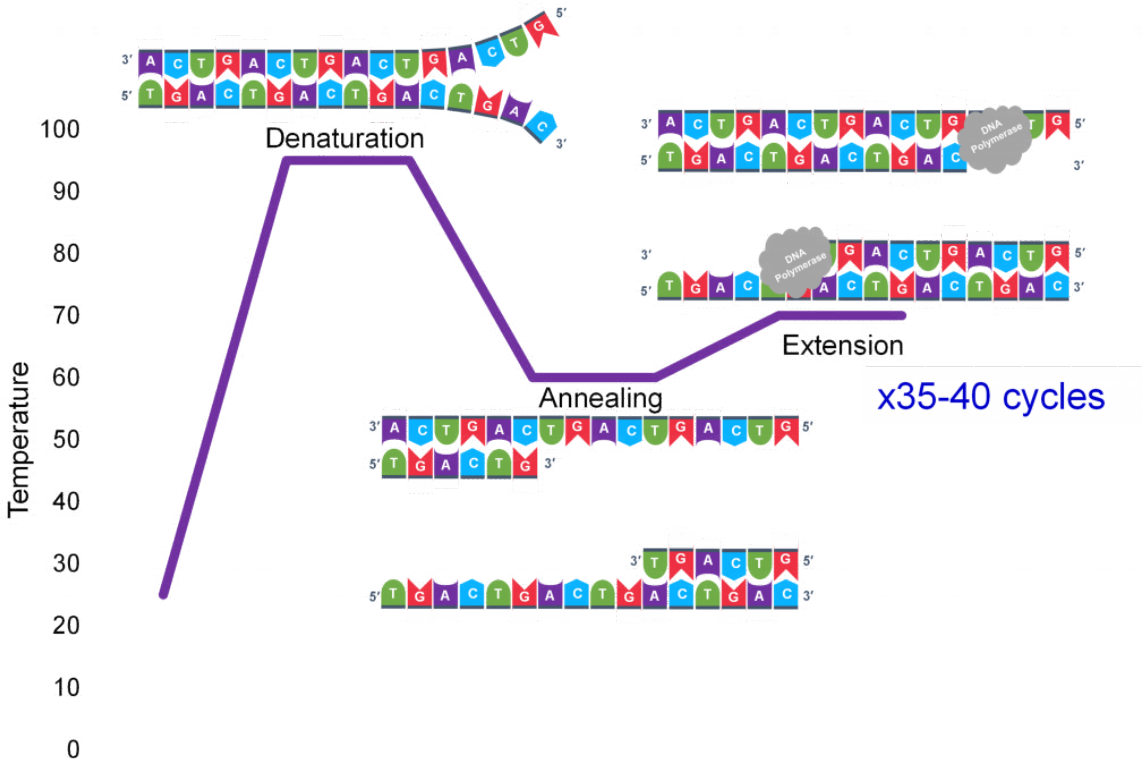


Figure 1. *The three temperature steps in a PCR cycle. Image by Marjorie Hanneman.*

In general, a single PCR run will undergo 25-35 cycles. The first step for a single cycle is the denaturation step, in which the double-stranded DNA template molecule (Figure 2) is made single-stranded(Figure 3). The temperature for this step is typically in the range of 95-100°C, near boiling. The high heat breaks the hydrogen bonds between the strands but does not break the sugar-phosphate bonds that hold the nucleotides of a single strand together (Figure 3).

Figure 2. *DNA template in the PCR reaction test tube is a double stranded molecule. The targeted sequence of nucleotides for this illustration is shown. The template DNA could be fragments from the chromosome of various lengths. The 5′ and 3′ orientation of both strands is shown. Image by Marjorie Hanneman.*

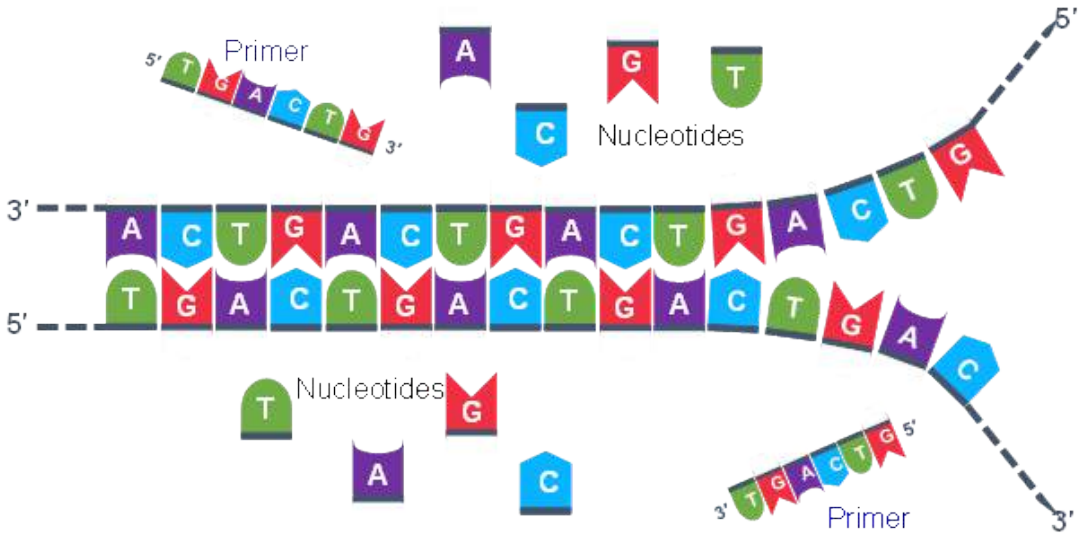


Figure 3. *Denaturation: The hydrogen bonds holding the strands of the double stranded template together are broken by heating the test tube. Image by Marjorie Hanneman and Donald Lee.*

Thousands of copies of the single stranded primers and the individual nucleotides were added to the test tube prior to beginning the cycles. Both the primers and nucleotides will become part of the new DNA strands. The second step in the PCR reaction is to cool the temperature in the test tube to 45-55°C. This is the primer annealing step in which the primers bind to **complementary** sequences in the single-stranded DNA template. The two primers are called the forward and the reverse primer and are designed because their sequences will target the desired segment of the DNA template for replication (Figure 4).

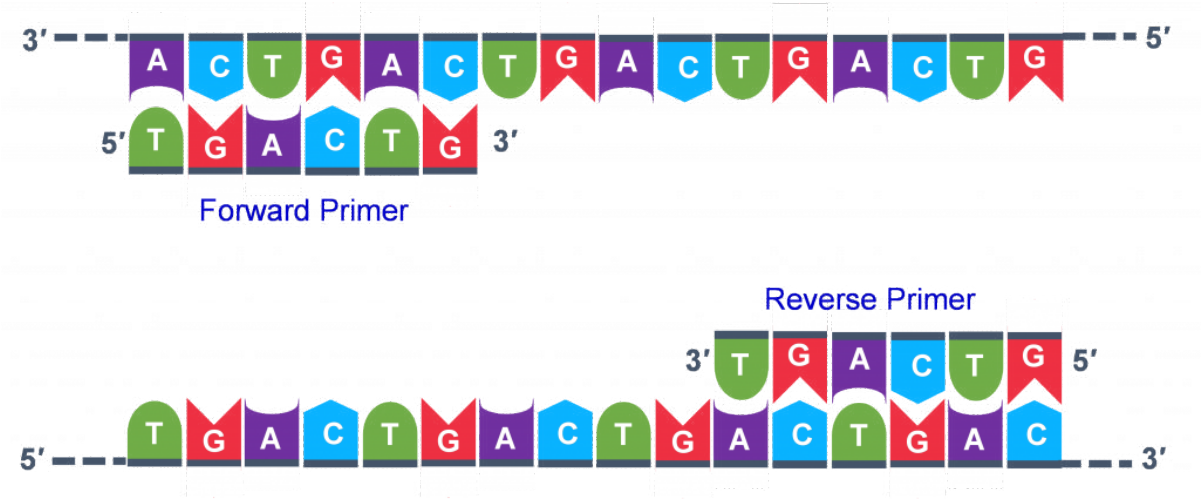


Figure 4. Annealing: *The second step of PCR is to cool the test tube which allows the forward and reverse primer to bind to the template. Image by Marjorie Hanneman and Donald Lee.*

The geneticist planning the PCR analysis must "design" the forward and reverse primers and then buy them from a vendor who can synthesize single stranded DNA that has a specific sequence and length. The two most important criteria for primer design are the following.

1. One primer must have a sequence that complements one of the template strands and the other primer must be complementary to the other strand. BOTH strands need to be primed for the replication process.
2. The primers must bind so that their 3′ ends are 'pointing' in the direction of the other primer. This ensures that the sequence between the primers is replicated in the PCR cycles.

Extension: The final PCR step is when the **DNA polymerase enzyme** reads the template and connects new nucleotidesto the primer's 3' end, extending a new complementary strand of DNA (Figure 5). Completion of the final step and the first cycle of PCR, will make two double stranded DNA copies from the original template DNA, doubling of the amount of DNA present. The test tube is heated to around 75°C, optimizing DNA pol. III activity and the newly synthesizing DNA strand is extended as the template strand is read by DNA pol. III. The Extension step will run for a few minutes and this step completes one PCR cycle.

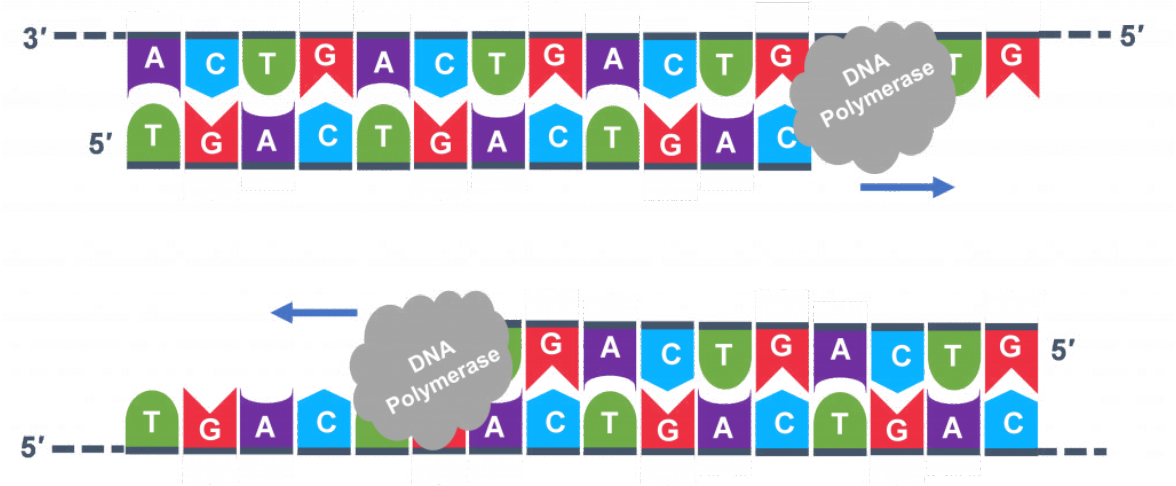


Figure 5. Extension: *This step is catalyzed by DNA pol III. The time allowed for this step determines how much of the template DNA pol III can read and replicate. Image by Marjorie Hanneman.*

For cycle 2, the denaturation, annealing, and extension steps are repeated (Figure 6 a, b, c). This time, though there will be twice as many DNA template molecules compared to what there was at the beginning of cycle 1. Copies are being made by reading the original template and copies are made by reading the copies made in the previous cycle. Therefore, if everything is working correctly, the DNA replication in the test tube is a chain reaction where at the end of a cycle, there is double the amount of that DNA sequence as what was found at the beginning of the cycle.

Figure 6a. Cycle 1 is complete*; two double stranded DNA molecules are made from the original double stranded template. Image by Marjorie Hanneman and Donald Lee.*

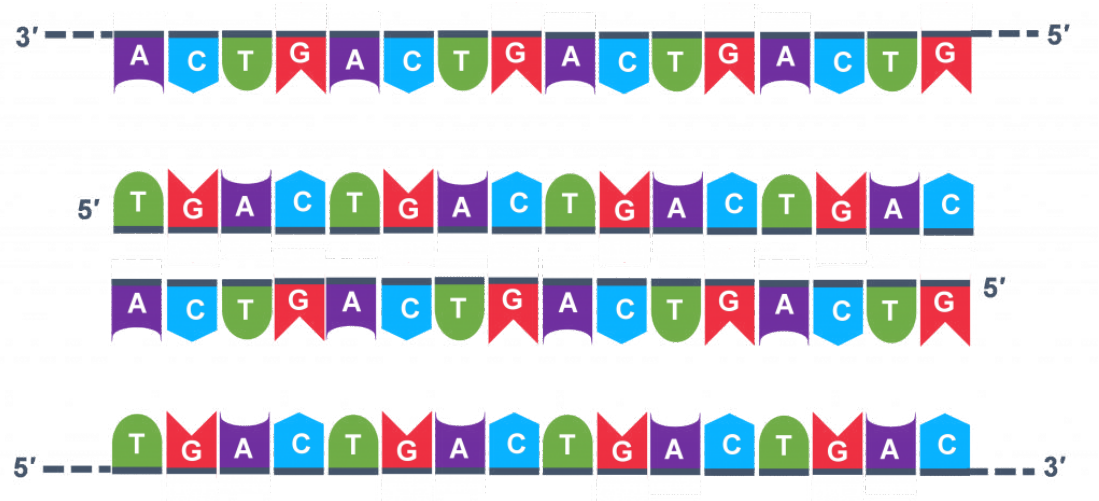


Figure 6b. Denaturation *step generates four single strand templates. Image by Marjorie Hanneman*

Figure 6c. Annealing and Extension*. All templates are primed and four 3′ ends provide a place of DNA pol III extension. Image by Marjorie Hanneman*

Because thousands of copies of the forward and reverse primer are added at the start of PCR, all the single strand templates, both the original, the copies in cycle 3 and beyond, and the copies of the copies made from previous cycles will be primed for the extension step of the cycles.

Thermal cycler

When PCR was first invented, scientists used water baths set at different temperatures and the hand transfer of test tubes at timed intervals to run the PCR reaction. Once the technique became a proven technology for DNA analysis, engineers went to work to create PCR machines. The instrument which heats and cools the DNA samples is called a **thermal cycler** (Figure 7). Each small tube or sample well in a plate contains all the chemical components needed for a PCR reaction. Adding a specific sample to the reaction mix provides the template DNA. A thermal cycler can be programmed for specific temperatures and the amount of time spent at each temperature. The engineered design of thermal cyclers to maximize the accurate replication of the targeted DNA in a small sample volume with the minimum amount of time can be critical in many applications of PCR.

Figure 7. The thermal cycler is used to carry out PCR reaction. Image by Walter Suza.

Taq DNA polymerase

When Dr. Kerry Mullis ran the first PCR experiments, he needed to add a new sample of DNA pol III after each denaturation step. This was because the high temperature needed to denature the double stranded DNA template also denatured the DNA pol III protein structure. The DNA pol III enzyme commonly available to molecular geneticists was from E. coli bacteria and this enzyme had no stability at near boiling temperatures. Fortunately, biologists had been investigating Thermus aquaticus, (Taq) a thermophilic eubacterium found in hot springs (Chien et al. 1976). The Taq version of DNA pol III does not easily denature in the hot temperatures required in PCR; plus, it has a good efficiency, able to add 60 base pairs/sec at 70°C. Like all other DNA polymerases, Taq DNA pol III cannot begin DNA replication without the addition of a starting primer. Thus, the discovery of Taq DNA pol III and the commercial availability of this enzyme made PCR a more reliable and doable technology which hastened its application to science investigation and diagnostic testing.

Visualizing the Results with Electrophoresis

Once a PCR reaction has been completed, we need to be able to see the results. To do this, a sample of the **PCR** mixture is loaded into an agarose gel for electrophoresis. The agarose gel contains a matrix of pores which enables it to separate DNA fragments based on their sizes. For details about setting up and running an electrophoresis gel, see Electrophoresis: How Scientist observe fragments of DNA

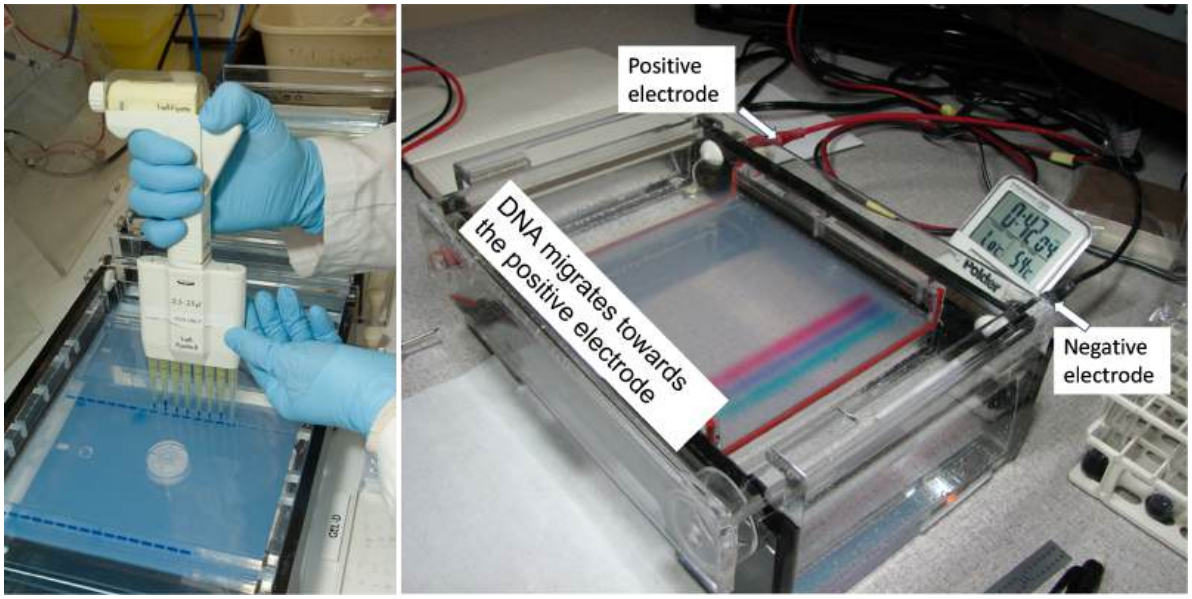


Figure 8. Electrophoresis: *A gel electrophoresis set-up with agarose gel with DNA and loading dye on the left and the power supply on the right. Image Source: Michael, CC BY 2.0, via Wikimedia Commons and U. S. Department of Agriculture, CC BY 2.0, via Wikimedia Commons.*

Figure 8 shows a picture of a gel electrophoresis gel that is running. The box on the right contains DNA loaded in the agarose gel. The gel placed in an aqueous solution of electrolytes. Depending on the type of dye used, color bands are a dye that was added to the PCR sample before it was loaded into the sample well. This allows for the tracking of the DNA's progression through the gel. Hooked up to this gel unit is an electrical power source which provides the force to move the DNA through the gel. Since DNA molecules are negatively charged, they will migrate towards the red, positive electrode. Shorter DNA fragments move faster than longer fragments through the pores in the gel.

After the gel is run, the DNA is stained with a chemical that binds specifically to DNA molecules and then will either reflect a specific color of visible light or fluoresce a specific color when viewed with ultraviolet light. A single 'band' contains 1000s of individual DNA fragments, all of the same length. Figure 10 illustrates the visual information that can be obtained from an electrophoresis gel after it has run. The electric current uniformly moves all the DNA fragments through a gel in the same direction. The sample wells at the top of the gel image thus establish lanes for the DNA samples to move.

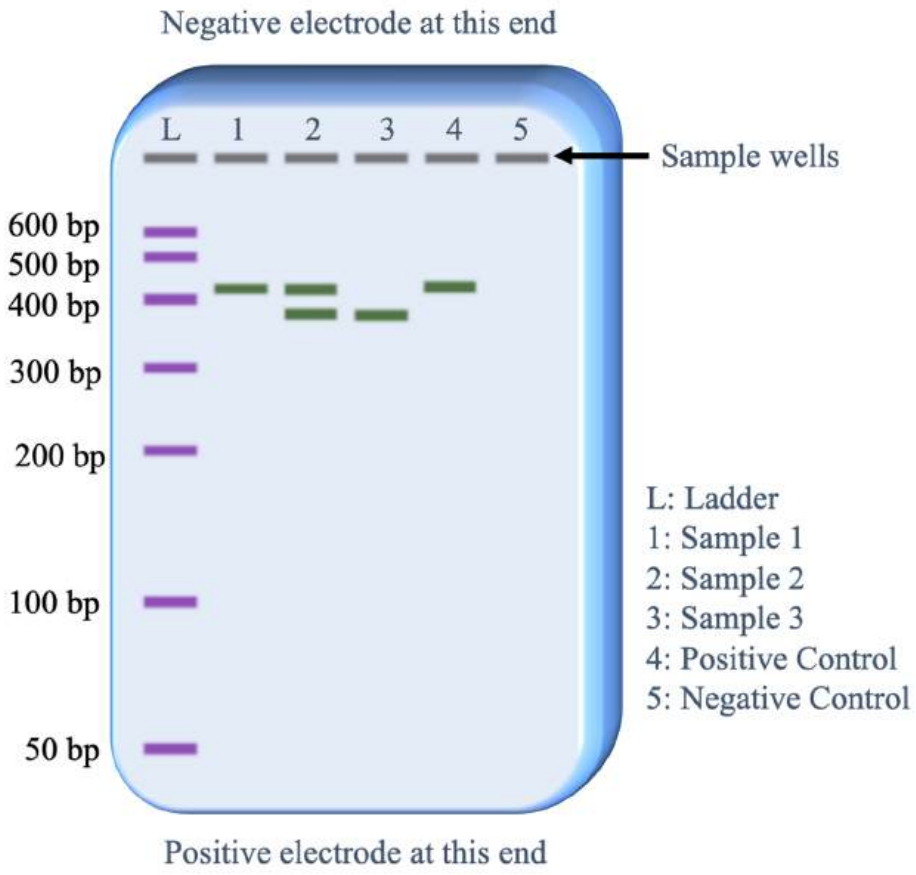


Figure 9. *Depiction of an electrophoresis gel with six sample wells that were loaded with either a DNA size ladder (lane L) or a sample from a PCR run (1-5.) The gel was subjected to a DNA staining dye. Image by Marjorie Hanneman.*

Below is a description of what information is revealed from each lane.

- Lane L: This was loaded with the DNA size ladder that contains copies of seven different lengths of DNA fragments. Commercial vendors of the DNA size ladder provide information on the lengths of each fragment in base pairs (bp). Running this lane provides an estimate of the DNA fragment lengths in the sample lanes (1-5).
- Lane 1: The PCR sample loaded in this lane has copies of a single length of DNA. The length is slightly more than 400 base pairs.
- Lane 2: The PCR sample loaded in this lane has copies of two lengths of DNA. One fragment is the same length as the fragments in lane 1 and the second fragment is slightly less than 400 bp.
- Lane 3: The PCR sample loaded in this lane has copies of one fragment that is the same length as the shorter fragments in lane 2.
- Lane 4: The positive control worked as predicted. The sample was set up with all the same reagents as the other PCR samples plus a DNA template that was known to contain the sequence targeted by the PCR primers that would generate a 410 base pair fragment.
- Lane 5: The negative control worked as predicted. The sample was set up with all the same reagents as the other PCR samples except no template DNA was added. Therefore, we would not expect the PCR reaction to work and the absence of a band of DNA fragments is as expected.

Because the positive and negative controls worked as expected, the biologist can be confident that the bands of DNA observed in lanes 1,2 and 3 reveal genetic information about the individual providing that DNA sample.

Advantages of PCR

PCR quickly became the method of choice for many types of DNA analysis because of several advantages over other DNA detection methods. The first; it is a simple procedure to set up and run. Fewer steps save time in getting a DNA analysis result. The second is the sensitivity. A very small amount of template DNA in the sample can be detected. Even just a few skin cells from one human hair contain enough DNA, making PCR useful in forensics. The third is that it can be designed to accurately differentiate genetic samples that are different by as little as single nucleotide in the targeted sequence. Finally, in application where large numbers of samples need to be subjected to the same PCR analysis, automation and robotic assistance allow the processing of many samples in a very short time. For example, an automated PCR would be critical for testing large numbers of people in hours during a pandemic.

Limitations of PCR

There are some drawbacks of using PCR that one should be aware of as well. First, the sequence of the gene or chromosome region being targeted is required. This limitation is rapidly diminishing as gene and genome sequencing technology and sharing of this sequence through Internet data bases has emerged as the norm in genetic analysis. Because of the size of the genomes of living things and a high conservation of gene sequences in many organisms, primers designed for a PCR test must be empirically tested with the proper controls. Biologists can easily generate false positive DNA from PCR that is caused by contamination or lack of specificity in primer design. There may also be a need to optimize concentrations of each chemical component. For example, changing the amount of DNA template, MgCl2 and Taq polymerase can affect both the quantity and quality of bands produced. Some studies have shown that even the brand of Taq polymerase can affect results (Holden et al. 2003). Likewise, the temperature cycles may need to be fine-tuned for a specific PCR test. Finally, as an in vitro DNA replication method, PCR cannot replicate entire chromosomes.

Summary

To briefly highlight the topics of this lesson, remember that **PCR** is a relatively easy lab technique which amplifies the amount of **DNA** present, much like living cells do in the beginning stages of a cell cycle. There are 5 chemical components of a PCR reaction: a **DNA template**, a **DNA polymerase**, primers, **nucleotides**, and a buffer. The 3 temperature steps for one cycle are the denaturation, primer annealing and extension steps. There are now many variations and uses of PCR ranging from forensics to genomic studies to identifying transgenic crops.

Lesson Activities

Watch these videos to learn more about PCR:

- Polymerase Chain Reaction (PCR) from DNA Learning Center
- Polymerase Chain Reaction (from UNL)

5. Gene Expression: Transcription

WALTER SUZA; DONALD LEE; PHILIP BECRAFT; AND MARJORIE HANNEMAN

Learning Objectives

1. Describe the roles that the promoter, coding region and untranslated regions of a gene play in gene expression.
2. Describe mRNA processing steps.
3. Draw the process of transcription and include the following in your drawing. DNA template and non-template strands, RNA polymerase, new RNA strand, and direction of RNA synthesis.
4. Draw the process of mRNA processing and include the following in your diagram, Gene (DNA), promoter, coding region, introns, exons, pre-mRNA, mature mRNA, poly A tail, cap.

Introduction

Genes are DNA sequences that control traits in an organism by coding for proteins (Figure 1). Organisms such as plants and animals have tens of thousands of genes. The impact that a single gene's information can have on an organism, however, is tremendous. Furthermore, organisms have all their genes in each of their cells, but they only need to use the information from a subset of these genes, depending on the type of cell and the cell's stage of development. Therefore, the key to gene function is controlling its expression.

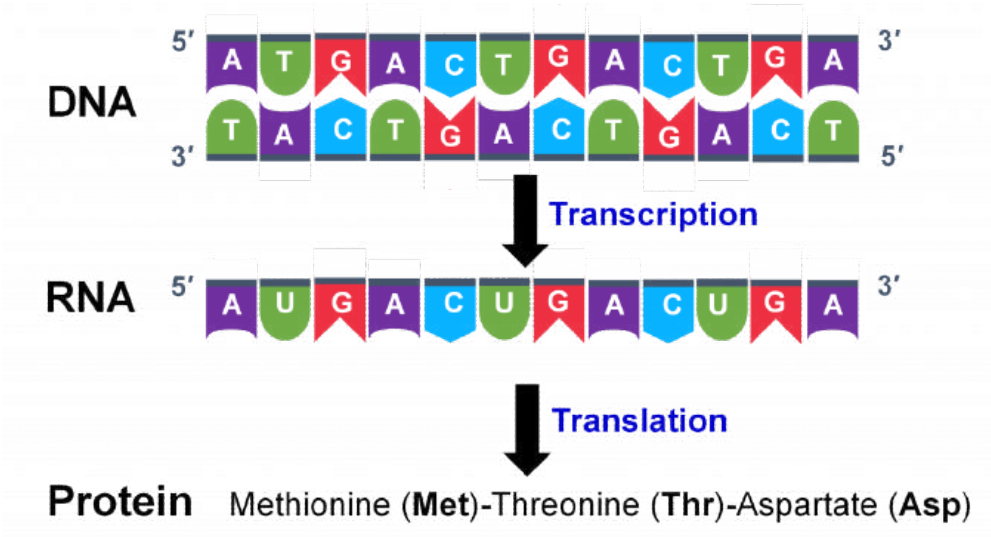


Figure 1. The central dogma of molecular genetics. Image by Marjorie Hanneman and Walter Suza.

Gene structure and transcription

The DNA sequence contains the information to control all biological functions, including the manifestation of traits important to agriculture (yield, drought tolerance, disease resistance, etc., etc.) How is the information contained in DNA sequences converted into the cellular activities necessary for plants and other organisms to function? DNA sequences are used to direct the synthesis of other molecules that actually perform these cellular functions.

Most typical genes encode proteins. The production of a protein from a gene involves several different processes (Figure 1). Transcription involves the copying of the DNA nucleotide sequence into an intermediate nucleotide molecule called RNA (ribonucleic acid). The primary RNA molecule is processed into a mature messenger RNA (mRNA) which then provides the information for the synthesis of a protein through the process of translation. Proteins are composed of amino acids connected by peptide bonds. The sequence of amino acids is determined by the sequence of nucleotide bases in the mRNA. The 20 amino acids have different chemical and physical properties and the sequence of amino acids determine the structure and function of the protein.

Gene structure

Only particular regions of chromosomal DNA are transcribed. A **gene** can be considered as the region of transcribed DNA, along with associated regions of DNA important for the regulation of transcription (Figure 1). A gene has several parts that are each important to the function of the gene.

The regulatory region (also known as the promoter) contains DNA sequence involved in the control of where and when the genes will be turned on to produce mRNA. The coding region is the part of the gene that is used as template to produce RNA molecules in a process called **transcription**. Some RNA molecules perform cellular functions directly while many others (messenger RNAs) are used to direct the synthesis of proteins in a process called **translation**.

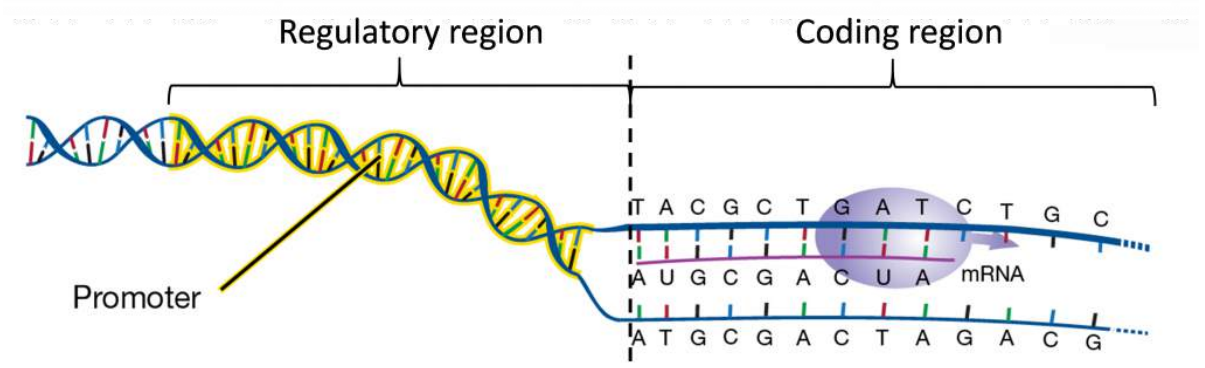


Figure 2. Every gene has a promoter and a coding region. Adapted from NIH-NHGRI.

a. Gene Promoter

The signals for starting and stopping transcription are located within DNA sequences. Specific nucleotide segments called promoters are recognized by RNA polymerase to start RNA synthesis. After the transcription of full-length RNA strand is completed, a second segment of DNA called terminator invokes termination of RNA synthesis and the detachment of RNA polymerases from the DNA template.

b. Protein-coding region

The protein-coding region of a gene is composed of the sequence of nucleotides that codes for **amino acids**. As described further in the section on **translation**, the coding region begins with an ATG **start codon** (AUG in RNA) and then ends with one of three **stop codons**. These sequences include only exons, but not all exonic sequences are protein coding as they may include **untranslated regions**.

Figure 3. The protein-coding region of a gene contains nucleotides that codes for amino acids. The 5′ and a 3′ UTR sequences do not code for amino acids but contain regulatory sequences that influence gene expression. In mRNA, translation start is AUG and translation stop can be UAA, UAG, and UGA. Image by Walter Suza.

c. Untranslated regions (UTRs)

Mature transcripts contain some sequences that do not code for amino acid sequences in proteins. These are referred to as untranslated regions or UTRs. Most mRNA transcripts contain a 5′ and a 3′ UTR. The 5′ UTR contains sequences toward the 5′ end of the mRNA sequence, before the **start codon**. These sequences can often be important for translational regulation, and sometimes other functions. The sequences following the **stop codon** are the 3′ UTR. The 3′ UTR may also have important functions regulating transcript stability or directing transcript localization within cells, or sometimes even transport (trafficking) between cells.

UTRs and introns are often useful in genetic studies. Protein coding regions are under strong selective pressure to produce functional proteins and so sequence variation is relatively rare. UTRs and introns on the other hand are under less stringent selection and are therefore sources of sequence variants that can be used to develop genetic markers.

Transcription

The genetic information of DNA is transferred to an intermediate molecule called RNA that is often translated to amino acid sequences used to build proteins. RNA is a nucleic acid, like DNA, but with some important differences. RNA contains a ribose sugar group instead of the deoxyribose found in DNA. RNA molecules are single stranded, instead of being double stranded. RNA contains a uridine (U) base and does not contain a thymidine base. The other bases (A, C, G) are contained in both RNA and DNA. U has the property of base-pairing with A.

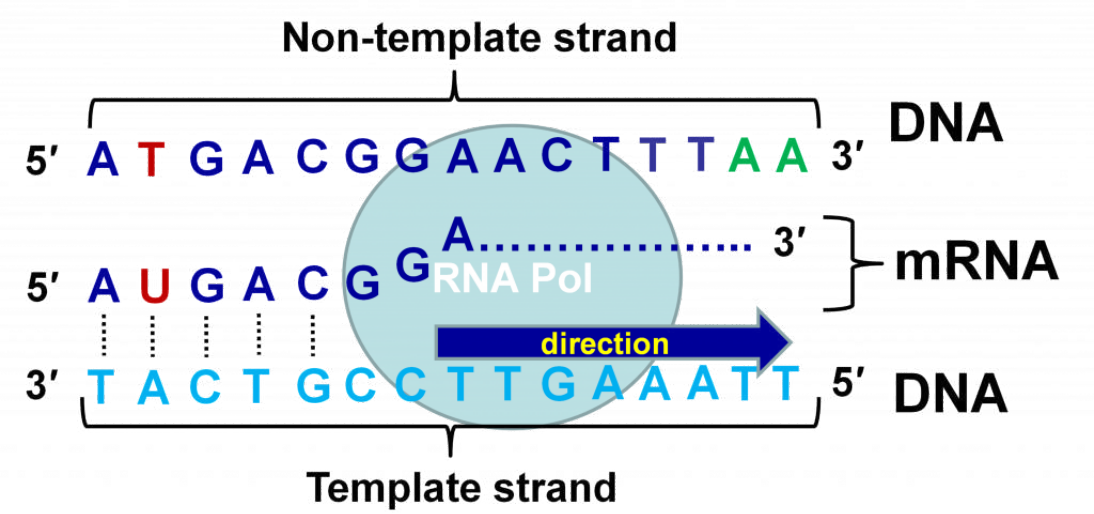


Figure 4. The enzyme RNA polymerase (RNA Pol) uses the template strand to synthesize RNA in the 5′ to 3′ prime direction. The base T is replaced with U in RNA. Image by Walter Suza.

RNA synthesis is directed by a DNA **template** in a process called **transcription**. A protein complex containing the enzyme **RNA polymerase** synthesizes an RNA molecule by adding nucleosides to the 3′ end of a growing chain. The principle of base pairing is used again and each nucleotide base added is complementary to the corresponding base on the DNA template. Thus, the RNA is complementary in sequence to the **template strand** of DNA, which is also referred to as "antisense" or "negative" strand (Figure 4). The RNA is identical in sequence (except U replaces T) to the other strand, which is called the "sense" or "positive" strand. Because RNA molecules are produced by the process of transcription, they are often referred to as **transcripts**.

RNA processing

Coding (transcribed) region

This is the region that is transcribed by RNA polymerase, also known as the RNA coding region. As described below, it may include **introns**, sequences that are removed from the mature RNA molecule during **RNA processing**. The transcribed region is demarcated by promoter and terminator sequences.

Introns and exons

As mentioned, and described in detail below, **introns** are sequences that are removed from transcripts during RNA processing. Sequences that are retained in mature transcripts are called **exons**. The corresponding stretches of DNA are typically referred to with the same terms. Introns are commonly found in genes of eukaryotes but are rare in prokaryotic organisms.

Intron splicing

The process of transcription produces pre-mRNA that contains both introns and exons. The process of splicing involves removal of introns from pre-mRNA and joining together the exons. A complex group of proteins that form a spliceosome perform the splicing reaction.

Introns sometimes serve as boundaries for sequences encoding functional protein domains, leading to possibility for new and variant proteins by exon shuffling. Also, introns can provide possibility for productions of variant RNA forms through alternate splicing allowing more than one gene product from a single gene. Some introns result from the insertion of transposable elements and may be spliced perfectly of imperfectly, offering more possibility for new genetic diversity.

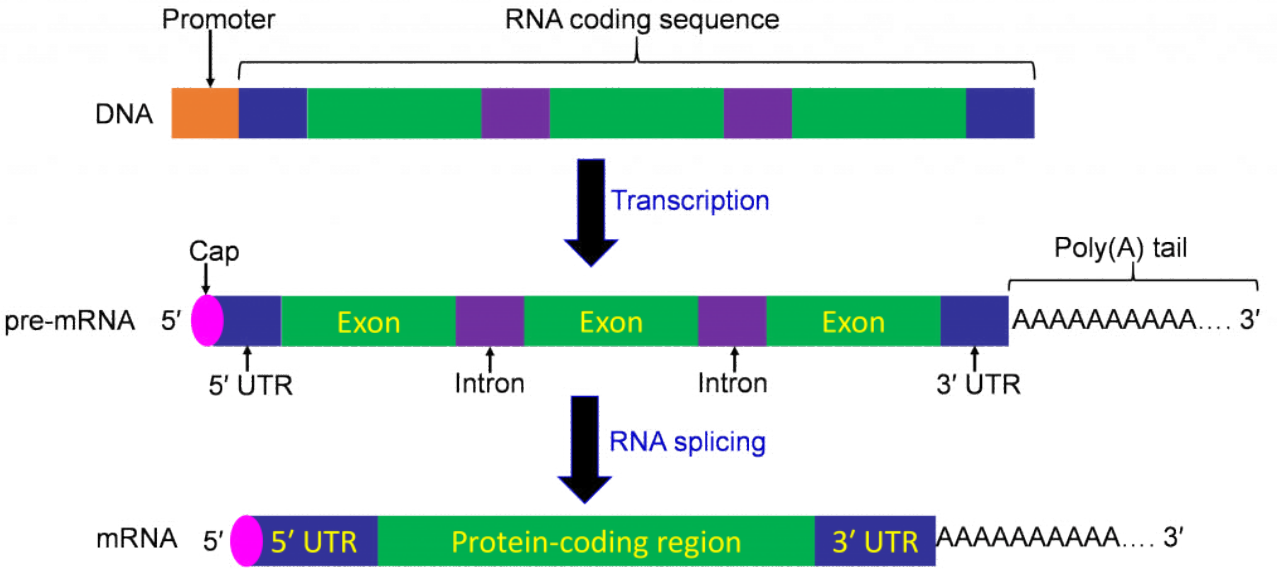


Figure 5. DNA (gene) transcription produces precursor-mRNA (pre-mRNA) that contains both introns and exons. The 5′ cap is 7-methyl guanidine. The enzyme poly(A) polymerase adds the poly(A) tail. The process of splicing involves removal of introns from pre-mRNA and joining together the exons to form mature mRNA. Image by Walter Suza

5′ Capping

The 5' capping is the addition of a 7-methyl guanidine to the first nucleotide of mRNA molecule, usually and adenine or guanidine. The phosphodiester linkage between 7-methyl guanidine and the target nucleotide is 5′-5′ instead of 5′-3′, and 3 phosphates rather than 1 are retained in the linkage. The cap stabilizes the 5′ end of the mRNA and plays a role in translation initiation.

Poly adenylation

The transcription of a gene may proceed beyond what ends up as 3′ end of mature mRNA. Thus the 3′ end of mRNA is formed after transcription. The enzyme poly(A) polymerase adds numerous adenosines to the 3′ end to result in what is called the poly(A) tail. The poly(A) tail is necessary for proper processing and transport of mRNA to the cytoplasm. The poly(A) tail is also important for the stability of mRNA, and initiation of translation in eukaryotic organisms.

Summary

The genetic information of DNA is transferred through transcription to an intermediate molecule called RNA. The signals for starting and stopping transcription are located within the DNA sequence and referred to as promoter and terminator sequences. The coding region of a gene is composed of a sequence of nucleotides that are transcribed into RNA. These sequences include exons and introns. Exons are the sequences that code for proteins. The coding region of a gene contains exons and introns. Also, pre-mRNA contains both introns and exons. The introns in pre-mRNA are removed through a process called intron splicing. The mRNA is processed by 5′ capping and addition of a poly(A) tail.

Learning Activities

Learning Activity 1

Given the following sequence of double-stranded DNA, predict the sequence of the RNA strand.

Learning Activity 2

Watch This Video: *Transcription Detail*

One or more interactive elements has been excluded from this version of the text. You can view them online here:
https://iastate.pressbooks.pub/genagbiotech/?p=24#video-24-1

6. Gene Expression: Translation

WALTER SUZA; DONALD LEE; PHILIP BECRAFT; AND MARJORIE HANNEMAN

Learning Objectives

1. Determine the amino acid sequence of a protein given the **nucleotide** sequence of a gene.
2. Recognize the biological connection among mRNA, tRNA, amino acids, and proteins.
3. Understand the concept of protein structure and its relation to protein function.

Introduction

The importance of gene expression is evident when you observe the changes plants go through during their lifecycle or during a season. Trees and bushes, for example have dormant buds through the winter. Environmental signals affiliated with the coming of spring induce genes in the buds to turn on and drive the dramatic changes of leaf development and flowering. The genes were always present in those bud cells but were controlled to turn on at the proper time. Understanding gene expression thus requires an examination of two processes, the activation of **gene expression** to make a "message" and the reading of this message to build a specific protein.

Amino acids are used to make proteins

Amino acids are the building blocks of proteins. Except for the amino acid proline, they all consist of a central carbon atom covalently bonded to an amino group, a hydrogen atom, a carboxyl group, and a variable R group (Figure 1). In the physiological range, both the carboxylic acid and amino groups are completely ionized allowing amino acids to act as either an acid or a base. Thus, amino acids do not assume a neutral form in an aqueous environment.

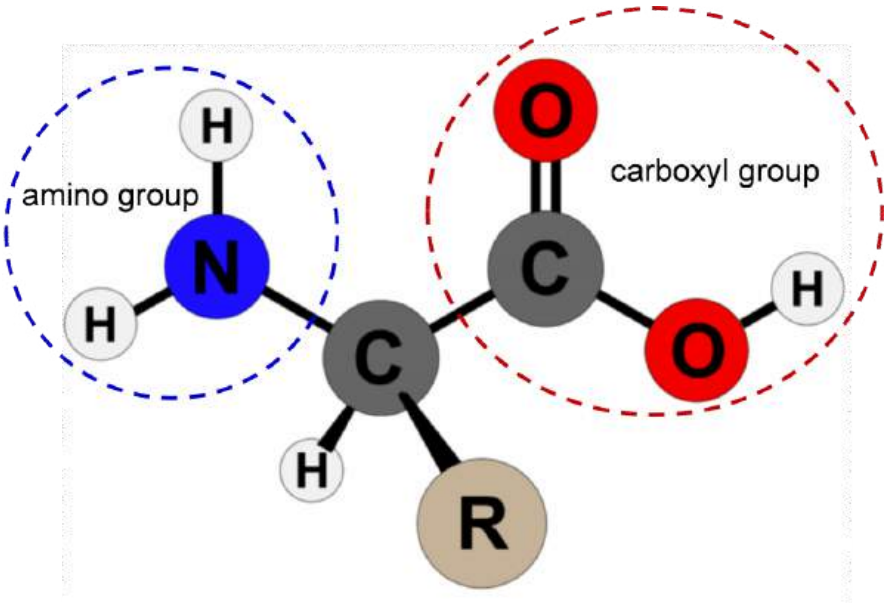


Figure 1. The structure of an amino acid. The amino (-NH2) and carboxyl (-COOH) are the same for all amino acids but the side chain (R group) is specific for each amino acid. Image Source: Techguy78, CC BY-SA 4.0, via Wikimedia Commons.

The chemical and physical properties of the side chains provide the functionally important properties to amino acids. Some of the important properties of side chains include acidic, basic, or neutral, hydrophobic vs. hydrophilic, as well as the size of side chains. They are categorized according to their side groups as nonpolar, polar, positively, or negatively charged or aromatic. All these properties and others affect how each amino acid contributes to the structure and function of a mature protein.

There are 20 different amino acids making up the subunits of **proteins** (Table 1). All proteins are made from differing combinations of the amino acids (Figure 2). Therefore, if you are going to make proteins you need to either make amino acids first (Figure 3) or consume amino acids in your diet. Animals consume some of their amino acids (the essential amino acids), but plants and bacteria are able make all their own amino acids.

Table 1. Amino acids and their abbreviations[1]

Full Name	Abbreviation (3 Letter)	Abbreviation (1 Letter)
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1. Table Source: Donald Lee, University of Nebraska-Lincoln

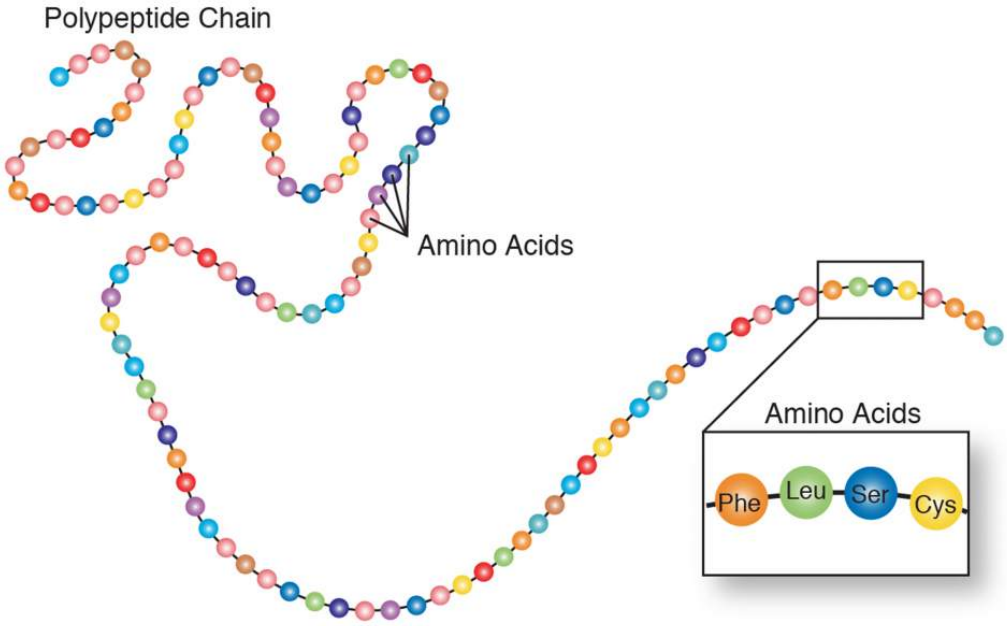


Figure 2. Proteins are polypeptides of different amino acids. Adapted from NIH-NHGRI.

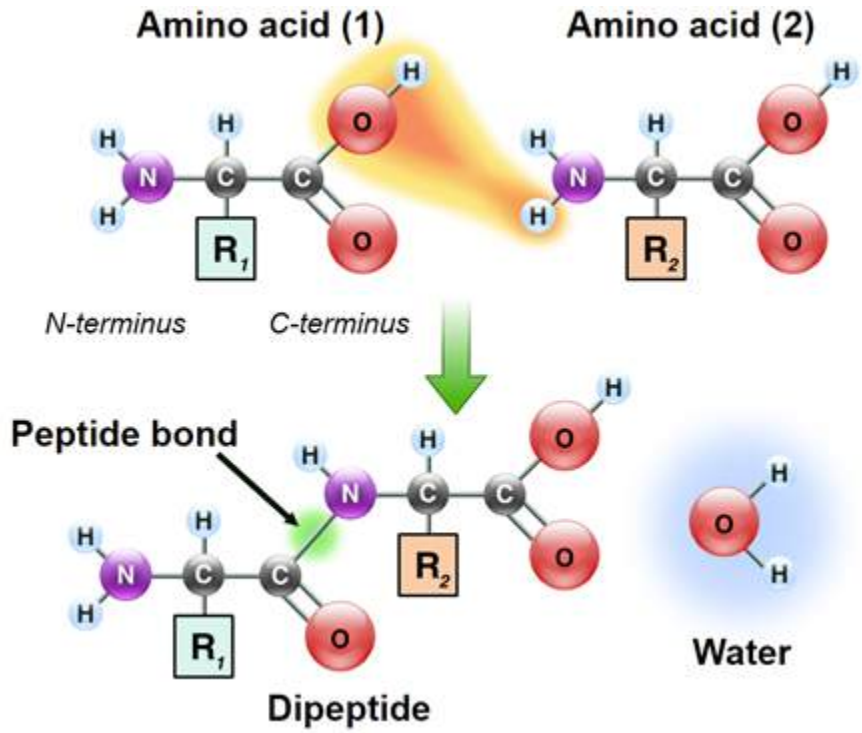


Figure 3. Proteins or polypeptides are made by connecting amino acids together with peptide bonds. The formation of peptide bond is a dehydration reaction that releases water with each new bond that is formed.

Codons and the Genetic Code

The search for the genetic code (Table 2) revealed that genetic information is stored in nucleotide triplets referred to as **codons**. The genetic code is degenerate because many amino acids are specified by more than one codon. Sixty one of the 64 possible combinations of the three bases in a codon are used to code for specific amino acids. Three **stop codons**, UAA, UAG, and UGA do not code for any amino acids but specify the termination of peptide chain synthesis during translation. The AUG **start codon** is used to initiate polypeptide synthesis and codes for the amino acid methionine.

Table 2. The genetic code. The 20 common amino acids are listed in their three and one-letter formats Stop codons are UAA, AUG and UGA.

First Position	Second Position				Third Position
	U	C	A	G	
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
					C
	Leu (L)		Stop	Stop	A
			Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
					C
			Gln (Q)		A
					G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
					C
			Lys (K)	Arg (R)	A
	Met (M)				G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
					C
			Glu (E)		A
					G

Why a Triplet Code?

Prior to understanding the details of **transcription** and **translation**, geneticists predicted that DNA could encode **amino**

acids only if a code of at least three nucleotides was used. The logic is that the **nucleotide** code must be able to specify the placement of 20 amino acids. Since there are only four nucleotides, a code of single nucleotides would only represent four amino acids, such that A, C, G and U could be translated to encode amino acids. A doublet code could code for 16 amino acids (4 x 4). A triplet code could make a genetic code for 64 different combinations (4 X 4 X 4) genetic code and provide plenty of information in the DNA molecule to specify the placement of all 20 amino acids. When experiments were performed to crack the genetic code, it was found to be a code that was triplet. These three letter codes of nucleotides (AUG, AAA, etc.) are called codons.

The genetic code only needed to be cracked once because it is universal (with some rare exceptions). That means all organisms use the same codons to specify the placement of each of the 20 amino acids in **protein** formation. A codon table can therefore be constructed and any **coding region** of nucleotides read to determine the amino acid sequence of the protein encoded. A look at the genetic code in the codon table below reveals that the code is redundant meaning many of the amino acids can be coded by four or six possible codons. The amino acid sequence of **proteins** from all types of organisms is usually determined by sequencing the gene that encodes the protein and then reading the genetic code from the DNA sequence

Ribosomes, tRNA and Anti-codons

The ribosomes are a large molecular complexes containing an assembly of several ribosomal RNAs (rRNA) and many proteins. Ribosomes are the machinery of protein synthesis, facilitating the ordered addition of amino acids to a nascent polypeptide chain under the guidance of an mRNA template (Figure 5A). Prior to the initiation of protein synthesis the ribosome occurs in two separate subunits of size 60s and 40s ("s" stands for Svedberg units and is a measure of a particles sedimentation rate during centrifugation).

The meaning of a codon for a specific amino acid is determined by the tRNA (transfer RNA). Each tRNA (Figure 5B) contains a triplet of nucleotides referred to as an anticodon, which is complementary to a specific codon. For each of the 20 amino acids, a specific enzyme (aminoacyl-tRNA synthetase) catalyzes its linkage to the 3′ end of its specific tRNA adapter. The function of aminoacyl-tRNA synthetase is critical as it provides a check for accuracy in protein synthesis by adding a specific amino acid to a specific tRNA molecule. In this way, one particular amino acid is targeted to each codon triplet of mRNAs.

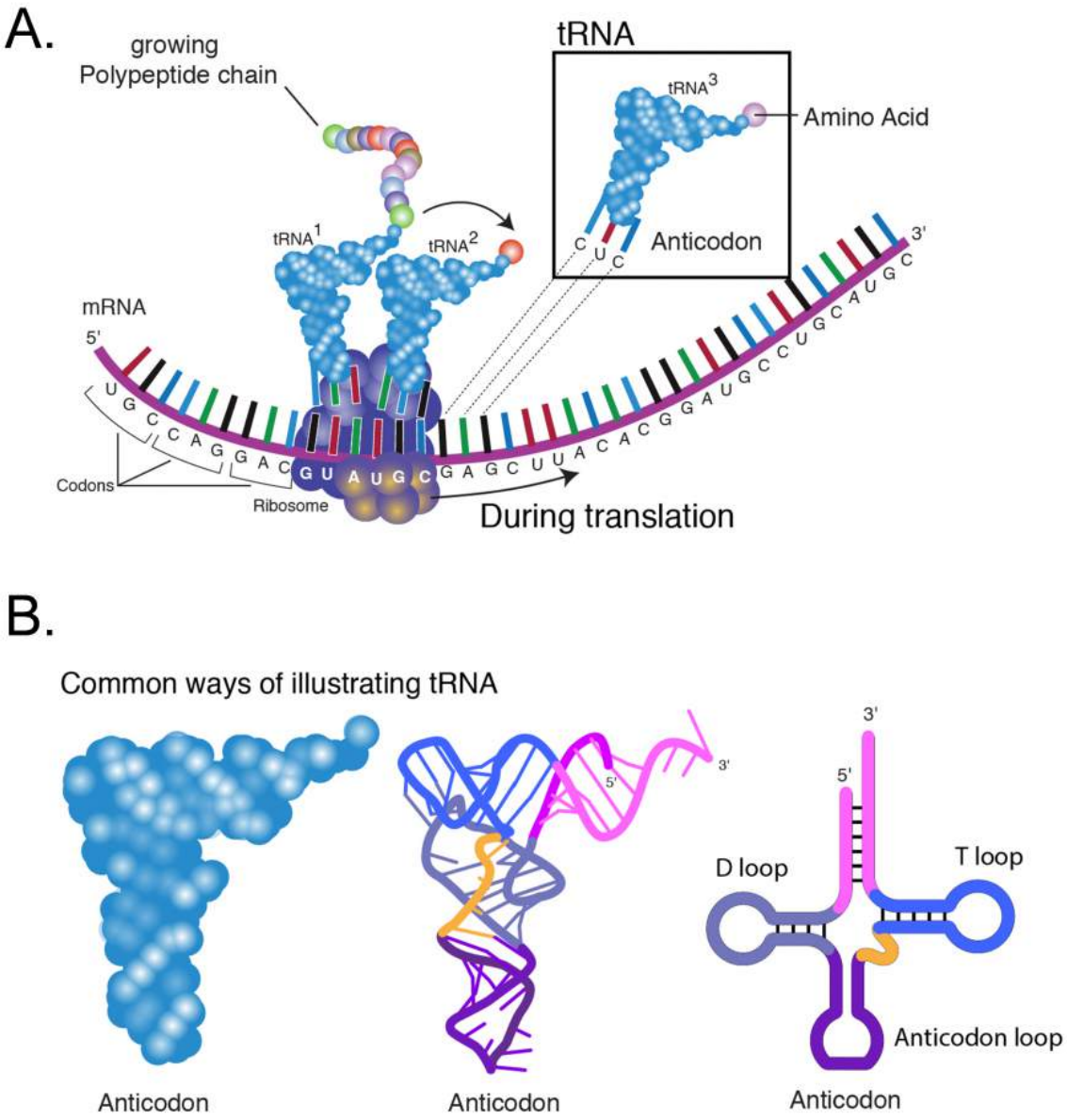


Figure 5: The key molecules needed for translation are the mRNA, ribosome, tRNA, and amino acids. Illustration by NIH-NHGRI.

Peptide chain formation

Protein synthesis is initiated by the 40s subunit through attachment it's to the mRNA, and recognition of the AUG codon. This is followed by the attachment of the 60s subunit to the 40s-mRNA complex to provide the structure necessary to align each successive aminoacyl-charged tRNA for transfer of its amino acid to the growing polypeptide. Successive amino acids are attached to the growing polypeptide by the formation of **peptide bonds** that form between the carboxyl group of one amino acid and the amino group of the next.

Like nucleic acids, polypeptides also have molecular directionality. One end of a polypeptide chain will have a free amino group and the opposite end will have a free carboxy group. As such, these ends are referred to as the amino-terminus or carboxy-terminus, respectively. During translation, mRNA templates are read from the 5' end toward the 3' end. Proteins are synthesized beginning at the amino terminal end going toward the carboxy terminus.

Protein structure and function

Proteins assume complex 3-dimensional structures that are essential for their various functions as enzymes, regulatory factors, or structural proteins. Overall structure is determined by complex physico-chemical interactions among amino acid side groups.

Several levels of structure are considered (Figure 6). The primary structure of a protein is the amino acid sequence of its polypeptide chains. The secondary structure is derived through the interactions between neighboring amino acids to form local structural elements such as beta sheets or alpha helices. The tertiary structure refers to the three-dimensional structure of an entire polypeptide and is determined by the diverse properties of amino acid side groups. For example, in an aqueous environment, hydrophobic amino acids will often interact at the core of a protein structure while hydrophilic groups may be exposed on the protein surface.

Many proteins are multimeric, composed of several polypeptide chains called "subunits", which associate non-covalently or in some cases via disulphide bonds. Such higher order spatial associations among polypeptides are also the result of interactions among amino acids and result in a quaternary structure. Some multimeric proteins consist of multiple copies of the same polypeptide.

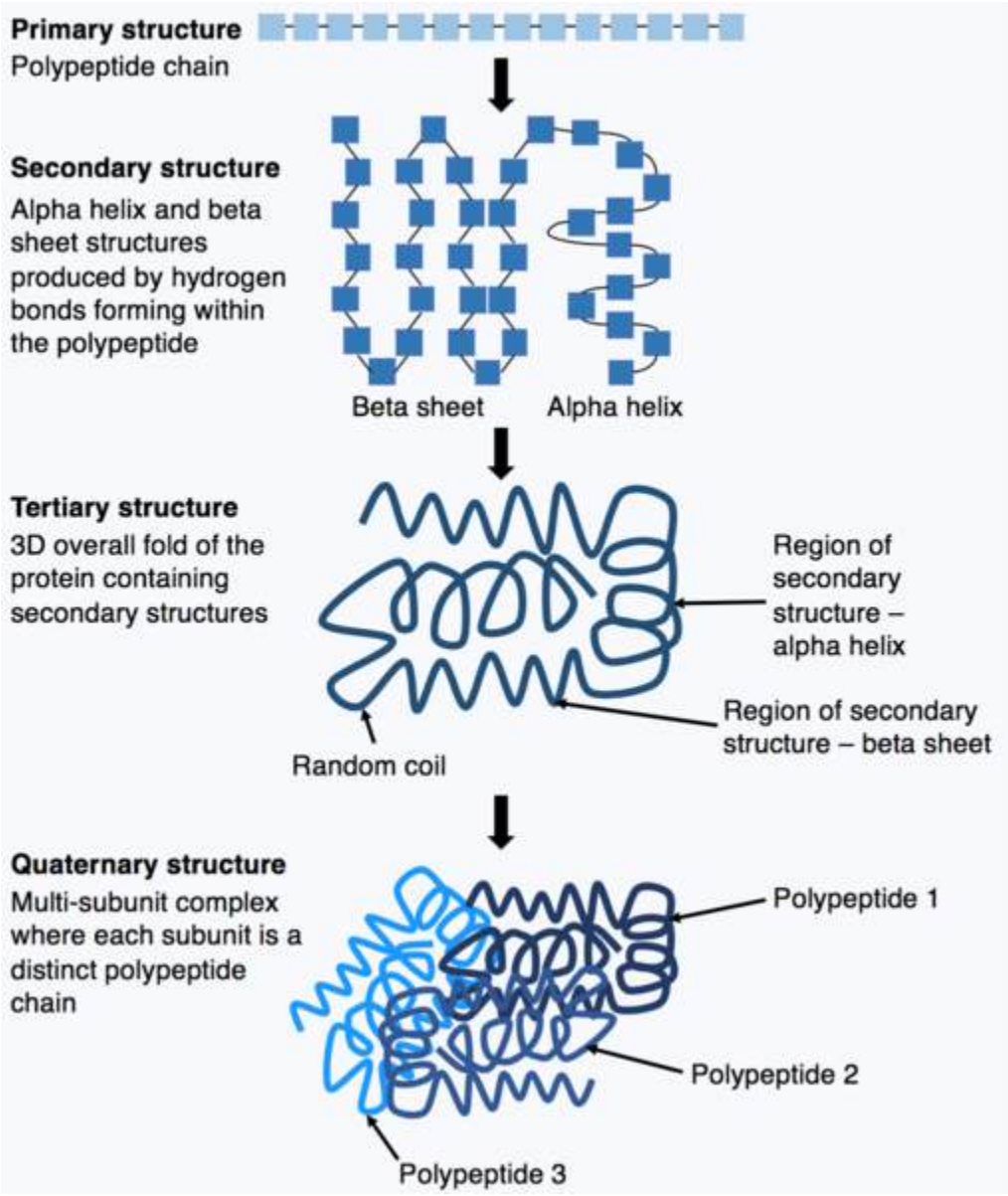


Figure 6. Levels of protein structure. Image Source: Kep17, CC BY-SA 4.0, via Wikimedia Commons.

Summary

The genetic information of DNA is transferred to through transcription to an intermediate molecule called RNA. The signals for starting and stopping transcription are located within the DNA sequence, and referred to as promoter and terminator sequences. The coding region of a gene is composed of a sequence of nucleotides that are transcribed into RNA. These sequences include exons and introns. Exons are the sequences that code for proteins. The coding region of a gene contains exons and introns. Also, pre-mRNA contains both introns and exons. The introns in pre-mRNA are removed through a process called intron splicing. The mRNA is processed by 5' capping and addition of a poly(A) tail. Mature RNA is then translated to amino acids used to build proteins.

7. Gene Expression: Applied Example (Part 1)

DONALD LEE; WALTER SUZA; MARJORIE HANNEMAN; AND PATRICIA HAIN

Learning Objectives

At the completion of this lesson you should be able to:

1. Define the roles of DNA and proteins in cell development and metabolism
2. Determine the amino acid sequence of a protein given the **nucleotide** sequence of a gene.
3. Describe the roles that the **promoter**, **coding region** and **termination sequence** of a gene play in gene expression.
4. Recognize the differences between the structure of proteins, **amino acids**, **genes**, and **nucleotides**.
5. Draw the process of gene expression and include the following in your drawing. Gene, **RNA polymerase**, promoter, coding region, termination sequence, cell, **nucleus**, cytoplasm, **mRNA**, tRNA, ribosome, anticodon, codon, amino acid, protein, peptide bond.

You have learned this fundamental biology fact: genes provide the information to instruct cells how to build specific proteins and these proteins control cell development and metabolism. Some biologists refer to this fact as the "Central Dogma". We prefer to refer to this process as gene expression. This lesson will integrate this biology fundamental with some gene expression examples. These examples emphasize an additional fundamental biology idea that life is a continuous cycle. In order for living cells to live, they require the genes, RNAs, and proteins that came from prior existing living cells.

Example #1: Gene Expression of the Plant Acetolactate Synthase Enzyme Gene

We have selected this gene because the protein the gene encodes is the target site for several types of herbicides and thus is an important example in agriculture. The gene also illustrates an important circle of life concept. All living things need amino acids to build proteins and some living things need proteins to build amino acids.

The ALS **gene** is not found in animals but is found in plants and bacteria. Plants and bacteria are able make all their own amino acids. Animals must consume a subset of the 20 amino acids (see Lesson on Translation) in their diet. One gene all plants and bacteria must have to live encodes the enzyme acetolactate synthase (ALS). ALS enzyme allows plants to make their own supply of three of the 20 amino acids; leucine, isoleucine, and valine. The ALS protein is an **enzyme** and is needed to catalyze a reaction in the chemical pathway that makes these three **amino acids**. Let us detail the information transfer steps needed for gene expression: use the genetic code of DNA sequences in the gene to encode the amino acid sequence of the ALS enzyme.

Three parts of the Gene and a Two-Step Process

The ALS gene consists of several thousand nucleotides of DNA and like all genes has three main parts, the **promoter**, **coding region**, and termination sequences (Figure 1, gene sequence abbreviated for demonstration).

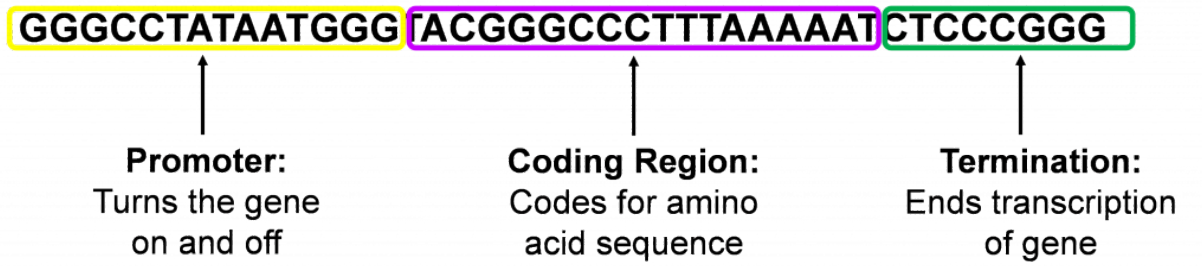


Figure 1. The parts of the ALS gene*; the promoter and termination sequences control transcription while the coding region encodes the protein. Image by Marjorie Hanneman.*

Each part of the ALS gene plays a role in controlling gene expression and this is a two-part process. The promoter is the on/off switch of the gene, the coding region determines the amino acids sequence of the **protein** that will get made and the termination sequence signals where the gene information ends. The DNA sequence that makes up a gene is a part of a much larger chromosome of continuous DNA. The gene sequence is organized to make an RNA (ribonucleic acid) copy of the gene's nucleotide sequence to serve as a mobile message for protein building. RNA building is called transcription because DNA nucleotides encode RNA nucleotides.

Step 1: Transcription: control from the promoter and termination sequence

The **promoter's** role is to selectively turn on the ALS gene in cells that will need the ALS enzyme. The promoter accomplishes this by its specific nucleotide sequences that react with regulatory molecules in the cell. We will discuss the regulation of gene expression in more detail later. What cells in a plant will need the ALS **enzyme**? All cells that are making proteins. This will be all cells in the plant at some stage of development. Therefore, the ALS gene has a promoter sequence that allows the gene to be turned on in all types of cells. When the ALS **gene** is turned on, transcription can take place. Like most chemistry in the cell, enzymes will be needed for **gene expression**. **RNA polymerase** is the *transcription* enzyme (Figure 2) which will bind to the DNA sequences in the ALS promoter and interact with one strand of the gene (the coding strand).

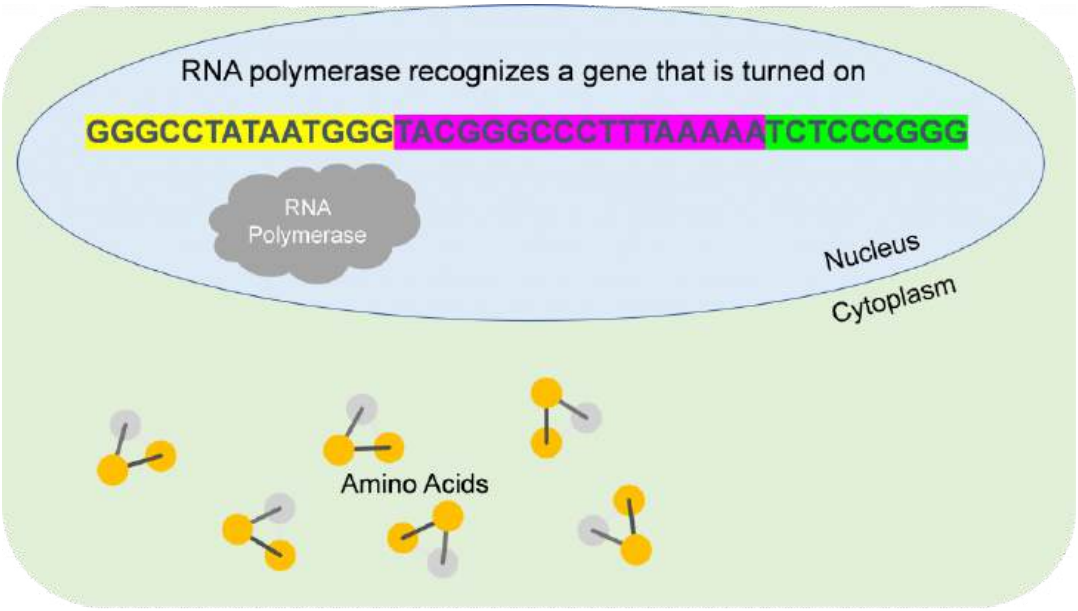


Figure 2. *RNA Polymerase binds to the promoter of the ALS gene to initiate transcription. Image by Marjorie Hanneman.*

RNA polymerase works by moving along the gene promoter until it encounters a series of A and T nucleotides called the TATA box (Figure 3). The TATA box signals RNA polymerase that it has reached the end of the promoter. Now the enzyme has the "green light" to read the coding strand **DNA** and make RNA. The procedure works much like DNA replication. RNA polymerase will read the DNA in a 3' to 5' direction and build the RNA 5' to 3'. Unlike DNA replication though, only one DNA strand is read and the transcription process must end once the RNA polymerase reaches the end of the ALS gene. If RNA polymerase kept going along that strand of the DNA making up the chromosome, other genes that are on the same **chromosome** might be expressed in the wrong cells or at the wrong times.

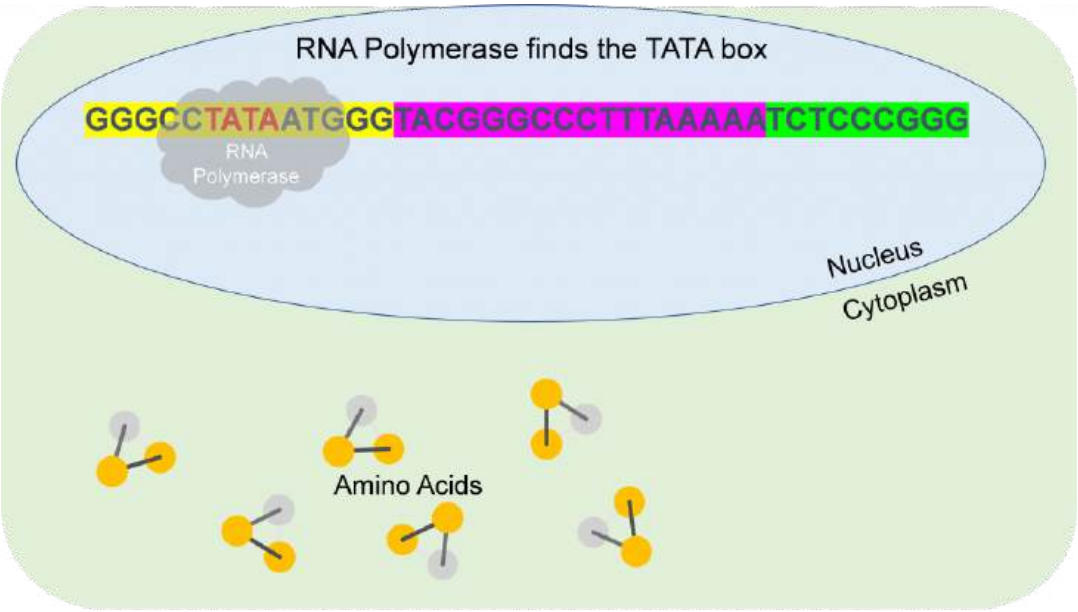


Figure 3. *RNA polymerase encounters the TATA box at the end of the promoter. This signals where to begin the reading of the coding region. Image by Marjorie Hanneman.*

How is the RNA polymerase signaled that it has reached the end of the gene? That is the role of a **termination sequence** in the gene (Figure 4). The termination sequences signal the end of the gene. One transcription termination strategy will be described here. The **nucleotides** that make up the termination sequence of the ALS gene could be ordered to form a palindrome. This means that the nucleotides being placed in the RNA can fold back on themselves and form a hairpin loop (Figure 5).

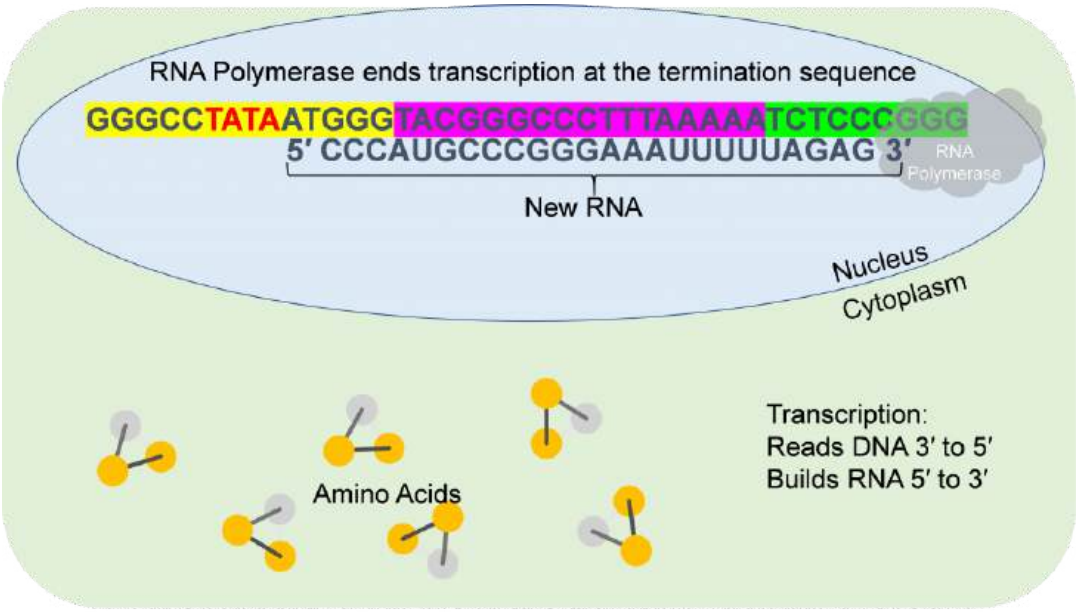


Figure 4. *When RNA polymerase reads the termination sequence it is signaled to quit reading the gene's coding strand. Image by Marjorie Hanneman.*

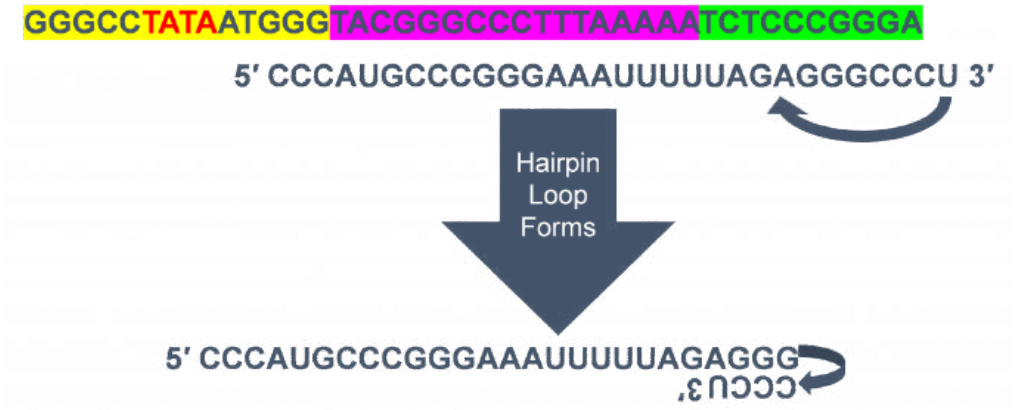


Figure 5. *The termination sequence is palindromic. The RNA being made has a sequence that can form a hairpin loop. The formation of the hairpin in the newly made RNA disrupts the RNA polymerase, halting transcription. Image by Marjorie Hanneman.*

While no geneticist has actually viewed this hairpin loop formation in action, it is believed that the hairpin formation snaps the RNA polymerase off the DNA coding strand and releases the RNA message. Transcription of an RNA that has the **coding region** information is now completed. Each time an RNA polymerase goes through the process, one copy of the RNA is made by reading the DNA template. The type of RNA made from transcription of the ALS gene is called a messenger RNA or mRNA for short. Some **genes** can encode transfer RNA (tRNA) or ribosomal RNA (rRNA). These RNA molecules have special roles in the next part of gene expression called translation. Once an **mRNA** is transcribed, some

modifications will occur in the nucleus. We will describe some of the post-transcriptional processes later. For now, let's follow the ALS mRNA onto the **translation** step of gene expression.

Video Examples

Watch this video from UNL for more information about the Transcription process

Step 2: Translation: reading the code to make proteins

The **translation** process requires a conversion of information from a chain of **nucleotides** in RNA to a chain of **amino acids** in **protein**. The translation term was coined based on the idea that we are converting to a new molecular language at this stage of **gene** expression. The conversion of information in translation is a little more complex than **transcription** and requires several biomolecules to interact and work together (Figure 7). We will describe the ribosomes and tRNAs molecules first and then describe how they work together to translate a **mRNA**.

- **tRNA:** The 't' prefix means transfer. The tRNA is a small RNA (70 to 80 nucleotides) and has a sequence that allows hairpins to form which gives a secondary structure to the molecule. Their small size makes them mobile to provide the cells means to transfer amino acids to the site of protein building. Their structure also makes them specific in how they deliver their cargo of a single, specific amino acid. A tRNA is about 50 times larger than an amino acid.

- **Ribosomes:** Ribosomes are macromolecules that are made of both RNAs (rRNAs) and proteins. These components assemble into a workbench for the translation process to occur. The ribosome is a big macromolecule: it can hold three tRNAs in position for the translation work and can catalyze bond breaking and bond formation.

We will use a simple representation of the mRNA, tRNA, ribosome, and amino acid to show how they work together (Figure 6).

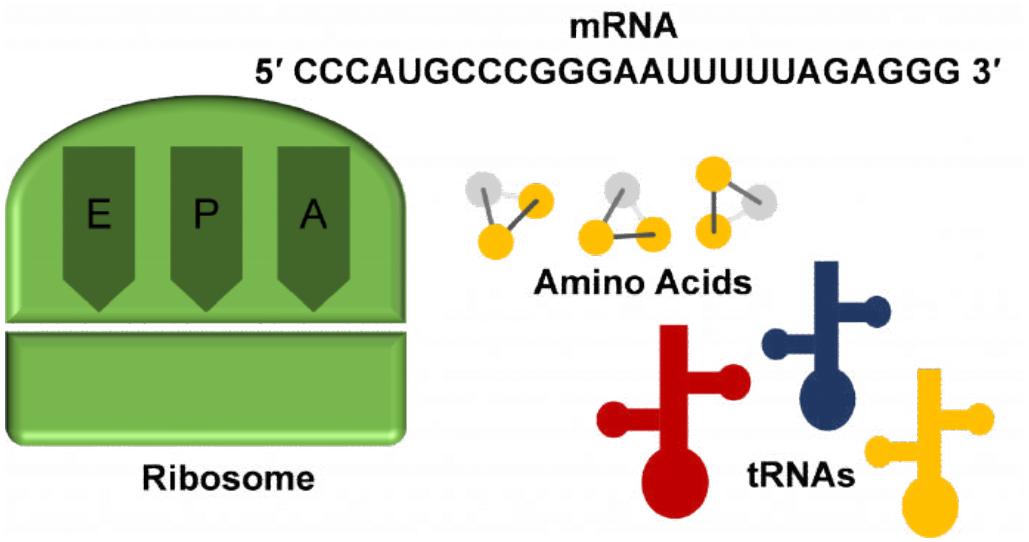


Figure 6. *The key molecules needed for translation are the mRNA, ribosome, tRNA, and amino acids. Image by Marjorie Hanneman.*

The tRNAs shown in Figure 6 are not ready for translation. The structure of the tRNA is recognized by special enzymes in the cell that attach the proper amino acid to the tRNAs. The tRNAs that are bound to their designated amino acid are called charged tRNAs (Figure 7). The near the middle of the 70-80 nucleotides in a tRNA are three nucleotides called the anticodon. When the tRNA folds, the anticodon becomes one end of the molecule and the other end will be attached to a specific amino acid. In the figures for this lesson, the tRNA is shown with its secondary structure and only the anticodon sequence is shown (Figure 7).

Let's describe how the key players get protein building started, how they keep it going, and how they end it. Active ribosomes are assembled from small and large subunits when the ribosome "workbench" is ready to translate a mRNA (Figure 7). When mRNAs are present in the cytoplasm, the small and large ribosome subunits assemble at the 5' end of these messages and are ready to read mRNAs like blueprints to build specific proteins.

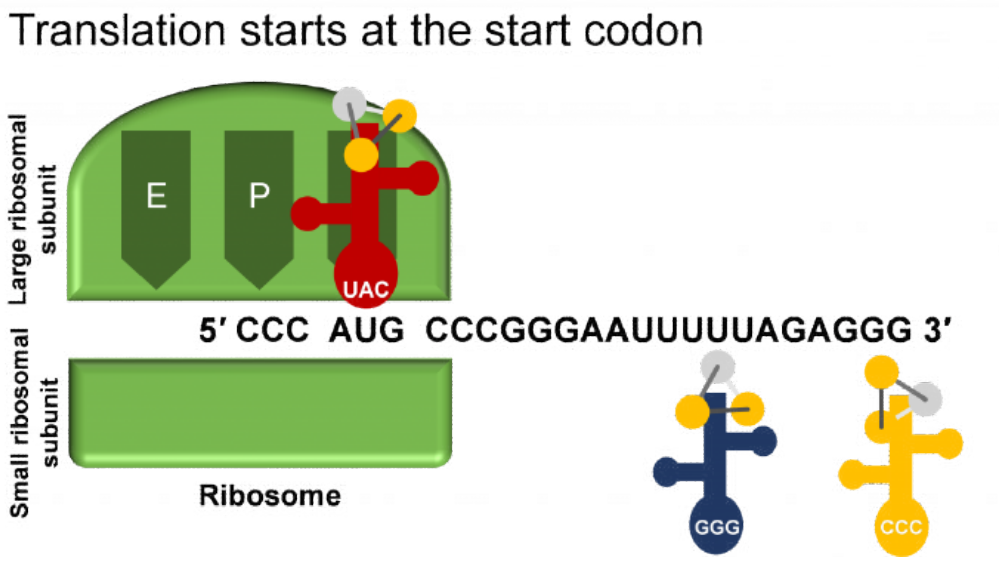


Figure 7*. The ribosome translates the mRNA 5' to 3' and starts by finding an AUG sequence. The ribosome allows a tRNA with the UAC anticodon to bind and bring in an amino acid. The tRNA is bound to the A site of the ribosome. Image by Marjorie Hanneman.*

The ribosome reads the mRNA quickly and accurately, but it must determine where to start. As the ribosome moves in the 5' to 3' direction along the mRNA it "looks" for the sequence AUG (Figure 7). This sequence is called the start codon and signals the ribosome that this is where to begin building the protein sequence.

Video Examples

Watch this UNL video on Translation for more information about these processes.

The Reading Frame, Codons and Anticodons

The mRNAs, tRNAs, ribosomes, and **amino acids** are now in position to combine their specific functions into the building of a protein. The AUG start codon establishes the beginning of what is called a reading frame on a mRNA. The nucleotides in the mRNA must have the specificity to code for 20 different amino acids to build the protein. The genetic

code is read as a three-letter code which provides this specificity. The ribosome will move along and read the sequence in the mRNA one codon (three nucleotides) at a time to follow the reading frame and build the correct protein (Figure 8).

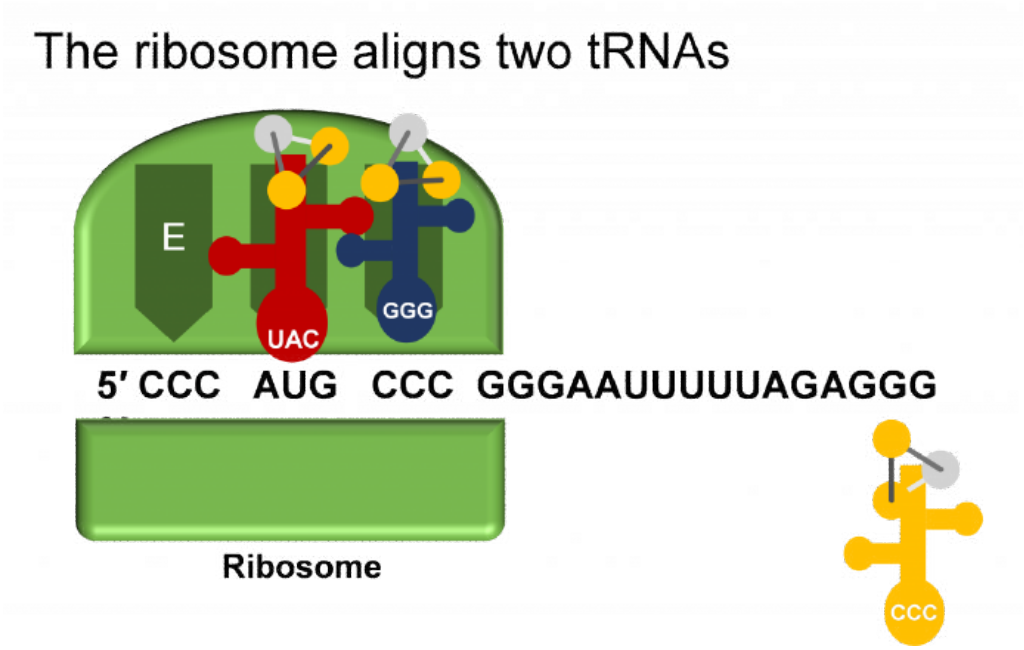


Figure 8. *The ribosome allows two tRNAs with complementing anticodons to enter the 'A' and 'P' sites and bring in their amino acids. Image by Marjorie Hanneman.*

Getting Translation Started with the Start Codon

The AUG start codon signals the ribosome to place in the amino acid methionine. The A, P and E sites on the ribosome site works somewhat like a vice on the workbench, holding stuff in place so it can be worked on. In Figure 9, the ribosome has moved in the 5' to 3' direction on the mRNA .mRNA. This allowed a tRNA with the anticodon GGG to move into the A site and placed two amino acids very close to each other on the ribosome. The ribosome will now move one more codon in the 5' to 3' direction which moves the two tRNAs now to the E and P sites (Figure 9). Now the ribosome is ready to start binding together amino acids with peptide bonds to start to form a protein. A third tRNA (with the CCC anticodon) can now arrive at the open A site, bringing the amino acid designated by the GGG codon.

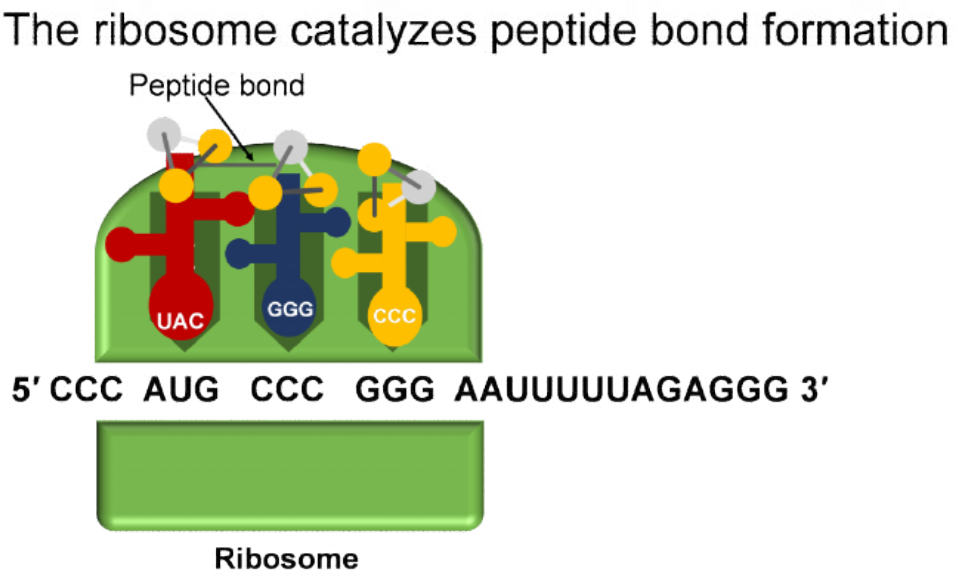


Figure 9. *The ribosome can link two amino acids together with peptide bonds because the tRNAs holding them are next to each other at the 'A' and 'P' sites of the ribosome. Image by Marjorie Hanneman.*

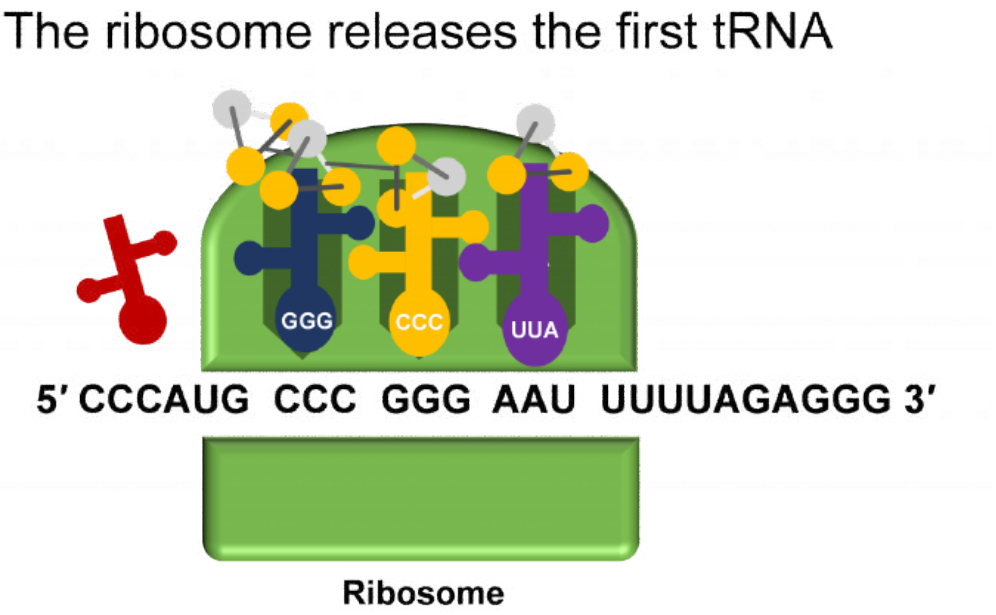


Figure 10: *The ribosome moves over to the 'GGG' codon, releases the tRNA with the 'UAC' anticodon and is ready for the next tRNA. Image by Marjorie Hanneman.*

Peptide Bond Formation and Protein Building

As you can see, a ribosome "workbench" has catalytic tools to go along with its 'P' and 'A' "vice sites". The ribosome has enzymatic functions that allow it to break and form bonds. The ribosome will break the bond that binds the amino acid to the tRNA at the 'E' site. Simultaneously the ribosome forms a peptide bond between the second and the third **amino acids** that were brought in by the tRNAs. The close proximity of these amino acids at the ribosome sites provides the opportunity for peptide bond formation and then release of the tRNA in a smooth continuous process (Figure 10 and 11).

Peptide bonds always bind the acid end of one amino acid with the amino end of the next amino acid. Peptide bonds are strong covalent bonds which keep the amino acids connected during and after **translation**. The tRNA at the 'E' site has now accomplished its task by bringing in the first amino acid. This tRNA will then move off the ribosome (Figure 11). The tRNA released from a ribosome is recycled by the cell. It can be charged again by binding another amino acid in the cytoplasm and contribute to the synthesis of another protein. The protein. The first three amino acids in the **protein** have now been put together.

The synthesis of our protein is far from over. The ribosome continues to shift one codon in the 5' to 3' direction as a tRNA leaves the E site .site. A tRNA can now come into the A site if it has the anticodon that is complementary to the next codon on the **mRNA**. **Proteins** are hundreds of **amino acids** long. In our example, the steps described continue until the complete mRNA information is read to build the 667 amino acid long ALS enzyme.

The End of Translation: stop codons looking for something they cannot find

Start codons start **translation** so it is logical that stop codons stop translation (Figure 11) How do stop codons do this? There are 61 tRNAs with different anticodons (see Appendix 1: Codon Table). That means there are three codons that do not have corresponding tRNAs with complementary anticodons. These three codons serve as stop codons. When a ribosome encounters a stop codon on a **mRNA** it will wait for a tRNA with the right anticodon to come over. It will not skip the codon or shift over one **nucleotide** to form a new reading frame. The ribosome waits for the right tRNA, but it does not wait for long. A stalled ribosome will quickly cleave off the bound tRNA with the newly built **protein** chain and then move on to translate another mRNA. No tRNAs in the cell have anticodons that complement any of the three possible stop codons. Therefore, stop codons are able to end the translation process when the completed protein is made.

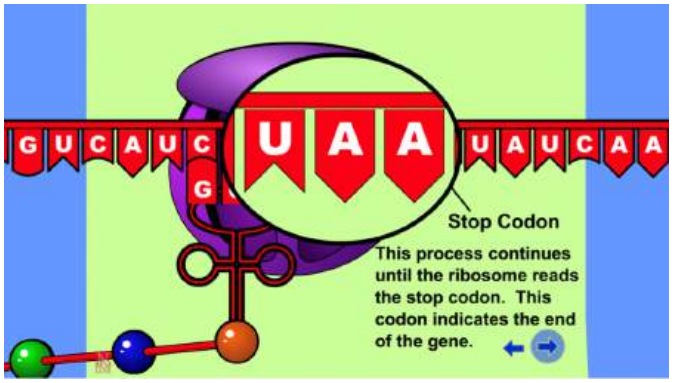


Figure 11: *Ribosome encounters one of the three sequences in the genetic code that function as stop codons. Image by Donald Lee.*

Sequence, structure, and function

In the case of our ALS protein, the assembly of all 667 amino acids into the polypeptide or protein chain in the correct order is important because the sequence of the amino acids determines how the protein folds. The folded structure of the protein determines its function. If some amino acids are changed and the structure changes, the ALS will not function properly as an enzyme in plant or microbe cells.

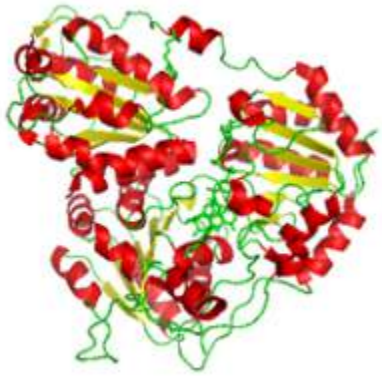


Figure 12. *ALS enzyme structure. The red and green stretches depict different kinds of protein structure features that provide the enzyme its ability to function as a specific enzyme. Image Source: Rj Joplin, CC BY-SA 3.0, via Wikimedia Commons.*

Summary

Organisms use the process of **transcription** to make **mRNA** as a temporary molecule to transport **protein** coding information from a **gene**. The **translation** process reads the mRNA and makes the intended protein, one amino acid at a time. This process is controlled by the **promoter** of the gene to ensure that **proteins** are made in the right cells at the right time to drive growth, development, metabolism, stress response and other processes needed to complete the organism's life cycle.

Video Examples

Watch this video from UNL for an overview of Transcription and Translation.

Circle of Life and Gene Expression

Let us revisit the idea that gene expression works because living cells give rise to more living cells. Imagine yourself when you were just a single cell zygote with a copy of all the genes in the human genome from your biological mother and a copy from your biological father. What else did you need to get from your parents before you could express your first gene and build your first protein? You needed to already have an RNA polymerase in the nucleus of the cell to transcribe your first gene and you needed to have ribosomes, tRNAs and amino acids ready for translating that mRNA in the cytoplasm. Luckily, the egg cell from your mother had the 23 human chromosomes plus all of these gene expression

biomolecules. Later, you could turn on your own genes for RNA polymerase, tRNAs and ribosomes but your mother's egg cell was a living environment that allowed you to start your life.

A plant treated with a herbicide which binds to and blocks the action of the ALS enzyme likewise demonstrates that gene expression has a cycle of dependency on living cells. Small plants in an environment for rapid growth will have their growth halted by the ALS binding herbicides. The plants do not die from the herbicide, they just stop growing. The reason for this is that plants need to build new proteins to grow and to build new proteins they need the ALS enzyme to be working and building new leucine, isoleucine, and valine amino acids. ALS herbicide treated plants stop making these three amino acids and translation stops for every protein in the cell that requires at least one of these amino acids. Living things need amino acids to build proteins and plants and microbes need the ALS enzyme to build three of the amino acids. The next lesson (**Gene expression part 2**) will apply the gene expression principles from this lesson to the observation that some plants can be resistant to ALS inhibitor herbicides.

8. Gene Expression: Applied Example (Part 2)

DONALD LEE; WALTER SUZA; MARJORIE HANNEMAN; AND PATRICIA HAIN

Learning Objectives

This lesson describes how changes in the DNA sequence of a gene can alter the synthesis of a protein and thus influence traits such as herbicide resistance.

In this lesson we will describe how changes in the gene can alter the gene expression process and influence traits in an organism. The specific example of ALS-inhibitor herbicide resistance is used to demonstrate the impact of genetic change on trait expression in a plant.

At the completion of this lesson you should be able to…

1. Explain how a mutation in a single gene can control resistance to herbicides such as ALS inhibitors.
2. Predict the impact of a specific mutation on gene expression and the eventual trait observed in a plant.
3. Distinguish between the molecular and classical definitions of an allele.
4. Describe the molecular basis of dominance or a lack of dominance between alleles.

Sequence Determines Structure, Structure Determines Function

Transcription and translation are fundamental to the production of all proteins found in all living things. That is why discovering the sequence of a gene can allow geneticists to predict the amino acid sequence of the protein that gene encodes. Remember, a completely synthesized protein will be hundreds of amino acids long and thus an average gene will have two or three thousand nucleotides in its sequence. As more is learned about how amino acid sequence dictates protein function, the value of gene sequence information to biologists increases. In this lesson, we will apply our understanding of the gene expression process to understanding the genetic basis of resistance to ALS herbicides in plants.

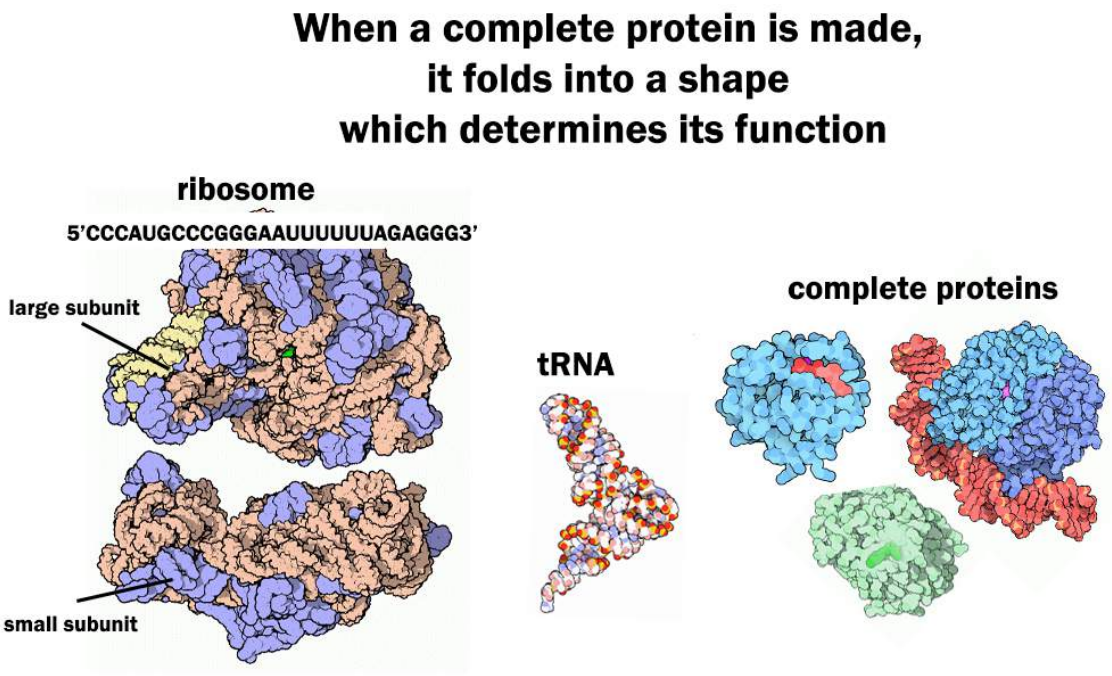


Figure 1. *A protein's function is dependent upon how it folds into a shape. The folding is a result of the sequence of amino acids present after translation. Based on an image by Donald Lee. Adapted by Abbey Elder with figures taken from PDB-101: "Molecule of the Month" collections, available under a CC BY 4.0 license.*

The ALS enzyme is about 800 **amino acids**. The specific order of amino acids in the protein chain dictates how the molecule can be folded up into a secondary structure (Fig. 1).[1] This is because each amino acid has a different chemistry that influences its interaction with other amino acids within the protein chain. The protein's secondary structure determines the function of the protein. We can relate this principle to functional objects that are more familiar to us. Why would someone need a set of wrenches? Every tool in the set is made of the same materials. However, the specific shape and dimensions of each tool allow it to perform a specific function. The ALS enzyme functions to catalyze reactions for amino acid synthesis in the chloroplast in a plant cell. Therefore, the plant ALS **enzyme** structure signals the cell to transport the enzyme to the **chloroplast** and then allows the enzyme to catalyze the formation of a bond between amino acid precursor molecules. Clearly it was important to assemble the amino acids properly when building the ALS enzyme to allow it to function properly.

ALS and Weed Control

The function of the ALS **enzyme** in plants is also an interesting issue in the area of weed control. Herbicides have been discovered with sulfonyl urea or imidazolinone chemistries that will bind to ALS enzymes in plants and disrupt their catalytic functions (Figure 2). Because the herbicides bind tightly to ALS and ALS is not found in high copies in the cell, the herbicides will effectively halt the synthesis of the **Leucine, Isoleucine, and Valine amino acids** (Leu, Ile, Val). How will plants respond to the shortage of these three amino acids? They quit growing because they cannot make three of the twenty amino acids.

1. Graphics used in this figure were taken from The RCSB PDB "Molecule of the Month": Inspiring a Molecular View of Biology D.S. Goodsell, S. Dutta, C. Zardecki, M. Voigt, H.M. Berman, S.K. Burley (2015) *PLoS Biol* **13**(5): e1002140. doi: 10.1371/journal.pbio.1002140

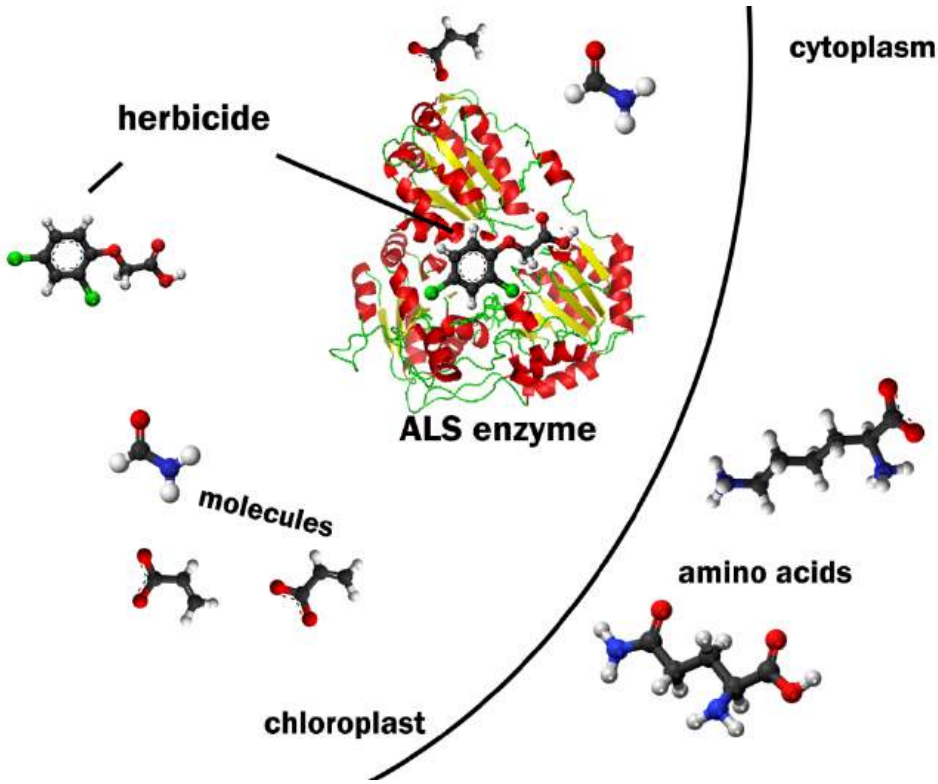


***Figure 2**. The ALS enzyme functions as a part of a metabolic pathway in the chloroplast that takes simple molecules and synthesizes amino acids. Herbicides can bind to the enzyme and block the amino acid synthesis. Based on an original image by Donald Lee, adapted by Abbey Elder. ALS image by Rj Joplin is used under a GNU General Public license. The amino acid Lysine representation by Ben Mills is used under a CC BY SA 4.0 license.*

The **proteins** that need these amino acids will be only partially synthesized. The partial proteins will not work because they lack the structure needed for their function. The plant cells cannot grow and divide and the plant's growth and development are halted. They will not necessarily die but they will no longer be able to compete for light and other resources if nearby plants are not affected by the ALS inhibitors. If the nearby plant that is not affected by the ALS inhibitors is a crop species such as corn or soybean, a valuable post emergence weed control strategy can be implemented.

ALS Alleles

Post emergence weed control by ALS inhibitors works in corn and soybean genotypes with altered ALS genes. The altered genes code for a slightly modified ALS enzyme which is resistant to ALS inhibitors. The molecular basis of this alteration can be determined and help us understand what alleles of a gene are. Even without molecular analysis, geneticists discovered alleles of the ALS gene that would make plants resistant to the ALS herbicides. How were these alleles discovered?

Corn breeders knew that ALS **resistance** did not naturally occur in their breeding lines. This discovery was easy to make. They planted thousands of different lines in short rows in the field, sprayed the field with the herbicide and observed the reaction of the lines. All of the lines tested had significant stunting. Variation did not exist in their lines so they decided to try to induce variation by mutagenesis. Pollen was collected from plants that had been exposed to a chemical mutagen. The chemical induced mistakes in DNA replication during pollen formation. The pollen was viable but carried mutant alleles. This pollen was applied to the silks of other corn plants to produce thousands of seeds. This seed was

planted and the field test for ALS resistance was repeated. This time some plants resistant to stunting from ALS were observed, tagged and self-pollinated. From this experiment, the IT (immi tolerant) allele was discovered. The IT allele appeared to be dominant over the normal ALS allele because plants with one copy of the IT allele per cell (heterozygotes) had the same level of resistance as plants that had two copies of the IT allele (homozygotes). Molecular analysis of these plants was done to reveal how the IT allele was different for the original version.

The IT Allele

The IT allele of the ALS **gene** has a change in the coding region sequence. This was discovered by making a gene library from the DNA of homozygous IT plants and selecting a clone of the IT gene. The gene clone was sequenced and compared to the normal susceptible ALS gene from corn. A single nucleotide substitution in the IT allele was responsible for endowing a corn plant with **resistance** to ALS herbicides. The nucleotide substitution was in the gene coding region so it changed a codon sequence in the mRNA, coding for a different amino acid in the protein. The amino acid substitution had a minor or negligible effect on the enzyme's ability to perform its normal functions. However, the amino acid substitution changed the site at which the ALS herbicide binds to the enzyme. The ALS herbicide will not interact with the ALS enzyme encoded by the IT allele, giving the plant resistance to the herbicide. This is an example of how a small change in a gene can have a big impact on a plant's phenotype in the right environment.

The IR Allele, Tissue Culture Selection for Mutations

ALS alleles that confer **resistance** to immidazolinone herbicides have also been selected from mutations that have occurred in tissue culture cells. DNA replication does not occur 100 percent without errors so it is not surprising that natural mutations occur in cells. A mutation, though, that gives a single cell in a corn plant resistance to ALS herbicides will not render the whole plant resistant. Geneticists have used the power of tissue culture to select for these mutations at the single cell level. Hundreds of thousands of corn cells can be grown in tissue culture plates that contain the herbicide. Most of the cells will quit growing and dividing when they take up the herbicide but some cells continue to grow and form clumps of undifferentiated cells called callus. Plants regenerated from these callus cells will often express the herbicide resistance as well. The IR allele of the ALS gene was selected in this way.

This allele also has an altered coding region sequence that blocks the ALS herbicides from binding. The ALS enzyme encoded by the IR allele, however, does not provide desired levels of herbicide resistance in a heterozygous plant. The molecular mechanisms causing this difference in the IR compared to IT allele are not clear. The implication of this difference in plant breeding are significant. The IT allele can be inherited from just one parent to give the 'Clearfield' or "Immi" hybrid herbicide resistance. Hybrids must be homozygous for the IR allele requiring both parents be homozygous IR inbreds (Figure 3).

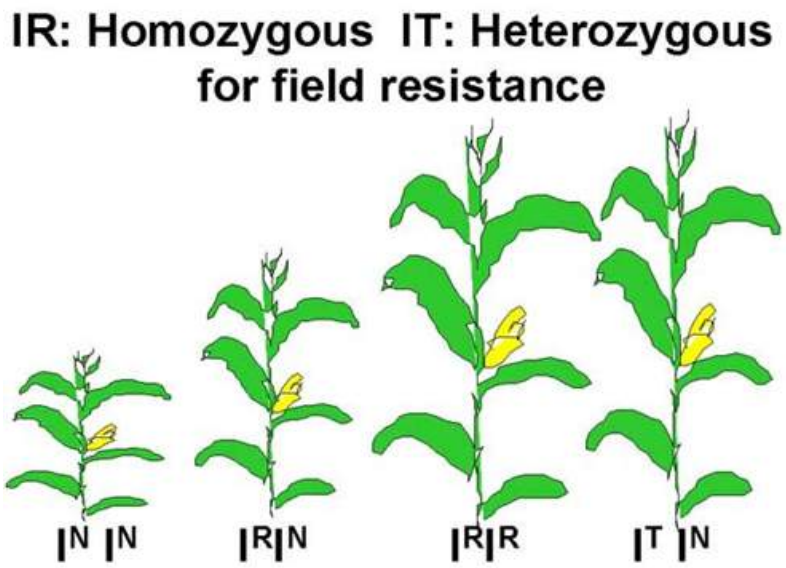


Figure 3. *Corn plants sprayed with an ALS inhibitor herbicide vary in their growth. The plant can have the IR, IT or IN (normal) alleles and the genotype of the plant for the ALS genes determines its phenotype in response to these herbicides. Image by Donald Lee.*

Molecular Basis of Dominance

The occurrence of the IT, IR and normal alleles of the ALS enzyme and their impact on plant phenotype provides us with an opportunity to think about dominance and lack of dominance at the molecular level. The IT allele has complete dominance to the normal allele because there is apparently no phenotype difference between homozygous dominant and heterozygous IT plants when they are sprayed with an ALS inhibitor. This suggests that in the heterozygous plants, one copy of the IT allele per cell encodes enough resistant ALS protein to drive the synthesis of sufficient quantities of Val, Leu and Ile for normal growth in the corn plant. However, the ALS enzyme encoded by the normal allele in these heterozygotes can be inhibited by the herbicide and will not contribute to amino acid synthesis.

In contrast, the IR allele has a lack of dominance to the normal allele (IN). Heterozygous plants that have one copy of the IR and one copy of the normal allele per cell do not maintain normal growth after ALS inhibitor treatment. One copy of the IR allele per cell does not encode sufficient copies of a resistant ALS enzyme to support normal growth. Instead, two copies of the IR allele per cell are needed so that all of the ALS enzyme made will not be bound by the herbicide.

The classification of the plant phenotype for characterizing possible genotypes at the ALS locus may depend on environment. In some growing environments it has been noticed that heterozygous IT hybrids will show a slight growth reduction in the sprayed plants compared to unsprayed. The difference is minor because the plants usually grow out of this difference later on and yield reduction is not a factor. This observation does emphasize that phenotype is a result of genotype and environment. Therefore, scoring traits in the proper environment will be critical to discovering the nature of gene action.

How Universal is the Genetic Code?

The table of codons used by organisms to translate mRNA into proteins is shown on the bottom of the page (Table 1).

As was mentioned earlier in this lesson, the genetic code needed to be cracked one time because all organisms used the same codons to encode amino acids. As scientists began to sequence the coding regions of genes from different organisms, they discovered something called a codon preference. When you look at the codon table, you can see that the genetic code is redundant. This means that more than one codon can encode the same amino acid. This is because there are 61 codons that code for the placement of 20 different amino acids. A codon will only work in coding if a tRNA with a complementary anticodon is also found in the same cell and has the appropriate amino acid to deliver. Therefore, there could be 61 different tRNAs, one to complement each codon. Each different tRNA needs to be encoded by a different gene. If that gene is not expressed in the cell, the tRNA will not be found and a codon that needs to be complemented by that tRNA will not be complemented. In this case, the codon will act like a stop codon. The ribosome will halt **translation** and the protein made will be a shorter version of the intended protein. Organisms would not benefit from this situation so there is a tight complementation between what tRNAs genes are present and expressed in an organism's cells and what codons are used to encode a specific mRNA. In this way the genetic code will have a dialect. The language is universal but certain words are used preferentially.

Table 1. RNA Codon Table. The genetic code. The 20 common amino acids are listed in their three and one-letter formats. Stop codons are UAA, AUG and UGA

First Position	Second Position				Third Position
	U	C	A	G	
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
					C
	Leu (L)		Stop	Stop	A
			Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
					C
			Gln (Q)		A
					G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
					C
			Lys (K)	Arg (R)	A
	Met (M)				G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
					C
			Glu (E)		A
					G

Scientists are not sure why codon preferences are a part of the **gene expression** process in organisms. It may provide another level for the organism to control the amounts and kinds of proteins made in its cells. Recent experiences in genetic engineering of plants and animals, however, has made codon preference an important consideration. For example, scientists have put genes from a soil bacterium called *Bacillus thuringiensis* (Bt) into corn plant cells in order to give the corn plant the ability to make a protein that is toxic to European corn borer, a common pest to corn producers. They found that the gene would be transcribed but the mRNA would not be translated to make the desired protein. One reason was codon usage. Some of the codons the bacteria use to encode amino acids are rarely used by corn. The corn plant either lacked the tRNA to complement the codon or make the tRNA at such low levels that there were not enough copies in the cell to accommodate translation of the Bt mRNA. Therefore, the genetic engineers needed to make synthetic coding regions that substituted codons preferred by corn for those preferred by bacteria. The end result was they were able to get higher levels of the Bt protein made once these changes were made in the gene. Codon preference thus makes the genetic engineering process more challenging.

Learning Activities

Watch this animation of the transcription process.

Watch this animation of the translation process.

Watch this video overview of Transcription and Translation.

9. Regulation of Gene Expression

WALTER SUZA; DONALD LEE; PHILIP BECRAFT; AND MARJORIE HANNEMAN

Learning Objectives

1. Understand the concept of gene expression.
2. Understand transcriptional regulation of gene expression.
3. Understand epigenetic gene regulation.
4. Understand post-transcriptional regulation of gene expression (RNA level).
5. Understand post-translational protein modification and regulation.

Introduction

Every cell in a plant contains the same genetic information, the same set of genes. Yet Therefore different sets of genes are required for the various functions of different cells or tissues, as well as for plant responses to environmental stimuli or stresses. This is achieved by regulating the activity of genes according to the physiological demands of a particular cell type, developmental stage, or environmental condition. This regulation of activity is known as **gene expression**.

The term **expression** can be used in different ways that are sometimes confusing. Typically, if a gene product is produced, the gene is considered "expressed". However, it sometimes occurs that a transcript might be produced but not a protein, or that a protein is produced but it is in an inactive state. In such cases, although a gene product is produced, the biological activity encoded by that gene is not present. For the purposes of this section, the key point is how the biological activity encoded by a gene is regulated.

The expression of genes in specific plant cells, tissues, and organs and the timing of this expression require a precise level of regulation. Expression, or genetic function, can potentially be regulated at any of the steps from transcription, RNA processing, translation, through post-translational protein modification, as discussed in lesson 1. Regulation can be qualitative (i.e. gene expression is either "on" or "off") or quantitative (i.e. expression levels can be modulated "up" or "down"). Fluctuations in the intensities of external stimuli coupled to changes that occur at the genomic level result in different developmental outcomes or physiological states. Regulation of gene expression at the level of transcription can be brought about through chromatin and histone modifications. Also, a gene sequence can be differentially spliced to produce mRNA products of variable lengths leading to new protein products with novel functions. Some genes do not encode proteins but short forms of RNAs with regulatory functions such as induction of flowering. Finally, proteins products can be subjected to modifications such as phosphorylation or dephosphorylation to alter their functions, or can be completely degraded to turn off a gene.

Transcriptional gene regulation

Since transcription is the first step in gene expression, it makes sense in terms of cellular economy, to regulate expression at this point, and in fact this is one of the most important regulatory points. We already described the involvement of RNA polymerase in the transcription process, but in fact there are other protein factors that are required. Proteins involved in transcriptional regulation are known as transcription factors. It is the interaction of these transcription factors with specific DNA sequences that regulate the process of gene transcription.

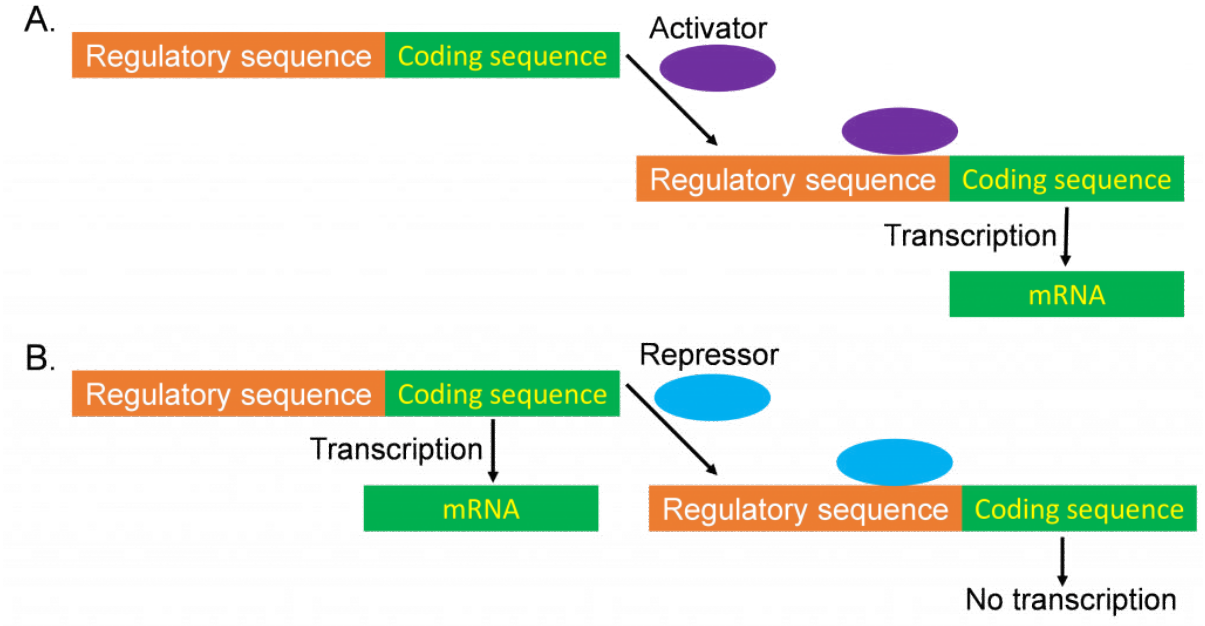


Figure 1. *Positive regulation of gene transcription (A) occurs when a transcriptional factor (activator) turns on gene transcription. Negative regulation of gene transcription (B) occurs when a transcription factor (repressor) turns off gene transcription. Image by Walter Suza.*

A. The concept of differential (regulated) gene expression

As just described, not every gene is expressed all the time. When a gene displays different levels of expression in different circumstances, this is known as **differential expression**. Circumstances that might apply include, but are not limited to, different plant tissues (root vs. leaf), different developmental stages (germination vs. reproductive development), or in response to different environmental stimuli (cold stress or pathogen attack).

The term differential expression can also be used to compare the expression of different genes. If two genes show different expression patterns (among plant tissues or in response to environmental stimuli), they are considered differentially expressed, whereas genes that showed very similar patterns of expression would be considered **co-expressed**.

B. Promoters

As mentioned, transcription is regulated through the interactions of proteins, transcription factors, with specific DNA

sequences. Most regulatory DNA sequences governing gene transcription are located on the 5′ border of the transcribed region. This region is called the gene **promoter**.

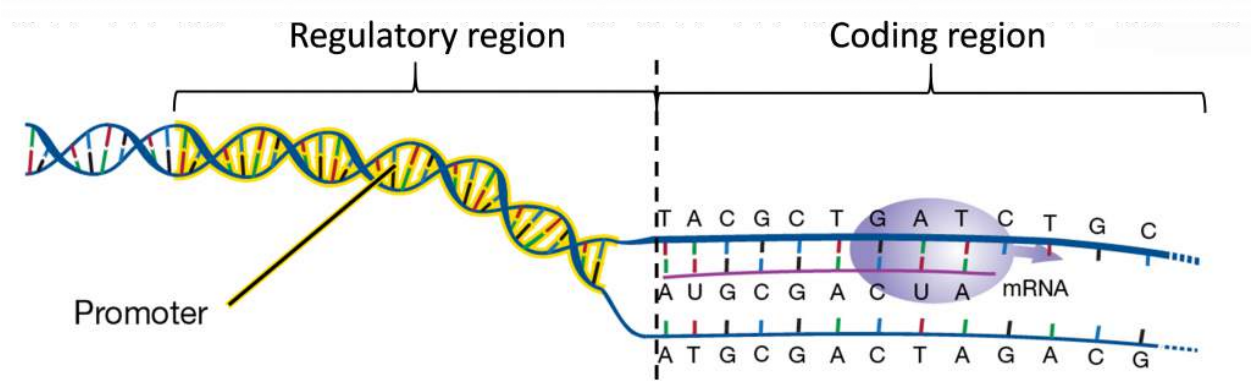


***Figure 2**. Promoters are part of gene regulatory sequence. Adapted from Illustration by NIH-NHGRI.*

Promoters contain a core, which is required for the binding of the "basal transcriptional machinery", including RNA polymerase. The "TATA" box, with a consensus sequence TATAA, is located within the core promoter, usually 25-30 nucleotides upstream of the transcription initiation site. Promoters also contain regulatory sequences that determine when, where, and to what level genes are transcribed. Promoters can vary in length from a hundred to a few thousand nucleotides.

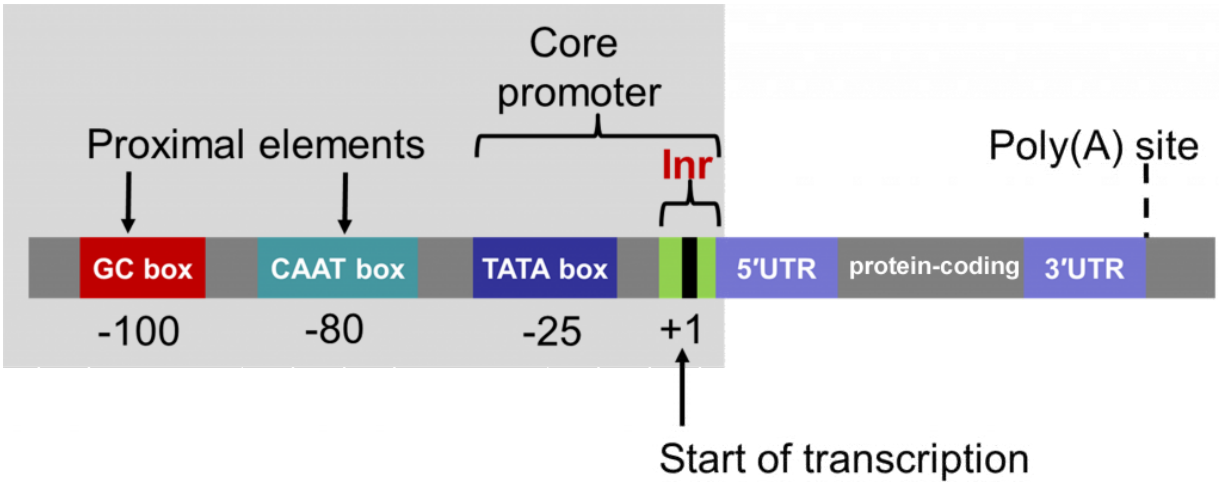


***Figure 3.** The promoter contains core and proximal elements. Image by Walter Suza.*

The full promoter sequences of different genes that are expressed in a similar manner may be different. However, such promoters often contain short sequence "motifs" that are similar, referred to as cis elements. Early work (Benfy and Chua, 1990) to understand the function of different promoter elements in regulating gene expression in plant cells and tissues revealed that various combinations of cis elements are able to be interpreted by the cell and control gene expression. Sometimes cis elements promote gene transcription and sometimes they function to restrict transcription of genes in particular cells and tissues.

Promoter analysis is facilitated through the use of **reporter genes**. The reporter gene produces effects that are easily identifiable and quantifiable, which can be used to determine the function of a regulatory region of another gene (promoter, promoter elements, or enhancers) in cells, tissues, or organs. Such analyses are critical to crop biotechnology where targeted expression of genes in particular tissues is often desirable. To test whether a promoter is effective in conferring the expression of a gene to a particular tissue, scientists fuse the putative promoter to a reporter gene

and introduce the promoter-gene fusion into plants. An example of a reporter is the GUS gene, which encodes a GUS enzyme (beta-Glucuronidase) that, when expressed, produces a blue color upon addition of a substrate. Figure 1 shows an example of GUS expression in maize seed using two promoters that act either in the endosperm or the embryo.

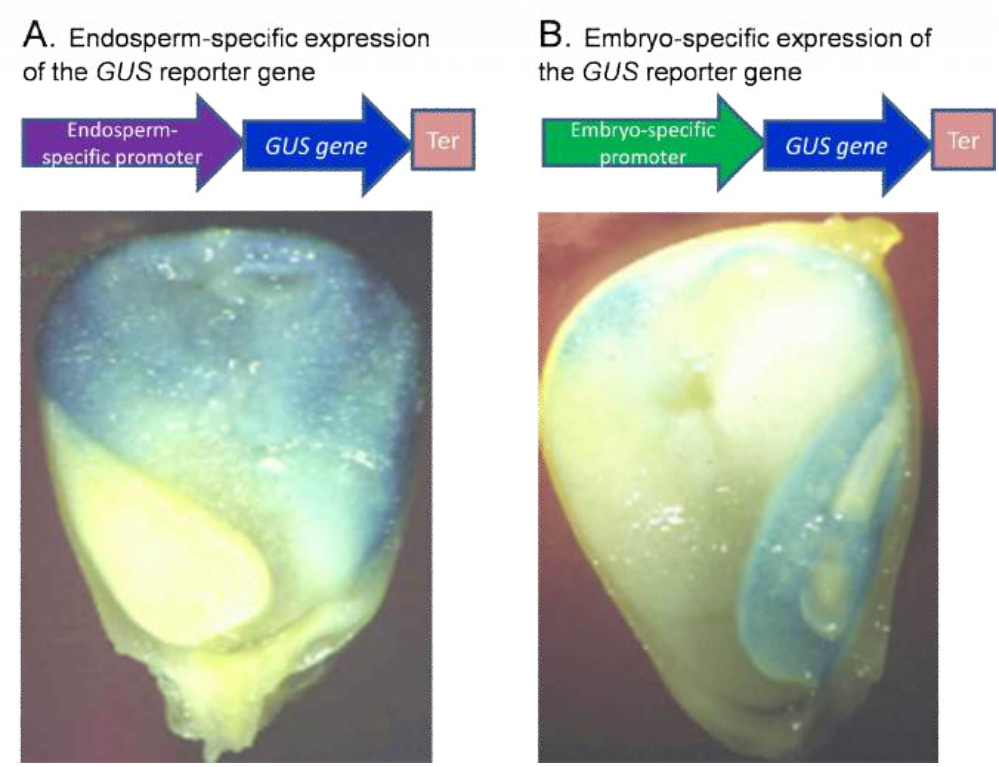


***Figure 4**. Promoters direct the expression of genes in plant tissues. (A) The expression of the GUS transgene is restricted to the endosperm of maize seed by an endosperm-specific promoter. (B) When the GUS transgene is fused to an embyo-specific promoter the activity of the GUS enzyme (production of the blue coloration after addition of a GUS substrate) is restricted to the embryonic tissue. Image by Shui-zhang Fei.*

C. Enhancers

Enhancers are DNA sequences that increase the rate of transcription of a gene when they are present, although alone they cannot cause transcription to occur. Enhancers are usually position and orientation independent. Although they are normally located upstream of the promoter, they can also be located on the 3' region of the gene or even with the coding region. Enhancers can increase the transcription when added to genes that they are not normally associated with. This is a useful property for biotechnology, allowing promoters to be manipulated for increased levels of transcriptional regulation. Some enhancers function at all times in all cells and tissues and they are referred to as constitutive. Other enhancers function in specific tissues at specific developmental stages. Some are active only in response to environmental signals. The AACCA enhancer on the promoter of soybean β-conglycinin gene encoding a seed storage protein functions specifically in seeds. Enhancers are thought to interact with specific nuclear proteins involved in transcription. For example, enhancers might facilitate the binding of transcription factors and direct these factors along the DNA strand to the direction of the promoter. Alternatively, enhancers may facilitate changes in DNA structure such as modification of the chromatin structure.

The counterpart of an enhancer is a **silencer**. Silencers have all the properties just described for enhancers, except they function to dampen, or decrease, the levels of transcription controlled by a promoter.

D. Transcription factors

RNA polymerase binds the promoter at the TATA box and in cooperation with other proteins drives gene transcription. The proteins that interact with RNA polymerase to facilitate its binding to the promoter and to regulate its activity are known as **transcription factors**. Some transcription factors, known as "basal transcription factors" are fundamental to RNA polymerase binding and function, and are expressed in all living cells.

Other transcription factors bind to cis regulatory DNA sequences of promoters, enhancers or silencers. These proteins interact in complex ways with the basal transcription machinery, to regulate the activity of RNA polymerase, and therefore gene transcription. Thus, transcription factors regulate when or where individual genes are expressed, and to what level.

Transcription factors can function as either positive or negative regulators. That is, they can either function to induce (increase) gene transcription or to repress it. The consequence of many transcription factors depends on their interactions with other proteins. Two factors together may be required for gene activity, and exclusion of one of the factors in space and time offers a mechanism for differential gene expression. Some transcription factors might function as a positive regulator in one context but as a negative regulator in another context, depending on what other cis elements and/or transcription factors might be present.

Epigenetic regulation

A. Chromatin structure and histone modification

At the molecular level chromatin is a product of an ordered and tight packaging of the double stranded DNA around nucleosomes (a core of proteins named histones), and an association with additional proteins.

Transcriptional regulation often involves modification of chromatin structure mediated by post-translational regulation (changes in acetylation and methylation) of histones. Histone acetylation involves the addition of acetyl groups and when histones are heavily acetylated the DNA is less tightly associated with them. This often correlates with increased transcriptional activity of specific genes. The idea is that when DNA is loosely associated with histones, it is more accessible to transcription factors that require interaction with the DNA to initiate transcription. Consequently, histone deacetylation (removal of acetyl groups) by histone deacetylase enzymes (for example, FLD, p462) stabilizes nucleosomes and represses transcription. On the other hand, histone acetylation by histone ecetyltransferase destabilizes nucleosomes and promotes transcription.

As described in the text, another form of histone modification is the addition of methyl groups to histone proteins, which similarly regulates chromatin condensation or decondensation.

B. DNA methylation

The nucleotide bases of DNA can be modified by the attachment of methyl groups at various locations. The addition of these methyl groups happens after the DNA has been synthesized and is controlled by enzymes that add methyl moieties to specific regions of DNA. The most common modified nucleotide base is C5 methylcytosine (m5C) due the activity of

the enzyme DNA (cytosine-5) methyltransferase (MET). MET recognizes the unmethylated newly-replicated DNA strand and incorporates a methyl group if the template strand was methylated (hemimethylated).

Methylation status is correlated with gene expression (Figure 5) as demonstrated by the low level of methylation in regions of the genome undergoing active transcription.

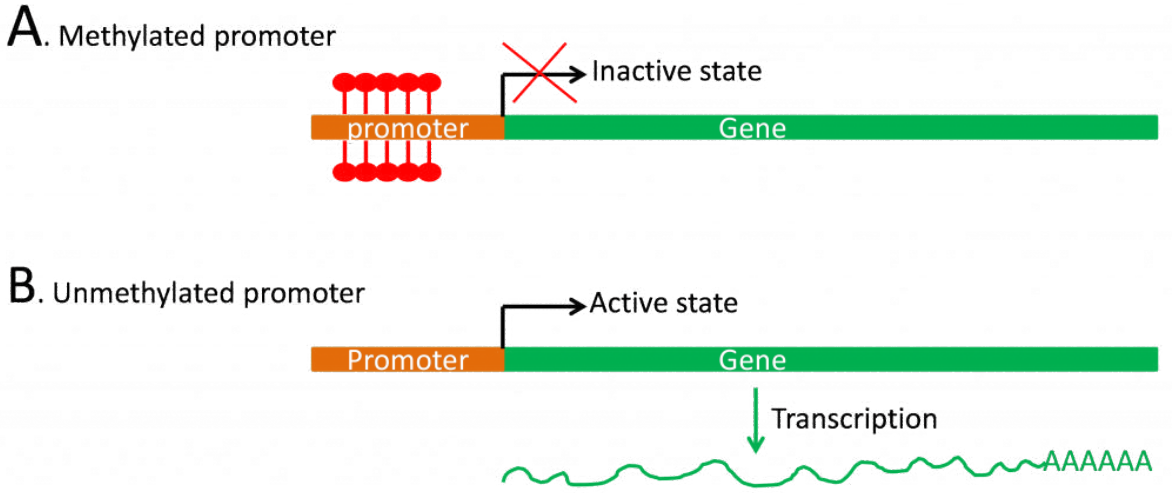


***Figure 5**. Absence of methylation correlates with gene expression. (A) The promoter region may be methylated during development or environmental conditions such that a gene is inactivated. (B) The same gene is activated upon the removal of the methyl groups.*

Studies have shown that altering a plant's overall DNA methylation status can affect growth and development. For example, cold treatment (vernalization) induces flowering in biannuals such as Arabidopsis and winter wheat, and lowers the level of methylation of particular genes. Also, treating plants with the drug 5 azacytidine prevents methylation at the 5 position of cytosine and stimulates flowering. Recent studies also suggest a relationship between epigenetic gene regulation and heterosis, with hybrids showing higher global levels of transcription, higher histone acetylation levels and lower DNA methylation levels (reviewed in He et al., 2011).

RNA-level regulation

Regulation of gene expression occurs at many levels, including post-transcriptionally. From the standpoint of the expression level for a given gene, a critical factor is the level of fully processed, mature mRNA. When we consider the level of the mature form a particular transcript at any given time reflects the steady state balance between synthesis, processing and degradation. The Post-transcriptional regulation of RNA occurs through several mechanisms.

RNA stability and degradation

The control of mRNA degradation rate, or turnover, is an important regulatory mechanism in gene expression. As mentioned, the level of a particular transcript at any given time reflects the steady state balance between synthesis and degradation. Thus, even though a gene may be transcribed at a high rate, a high rate of RNA turnover could effectively shut off the gene. The stability of mRNA can be regulated globally, to affect all or most transcripts, or it can be very specific to a particular mRNA. Regulated mRNA turnover may be particularly important in plants, since plants cannot

move to avoid harsh environmental conditions and stresses are known to induce specific changes in RNA stability. Other factors such as light and plant hormones are also known to regulate mRNA stability.

The degradation of mRNA is catalyzed by enzymes called ribonucleases. Ribonucleases include exonucleases, which degrade only from one end of the transcript, and endonucleases, which can attack the mRNA molecule internally.

Therefore, controlling ribonuclease access to the mRNA substrate is an important regulatory mechanism in gene expression. This control is achieved through various mechanisms including, controlling the amount of nucleases present, regulating the activity of the nuclease, sequestering the nuclease to particular cellular locations to restrict its access to RNA, and to control the stability of the mRNA by decreasing accessibility to nuclease.

Four regions of an mRNA are important for its overall stability. These include, the 5′ untranslated region and the cap, coding region, 3' untranslated region, and the poly(A) tail.

The cap at the 5′ end protects the mRNA from exonucleases that degrade RNA from the 5′ to the 3′ direction.

The 3′ untranslated region may contain short sequences that influence mRNA stability. For example, the repetition of the sequence AUUUA in the 3′ untranslated region of many animal and plant genes is associated with mRNAs with short half-lives (the time it takes for half the RNA to be degraded, after transcription is stopped).

The poly(A) tail increases mRNA stability but is not sufficient alone. It only provides stability when bound by proteins such as poly(A) binding protein (PABP). Test-tube experiments have shown that, removing PABP from a stable mRNA destabilizes it, while adding back purified PABP restores stability.

Translational regulation

As described, for most genes it is the protein product that performs the biological function. As such, regulating the amount of protein production effectively regulates gene expression. There are several known mechanisms by which the process of translation is known to be regulated. A detailed consideration of these mechanisms is beyond the scope of this course, but in general it is important to bear in mind their importance. For example, under certain stress conditions, "normal" translation is halted and only mRNAs related to stress tolerance are selectively permitted to be translated into proteins. This selective translation is important to allow plants to quickly respond to stress and to conserve energy under stress conditions.

Translation of mRNAs can play an important role in determining their overall stability. Mutations that add premature stop codons often lead to rapid degradation of the mRNA. Ribonucleases (figure 9) or other factors involved in degradation may recognize the number or spacing of ribosomes on an mRNA, and degrade those that are not produced properly. This may help prevent synthesis of proteins with incorrect functions which will negatively impact cellular processes. Errors during transcription could add or omit nucleotides which would alter the proper codon sequence creating mutant proteins.

Protein-level regulation

After translation, proteins are subject to a variety of modifications that can regulate their activity. There are many different ways by which proteins can potentially be modified and thereby regulated, and only a few of the basic mechanisms will be considered here.

Common types of post-translational modifications

The first mechanism is covalent modification by the addition of various chemical groups. A wide variety of groups can be involved, including small organic groups (methylation, acetylation) lipids (myristoylation, farnesylation, palmitylation), carbohydrates (glycosylation, glucosylation), small proteins (ubiquitination, sumoylation) and inorganic molecules (phosphorylation, sulfation). Such covalent modifications are generally accomplished by the activity of enzymes specialized to perform these modifications. Many of these modifications are also reversible; that is these groups can be added to a protein and subsequently removed. Phosphorylation is a particularly noteworthy reversible modification that is common in the regulation of many proteins. Enzymes that add phosphate groups to other proteins are called protein kinases, and those that remove phosphates are called phosphatases. As such, protein kinases and phosphatases are central to many cellular regulatory systems.

A second common mechanism of protein modification is through proteolytic cleavage. Proteolysis occurs as part of the general turnover of cellular proteins, which is required to eliminate damaged proteins and recycle amino acids. Proteolysis is important for the processing of certain proteins, for example in the removal of the signal peptide of proteins targeted to specific cellular compartments. It can also occur in a highly specific manner whereby particular proteins are targeted for degradation or for cleavage at a specific site within the protein. Proteolytic modifications are non-reversible.

Proteins can undergo modification through complex formation. Such complexes can occur among proteins or between a protein and a cofactor.

Proteins can also be modified according to conditions of the cellular environment. The redox state can result in oxidation or reduction of proteins, particularly of sulfhydryl side groups. Cellular pH can affect the charge of ionizable side groups.

Common types of protein regulation

All of the above mentioned types of protein modification can alter protein conformations and thus have regulatory consequences on protein function or activity. Protein regulation is highly complex and there are a myriad of different ways by which this occurs. Again we will just briefly consider a few of the more common mechanisms.

One major way protein modification can regulate protein function is by **altering their activity**. This can be true for many types of proteins including, but not limited to, enzymes, transcription factors, signaling proteins, and structural proteins.

Cells are **compartmentalized** into several membrane bound organelles, including the nucleus, chloroplasts, mitochondria, peroxisomes, endoplasmic reticulum (ER), golgi, and vacuoles. Each of these compartments performs unique metabolic functions that require a set of proteins. Compartmentation is regulated for some proteins. For example, upon exposure to light the phytochrome protein moves into the nucleus where it affects the expression of light-regulated genes.

Proteolytic processing is involved in several important regulatory processes. Many proteins are synthesized in an inactive form that requires proteolytic cleavage for activation. The full-length translation product before processing is often called the precursor, or a preprotein. As mentioned, cells contain membrane-bound compartments. Since membranes are impermeable to most proteins, an active mechanism is necessary to move a protein across a membrane. For proper delivery to their organellar destinations, specific amino acid sequences, called target signals must be present

in a protein (to serve as an "address"). For example, entry into the chloroplast is achieved by the presence of a target signal called the transit peptide. This is at the amino terminal end of the protein and is proteolytically cleaved during import.

Another example of proteolytic processing is seen in a plant defense response in members of the solanaceae (for example, tomato and potato). An 18-amino acid peptide hormone called systemin is secreted by plant cells that are damaged by insects or mechanical wounding. Systemin production by wounded cells is required to induce the synthesis of proteins involved in defense. Systemin induces defense responses in wounded cells, and throughout the plant. Analogous to animal peptide hormones (e.g., insulin), systemin is initially synthesized as a much larger (200 amino acids) precursor called pro-systemin. Pro-system is inactive; however, upon wounding it undergoes proteolytic cleavage to produce activated systemin.

Targeted protein degradation

The amount of protein present in a cell or tissue is determined by both its rate of synthesis and its rate of degradation. Therefore, protein degradation is an important mechanism by which the plant can regulate biological activity (i.e., a genetic function). For example, one way to shut down a metabolic pathway is by degrading one of the key enzymes controlling the rate of the entire pathway. Therefore, protein degradation is an essential component of gene regulation to meet cellular demands for growth, development, and defense.

Protein degradation must be carefully controlled to fine tune gene expression to allow plants to adapt to new environmental conditions. Often, cells will adopt several complex mechanisms for proteolytic degradation of proteins. Enzymes that cleave or degrade proteins are referred to as proteases. For example, plant vacuoles are rich in proteases that play a similar function in protein degradation as that of lysosomes in animal cells. Protease activity must be tightly regulated to prevent accidental degradation of essential proteins. Sequestering proteases in particular organelles like the vacuole separates them from other organelles and is one of to control their activity.

An important mechanism by which specific proteins are targeted for degradation is through ubiquitin-mediated proteasomal degradation. The proteasome is a large complex of multiple protein subunits that have a protease activity to degrade proteins. Proteins get marked for proteasomal degradation with a small protein called ubiquitin. Ubiquitin is covalently attached to specific proteins in response to environmental or developmental signals. This allows plants to quickly adapt to changing conditions by eliminating proteins whose functions are not advantageous under the new conditions. For example, the photoreceptor phytochrome mentioned above becomes targeted for degradation when light is no longer available. This allows plants to change their physiological functions going from daylight to night conditions. Protein degradation by the proteasome system is also an important regulatory mechanism for plant hormone signaling, for example the signaling of the defense hormone jasmonic acid, and growth hormone gibberellic acid.

Proteins have lifetimes that range from a few minutes to weeks or more. Cells continuously make proteins from and break them down to amino acids. One of the functions of protein degradation is to eliminate aberrant or damaged proteins which could harm the cell. The second function is to facilitate the recycling of amino acids. For example, most of the amino acids required for growth of the seedling are derived from the degradation of seed storage proteins. Conversely, in annual crop plants, many of the amino acids in seed storage proteins are derived from proteins degraded in leaves and other plant parts during senescence.

Lesson Summary

The expression of genes in specific plant cells, tissues, and organs and the timing of this expression require a precise level of regulation. A single promoter may not be sufficient to regulate the expression of such gene(s) in space and time. Therefore, coding regions with the same function may have different promoters, and such genes are referred to as differentially regulated. Most regulatory sequences governing gene expression are located on the 5' border of the coding region. Transcription is often initiated between 20 and 60 nucleotides upstream of the ATG start site. Enhancers are usually position and orientation independent. Although they are normally located upstream of the promoter, they can also be located on the 3' region of the gene or even with the coding region. Enhancers can increase the transcription when added to genes that they are not associated with. RNA polymerase binds the promoter at the TATA box and in cooperation with other proteins drives gene transcription. The proteins that interact with RNA polymerase bind to regulatory sequences upstream and downstream of the transcription site.

Transcription factors can play a regulatory role by determining where individual genes are expressed. Transcriptional regulation often involves modification of chromatin by changes in acetylation and methylation of histone. The nucleotide bases of DNA can be modified by the attachment of methyl groups at various locations. Methylation status is correlated with gene expression. Alternative splicing describes an alternative mechanism of pre-mRNA processing to generate mRNAs that have different combinations of exons. The control of mRNA degradation rate, or turnover, is an important regulatory mechanism in gene expression. RNA interference (RNAi) is a post-transcriptional process involving the degradation of mRNA initiated by the formation of double-stranded RNA (dsRNA) of the target mRNA. Translation of mRNAs can play an important role in determining their overall stability. Mutations that add premature stop codons often lead to rapid degradation of the mRNA. Protein degradation is an essential component of gene regulation to meet cellular demands for growth, development, and defense. The plant can alter the activity of a metabolic pathway, by degrading one of the key enzymes controlling the rate of the entire pathway.

Exercises

1a. Which of the following could alter gene regulation

i. Deleting a promoter
ii. Altering fertilizer application rate
iii. Raising the temperature of the greenhouse
iv. Herbicide application
v. Reduce irrigation frequency

1b. The detection of a gene product, for example, RNA or protein at any one moment is a reflection of the "steady state" of the product. Describe the term "steady state" in this context.

2. You are studying expression of a gene responsible for resistance to a pathogen. You cloned the gene and made antibody to detect the protein it encodes by a procedure called Western blot analysis, and the mRNA by a procedure called reverse transcription PCR. Explain the type of regulation from the scenarios in the table.

Scenarios	**Not treated with pathogen**		**Treated with pathogen**	
	mRNA	**Protein**	**mRNA**	**Protein**
1	Absent	Absent	Present	Present
2	Present	Absent	Present	Present
3	Present	Present	Absent	Absent
4	Present	Present	Present	Absent
5	Present	Present	Present	Present

3. You are studying the expression of a gene controlling height in your crop species. You have cloned the gene and are interested in determining its expression in different parts of the plant. You use the RT-PCR procedure as in the previous problem. After the PCR part, you load products on a gel and observe the following pattern. Explain your results in the context of gene regulation.

References

Benfey, P. N., N. Chua. 1990. The cauliflower mosaic virus 35S promoter: Combinatorial regulation of transcription in plants. Science 16:959-966.

Chung, H. S., G. A. Howe. 2009. A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in Arabidopsis. Plant Cell 21:131-145.

He, G., Elling, A.A., and Deng, X.W. (2011). The Epigenome and Plant Development. Annual Review of Plant Biology 62, 411-435.

ISHIHAMA, N., R. YAMADA, M. YOSHIOKA, S. KATOU and H. YOSHIOKA, 2010 Phosphorylation of the Nicotiana benthamiana WRKY8 Transcription Factor by MAPK Functions in the Defense Response. Plant Cell 23: 1153-1170.

Nakaminami, K., Matsui, A., Shinozaki, K., and Seki, M. (2011). RNA regulation in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA) – Gene Regulatory Mechanisms.

Sheen, J., 1991. Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. Plant Cell 3:225-245.

10. Genetic Pathways

WALTER SUZA; PHILIP BECRAFT; DONALD LEE; AND MARJORIE HANNEMAN

Learning Objectives

1. Understand how multiple genes often function together to control a trait
2. Understand the concept of a biosynthetic pathway
3. Understand the concept of a regulatory pathway
4. Understand the concept of a network
5. Understand how mutations in different genes of a pathway can cause the same phenotype
6. Understand how mutations in different genes of a pathway might cause different phenotypes

Introduction

Genes do not function in isolation but rather suites of genes act in concert to perform biological functions. When different genes function in different sequential steps of a biological process, this is known as a **genetic pathway**. Perhaps the most conceptually intuitive type of pathway is a **biosynthetic pathway**, where a precursor molecule is chemically modified through a series of enzymatically catalyzed intermediate steps to produce a bioactive product. Since enzymes are (proteins), the genes that encode these enzymes are considered a genetic pathway. In a **regulatory pathway**, some type of stimulus leads to a change in the expression or activity of a particular gene product, which in turn acts to alter the expression or activity of another gene product or products, which in turn could regulate yet another level of activity.

Ultimately, these regulatory changes in expression or activity lead to a response to the stimulus, typically involving changes in gene expression. Again, the genes that encode the various types of regulatory molecules, mainly proteins, constitute a genetic pathway. Pathways are often targets in the biotechnological manipulation of traits, which requires an understanding of how pathways behave. Also note that many biosynthetic and regulatory systems are better described as networks because of often, pathways branch or converge and interact with other systems. However, for simplicity such complex networks will not be discussed here.

Biosynthetic pathways

Genes encode enzymes that catalyze steps in the synthesis of a compound. A **biosynthetic pathway** actually describes a process of converting a precursor substrate into a product. In addition to the precursor substrate and product, it includes the enzymes, the chemical reactions catalyzed by the enzymes and all the intermediate compounds. A pathway consists of a series of steps where a precursor molecule acts as a substrate for an enzyme, which catalyzes a chemical reaction to produce a product, which then serves as a substrate for a subsequent step.

We have already discussed the one gene, one enzyme hypothesis which was an important advance in the discovery of how genes encode proteins. This analysis also provides a nice example of the concept of a biosynthetic pathway. This hypothesis was discovered by analyzing **auxotrophic** neurospora mutants that were defective in the production of amino acids. That is, the biosynthetic pathways producing particular amino acids were disrupted by mutations. The text discusses the analysis of the arginine pathway. Mutants that disrupt this pathway cannot grow on minimal medium lacking arginine but can grow if arginine is supplied in the medium. The same fundamental pathway also produces arginine in plants.

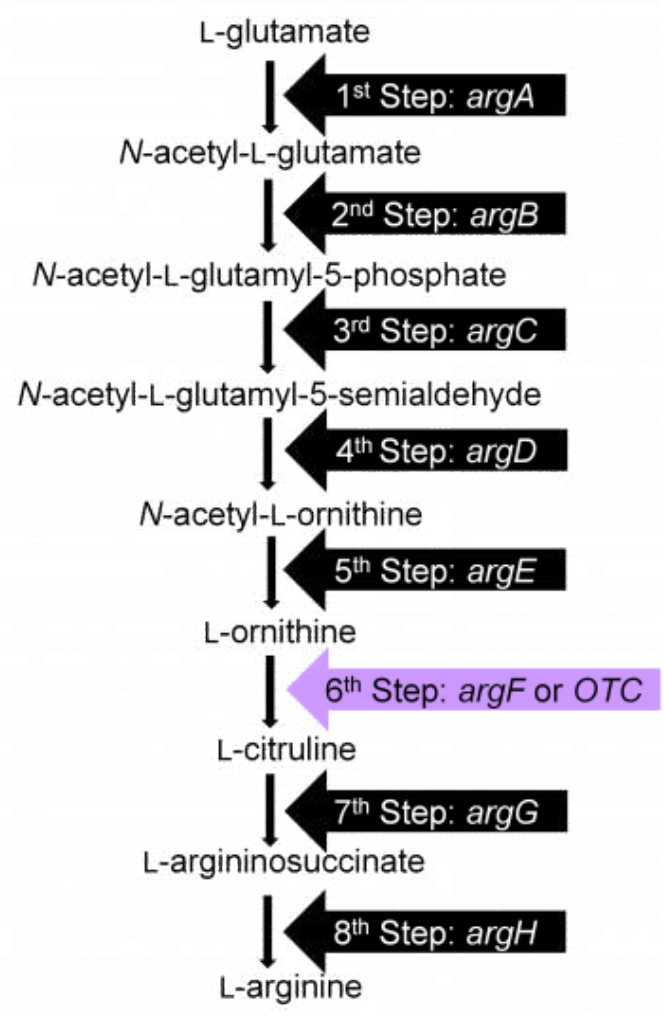


Figure 1. The arginine biosynthetic pathway. The labeled arrows represent chemical reaction steps. The italicized 4-letter symbol (e.g., argF) represents the gene that encodes the enzyme abbreviated to the right that catalyzes each step (e.g., OTC).

Mutations that disrupt the function of particular enzymes in the pathway block the progression of the pathway at that corresponding step. For example, a mutation in the argF gene would disrupt the enzyme ornithine carbamoyltransferase (OTC). This would block step 6 in the pathway, the conversion of L-ornithine to L-citruline. As a result, the intermediate compound L-Ornithine will likely accumulate, whereas all the compounds that occur later in the pathway (i.e., L-Citruline, L-argininosuccinate and L-arginine) will be deficient.

The anthocyanin pathway

The anthocyanin pathway is a slightly more complex pathway, although still simple as far as these things go. It is of significant interest to horticulturists because anthocyanin pigments are responsible for many plant colors, particularly in flowers and fruits (think red wine). Anthocyanin pigments range from orange to red, purple, and blue. In addition to the attractive colors, anthocyanins provide, there is also considerable recent interest in anthocyanins for their dietary value, acting as strong antioxidants and possibly anti-cancer agents.

Effects of mutations at different steps in anthocyanin biosynthesis.

Of course, each enzyme in the anthocyanin pathway is encoded by a gene. As such, mutations can generate variation or disrupt the pathway. The maize anthocyanin pathway and examples of mutants are shown in Figure 2. The ***a2*** mutant causes a deficiency in the anthocyanidin synthase enzyme, blocking the production of any pigmented compounds. The ***a1*** and ***c2*** mutants are similarly colorless. The ***bz2*** mutant is deficient in an enzyme related to glutathione-S-transferase, which is required for transport of anthocyanidins into the vacuole, resulting in the bronze-colored pigment due to the more alkaline environment of the cytosol. The ***bz1*** mutant (not shown) produces a similar phenotype. Mutation of the ***pr1*** gene eliminates just one class of anthocyanins. The ***pr1*** gene encodes flavanone 3′-hydroxylase, required for the production of cyanidin, which generates purple anthocyanin. In the mutant, only pelargonidin is synthesized, producing a red pigment.

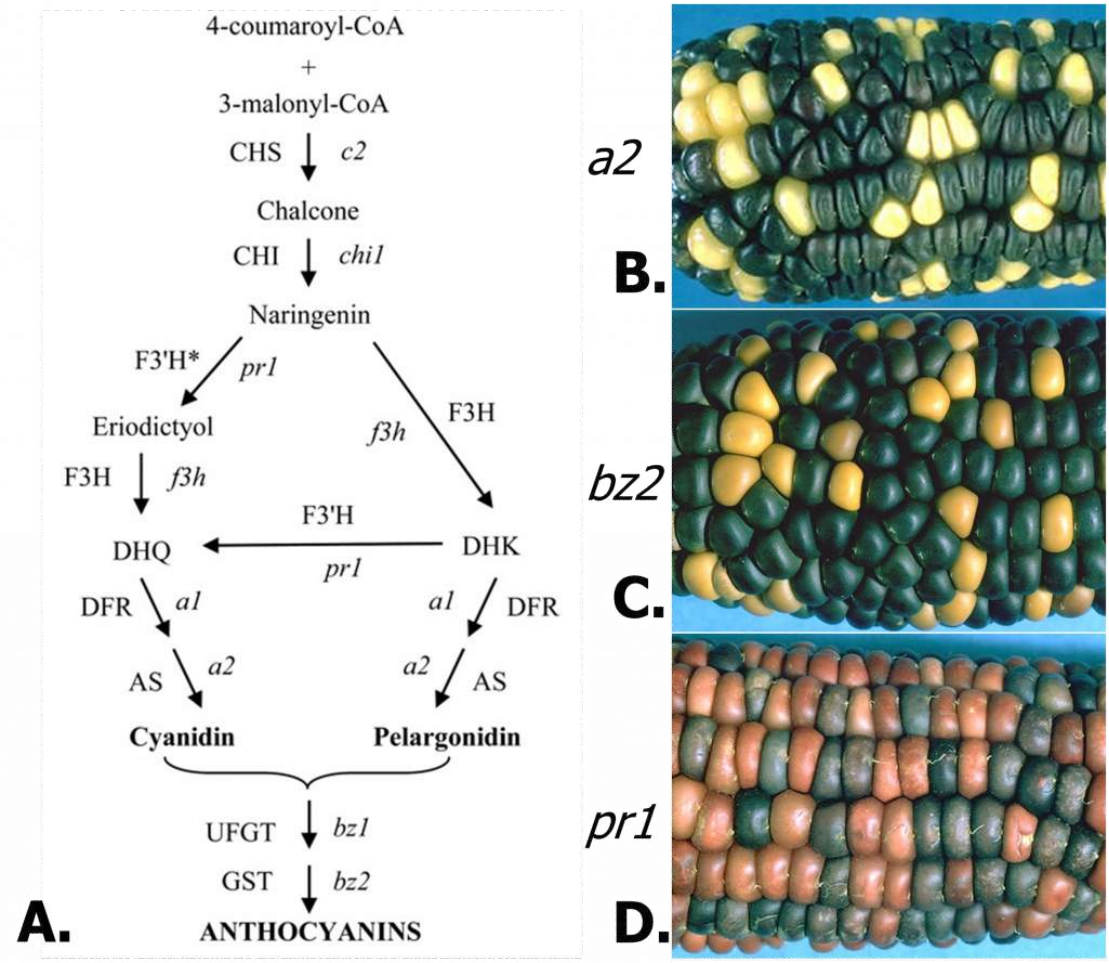


Figure 2. *Maize anthocyanin mutants. A. The maize anthocyanin pathway. (Modified from Sharma et al., 2011). B-D. Ears segregating a2 (anthocyaninless2), bz2 (bronze2), and pr1 (red aleurone1) mutant kernels, respectively. Normal kernels are dark purple, nearly black, and the mutant kernels are the lighter colors. Images in B-D courtesy of M.G. Neuffer and MaizeGDB.*

Regulatory pathways

A regulatory pathway refers to a situation where one gene or gene product controls the expression or activity of another gene or gene product, which may in turn also perform a regulatory function. As we discussed, gene function can be

regulated at many points from transcriptional regulation to post-translational regulation. We will examine a couple of specific examples of regulatory pathways but first let's consider some of the general features of regulatory pathways. Understanding the logic of regulatory pathways will become important later when we seek to use biotechnology to manipulate traits by altering their regulation.

A simplistic but useful way to help understand the logic of regulatory pathways is to consider them as a series of switches with on and off states.

Positive vs. negative regulators

A key feature of regulatory pathways is that regulators might function either positively or negatively. A positive regulator is one that functions to activate the next component in the pathway, whereas a negative regulator would inhibit the next component. For example, a transcription factor that activated the transcription of a gene would be considered a positive regulator, while one that repressed the transcription of a gene would be a negative regulator. Another common type of regulator is a protein kinase. Protein kinases are proteins that add a phosphate group to another protein. Phosphorylation is a common type of post-translational modification that can regulate the activity of proteins; sometimes the phosphorylated protein becomes activated and sometimes it becomes repressed.

A positive regulatory step is represented with an arrow. For example, "protein A activates protein B" would be represented like:

$$\mathrm{A} \rightarrow \mathrm{B}$$

A negative regulatory step is represented with a bar. "Protein C represses protein D" would be represented like:

$$\mathrm{C} \dashv \mathrm{D}$$

Active vs. inactive states of regulators

A second point to consider in the logic of regulatory pathways is the activity state of each component. A useful way to think about this is that each component can have an "on" or an "off" state. Of course, in reality, many proteins can have intermediate levels of activity but for the sake of simplicity, we will just consider the on and off states.

The effect of a regulator being in the on or off state depends on whether it is a positive or negative regulator.

If a positive regulator is in the on state, it will function to activate the next factor. In the example of $\mathrm{A} \rightarrow \mathrm{B}$, if **A** is in the on or active state, it will function to switch **B** to the on or active state.

If a negative regulator is in the on state, it will function to switch the next factor to the off or inactive state. For $\mathrm{C} \dashv \mathrm{D}$, if **C** is in the on or active state, it will switch **D** to the off or inactive state.

Regulatory pathways sometimes contain many, sometimes few, steps. They can also contain a mixture of positive and negative regulators. Let us explore a few hypothetical examples to see how the pieces fit together to regulate plant responses to stimuli.

When a pathway is depicted, all the steps are shown. For the sake of simplicity, it is assumed that the default activity state of any given component is such that the preceding step functions to change it. We will begin with a simple example containing several positive regulatory steps that activate a cellular activity, **Activity 1**, in response

to **Stimulus 1**. In the pathway shown below, the **Stimulus 1** activates **A**. Therefore, we assume that in the absence of this stimulus, **A** is in the inactive or off state. Since **A** is inactive, it is not functioning to activate **B**, which is therefore also in the inactive state, and likewise for **C**. Ultimately **Activity 1** does not occur in the absence of **Stimulus 1** but does occur in response to the stimulus.

Stimulus 1 → A → B → C → Activity 1

We can use color coding to help visualize this. Light grey represents the **OFF** state and blue represents the **ON** state.

No Stimulus 1 → A → B → C → Activity 1 (OFF)

Stimulus 1 → A → B → C → Activity 1 (ON)

Or we can make a table to help keep track of the states of each component in the presence or absence of stimulus.

Stimulus 1 →	A →	B →	C →	Activity 1
NO	OFF	OFF	OFF	NO
YES	ON	ON	ON	YES

Now let us look at another example containing both positive and negative regulators.

Stimulus 2 → D → E ⊣ F → Activity 2

In this case, components **D** and **E** would be in the inactive states in the absence of stimulus, because they would not have been activated. But **E** functions as a negative regulator of **F**. Since **E** is inactive, it is not functioning to inhibit **F**, which then remains in the active state promote cellular **Activity 2**. When **Stimulus 2** is present, **D** and **E** become activated, and **E** functions to inactivate **F**. Since **F** is off, **Activity 2** does not occur. Thus, the net response to **Stimulus 2** is the *repression* of **Activity 2**.

No Stimulus 2 → D → E ⊣ F → Activity 2

Stimulus 1 → D→ E ⊣ F → Activity 2

OR

Stimulus 2 →	D →	E ⊣	F →	Activity 2
NO	OFF	OFF	ON	YES
YES	ON	ON	OFF	NO

Now that we have considered regulatory pathways from a hypothetical perspective, let's consider a couple real ones. As mentioned, regulatory pathways can take on many forms. They can consist of a series of transcription factors that regulate the expression of one another's genes, ultimately resulting in the regulation of genes that effect some biological response. Here, we will look at a signal transduction pathway consisting of a series of post-translational modifications that occur response to a hormone stimulus and culminate in gene expression changes to effect a response.

Signal transduction pathway for hormone signaling

Brassinosteroids (BRs) are a class of plant steroid hormones that control many aspect of plant physiology. They are most

noted for their role in promoting plant growth and mutants deficient in BR biosynthesis or signaling show dwarfism. BRs also promote a range of other responses, including yield and increased stress resistance.

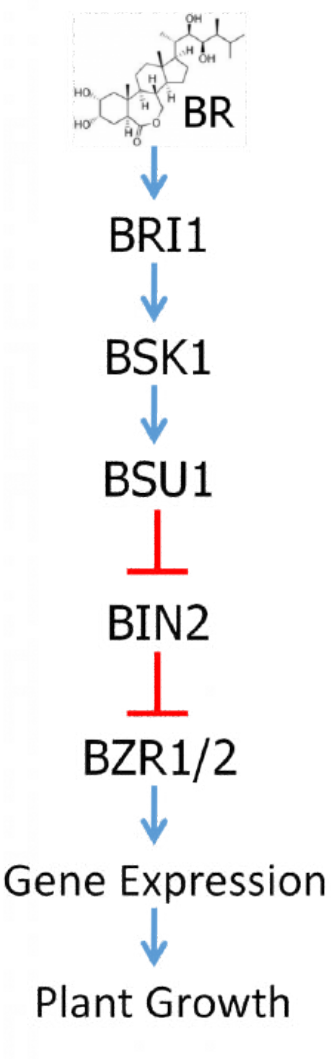


Figure 3. The brassinosteroid signaling pathway. Diagrammatic representation of the regulatory relationships among BR pathway components.

BR hormones, of which brassinolide (BL) is considered the most active, are recognized by a receptor called BRI1. BRI1 is a type of receptor known as a receptor kinase, which spans the cell membrane. On the outside of the cell is a receptor domain which specifically recognizes and binds to BR. There is also a membrane-spanning domain and then a protein kinase domain inside the cell. Binding to BR outside the cell then activates the protein kinase domain inside the cell. Protein kinases are enzymes that function to add phosphate groups to other proteins. As previously discussed, phosphorylation can alter the activity of a protein. Activation of the BRI1 receptor kinase then results in the activation of a signaling pathway that ultimately regulates the activity of transcription factors inside the nucleus, changing the expression of genes that regulate growth and stress resistance.

The BR signal transduction pathway is well studied and is summarized in Figure 3. Do not be concerned with memorizing the names of all these factors, but just focus on understanding the regulatory logic. The overall regulatory relationships of the pathway are depicted in Figure 3. The main target of regulation is a pair of closely related transcription factors called BZR1 and BZR2. BZR1/2 are negatively regulated by a protein called BIN2 in the absence of BR. BIN2 is a protein kinase that phosphorylates BZR1/2, which results in BZR1/2's exclusion from the nucleus and their proteolytic degradation. In the absence of BR, BIN2 contains a phosphate group that is required for its activity.

Binding of BR to the BRI1 receptor domain outside the cell results in the activation of the kinase domain inside the cell. Activated BRI1 then phosphorylates a protein called BSK. Phosphorylated BSK binds yet another protein called BSU1, which is a protein phosphatase. Protein phosphatases are enzymes that remove phosphate groups from other proteins. BSK binding activates BSU1 which then catalyzes the removal of the phosphate on BIN2, thereby inactivating BIN2. Thus, BSU1 is a negative regulator of BIN2. When BIN2 is inactivated, that allows BZR1/2 to accumulate, to enter the nucleus, and to regulate the expression of genes that effect the BR response, including promoting plant growth and yield.

Effects of mutation on different steps

Mutations can affect regulatory pathways in multiple ways. Most mutations are **loss-of-function** mutations where the function of the gene or gene product is partly or completely impaired. Considering the BR signaling pathway, the phenotypes of mutation will differ depending on whether the mutation affects a positive or negative regulator of the pathway. A loss-of-function mutation in the bri1 gene will block the perception of the hormone and therefore the pathway will not be activated; a dwarf phenotype will result. On the other hand, a loss-of-function in the bin2 gene will release BZR1/2 from regulation. The result will be constant BZR1/2 activity regardless of whether BR hormone is present. This might be expected to generate giant plants but in fact the effects of deregulating hormone responses are more complex—let us just call it unregulated growth.

Mutations can also be **gain-of-function**, where the gene product assumes a form that is constantly in the active state. Such mutations are typically dominant and are much less common than loss-of-function. Again, the outcome of a gain-of-function mutation on the BR signaling pathway will vary depending on whether it affects a positive or negative regulator of the pathway. A gain-of-function mutation in a positive regulator would cause deregulated hormone responses, whereas such a mutation in a negative regulator would permanently shut down the pathway and cause dwarfism.

Complementary gene action and genetic epistasis in the context of pathways

Considering that biological processes are controlled by multiple genes acting together, several genetic principles make more sense. First, it becomes clear how mutations in different genes can produce the same phenotypic effects. Since multiple genes are often required for a single biological process, disruption of different genes in the process might have the same net effect. From this follows the concept of **complementary gene action**. Mutant plants with the same phenotype are crossed together and the offspring is normal looking. The mutations in the parent plants likely affected different genes in the same pathway. Finally, the concept of **epistasis** can be best understood in the context of pathways. Note that the term epistasis is used differently by different geneticists. In quantitative genetics, epistasis refers to any gene interaction. Here, the term epistasis refers to a specific type of gene interaction where the action of one gene is masked by the action of another gene. In other words, if two mutations are combined in an individual, only one of the phenotypes is apparent. The mutant whose phenotype is apparent is said to be epistatic to the one that is masked.

Complementary gene action in a regulatory pathway

The concept of complementary gene action is no different in a regulatory pathway than in a biosynthetic pathway. Mutations in different genes that have similar roles in the pathway (e.g., are both positive regulators of the pathway) would be expected to cause similar mutant phenotypes. If such mutants in different genes were crossed together, the F1 progeny would be normal because both genes would be heterozygous and therefore a functional copy of both factors would be present.

Epistasis in a regulatory pathway

Epistasis can be quite complex, especially in regulatory pathways. Unlike complementary gene action, epistasis is not restricted to recessive loss-of-function mutations. The situation is further complicated by the presence of positive and negative regulators in a pathway. In contrast to a biosynthetic pathway, the rule of thumb for a regulatory pathway is that the more "downstream" gene is epistatic to the more "upstream" one. Let's turn again to the BR signaling pathway and look at a couple of examples to understand why.

First consider loss-of-function mutations in the *bri1* and *bin2* genes, which in single mutants cause a dwarf phenotype and an unregulated growth phenotype, respectively. If the two mutations are combined into a double mutant plant, what do we expect? The *bri1* mutations disrupt the receptor function required to initiate signaling through the pathway. However, BIN2 functions to inhibit the activity of pathway by repressing the function of BZR1/2. If BIN2 is rendered non-functional, then BZR1/2 will be active and promote the growth response. This will be true even if previous steps are blocked by the *bri1* mutation. So the double mutant would show the *bin2* unregulated growth phenotype making *bin2* epistatic to *bri1*.

What about a gain-of-function mutation in *BZR1* or *BZR2* combined with a loss of function mutation in *bri1*? A gain-of-function mutation would render *BZR1/2* constitutively active, regardless of the activity state of BIN2, resulting in the unregulated growth phenotype. Since BIN2 activity would be irrelevant, upstream activities would also be irrelevant. Thus, the double mutant would show unregulated growth and *BZR1/2* would be epistatic to *bri1*.

Lesson Summary

We saw that genes do not act alone but function in concert with other genes to perform various biological functions such as the biosynthesis of a compound or the regulation of a response to a stimulus. When genes control a sequential series of steps, that is called a pathway. But pathways typically branch and intersect with other pathways to form networks. Mutations affect the activity of pathways in various ways, depending on the nature of the mutation (loss- vs. gain-of-function) and the function of the gene product (positive vs. negative effector).

When mutations affect different steps in a pathway, they can often have a similar overall effect on the pathway and therefore lead to similar phenotypic consequences. Such mutations will complement one another even though their phenotypes are identical. It is also possible that mutations in different steps of a pathway can cause different phenotypes. When such mutations are combined into a double mutant individual, the phenotype of one is often masked by the phenotype of the other epistatic mutation. When pathways are linear, the "upstream" mutation is generally epistatic in biosynthetic pathways and in regulatory pathways it is generally the "downstream" mutation that is epistatic. In branched pathways or networks, it is more difficult to generalize.

Complementary gene action is important for geneticists trying to identify all the steps in a pathway. This is the basis of the classic complementation test used to determine if mutations are allelic or affect independent genes. Epistasis is also useful to geneticists who want to understand whether mutations with different phenotypes affect the same pathway, and if so, to determine the order of gene action.

When biotechnology is used to manipulate a trait, it is really the activity of a pathway or network that is being targeted. The same genetic principles outlined in this section are used to design biotechnological strategies to alter pathway activities so as to produce the desired outcome on the trait of interest.

Activity 1

To answer the following questions, refer to the pathways illustrated in Figure 3. They are labeled slightly differently but a1 encodes DFR, which catalyzes the production of leucoanthocyanidins and a2 encodes AS or LDOX/ANS, which subsequently catalyzes the production of anthocyanidins. Assume that normal, wild-type maize kernels are the deep purple/black color as in Figure 3.

1. What would you predict happens to the levels of leucoanthocyanidins in an *a2* mutant compared to normal?

a. no change
b. increase
c. decrease
d. can't predict

Show Answer
Answer: b or d.

The expectation would be b, that leucoanthocyanidins would increase. Because the pathway is blocked at the subsequent step, leucoanthocyanidins cannot be converted to anthocyanidins. But since all the preceding enzymes remain functional, those reactions would be expected to continue resulting in the accumulation of leucoanthocyanidins.

d is also a correct response. Biosynthetic pathways do not always behave as expected. For example, there could be feedback inhibition of one or more enzymes by intermediate compounds. Bottom line, we won't know for certain until the levels are measured experimentally.

2. What would you predict happens to the levels of pelargonidin in an *pr1* mutant compared to normal?

a. no change
b. increase
c. decrease
d. can't predict

Show Answer
Answer: b or d.

Again, the expectation would be b, that pelargonidin would increase. Normally, some fraction of the intermediate compounds are channeled through the pathway branch leading to cyaniding. Because the pathway leading to cyanidin is blocked it might be expected that all the intermediates would be converted to pelargonidin, causing an increased accumulation.

d is also a correct response. As discussed in problem 1, biosynthetic pathways do not always behave as expected. Feedback inhibition could be a factor or it may be possible that the enzymatic capacity of the pelargonidin branch of the pathway is already saturated. To increase flux through the pathway, it could be required to increase the levels or activities of one or more rate-limiting enzymes. Again, we won't know for certain until the levels are measured experimentally.

Activity 2

Predict the outcomes of the following mutations on the signaling output, or plant response for the BR signaling pathway.

1. A gain-of-function mutation in the BZR2 gene

a. increased output
b. no change
c. decreased output
d. can't tell

Show Answer
Answer: a. Since BZR2 is a positive regulator of BR signaling, increased output from the pathway would be expected from a gain-of-function mutation.

2. A loss-of-function mutation in the bsu1 gene. (Study the pathway and think about this one!)

a. increased output
b. no change
c. decreased output
d. can't tell

Show Answer
Answer: c. The output from the pathway would decrease causing decreased plant growth. Even though BSU1 is a negative regulator, its target is BIN2 which is another negative regulator. Without BSU1 function, BIN2 will always be in the ON state and inhibit the activities of BZR1/2, therefore decreasing the response to BRs.

Activity 3

1. Predict the phenotype of a BSK gain-of-function.

a. unregulated growth
b. no effect
c. dwarf
d. can't predict

Show Answer
Answer: a. BSK is a positive regulator of the pathway that functions to activate BSU1. Therefore a gain-of-function would cause it to positively activate the pathway, even in the absence of BR hormone, causing unregulated growth.

2. Now predict the phenotype of a double mutant between the BSK gain-of-function and a BIN2 gain-of-function.

a. unregulated growth
b. no effect
c. dwarf
d. can't predict

Show Answer
Answer: c. The gain-of-function BIN2 mutation locks BIN2 in the ON state where it would function to inhibit the activity of BZR1/2. It would be impervious to regulation by the upstream BSU1, so even if BSU1 were continually active due to the gain-of-function BSK, the BIN2 mutant would inhibit growth and therefore be epistatic.

Activity 4

Now you try a couple. Fill in each of the following tables to determine whether the hypothetical Activity will be activated or repressed in response to each Stimulus.

References

Sharma, M., Cortes-Cruz, M., Ahern, K.R., McMullen, M., Brutnell, T.P., and Chopra, S. (2011). Identification of the Pr1 Gene Product Completes the Anthocyanin Biosynthesis Pathway of Maize. Genetics 188, 69-79.

11. Recombinant DNA Technology

WALTER SUZA; DONALD LEE; MARJORIE HANNEMAN; AND PATRICIA HAIN

Learning Objectives

1. Understand the importance of recombinant DNA technology.
2. Learn isolation of DNA and its separation on an agarose gel.
3. Understand restriction and ligase enzymes and their application in gene cloning.
4. Understand vectors and their application in gene cloning and expression.
5. Understand polymerase chain termination reaction (PCR).

Introduction

Recombinant DNA (rDNA) technology has resulted in breakthroughs in crop and animal biotechnology. The power of rDNA technology comes from our ability to study and modify gene function by manipulating genes and transform them into cells of plant and animals. To arrive at this several tools of molecular biology are used including, DNA isolation and analysis, molecular cloning, quantification of gene expression, determination of gene copy number, transformation of the appropriate host for replication or transfer into crop plants and analyses of transgenic plants.

Definition and background

Recombinant rDNA technology involves procedures for analyzing or combining DNA fragments from one or several organisms (Figure 1) including the introduction of the rDNA molecule into a cell for its replication, or integration into the genome of the target cell.

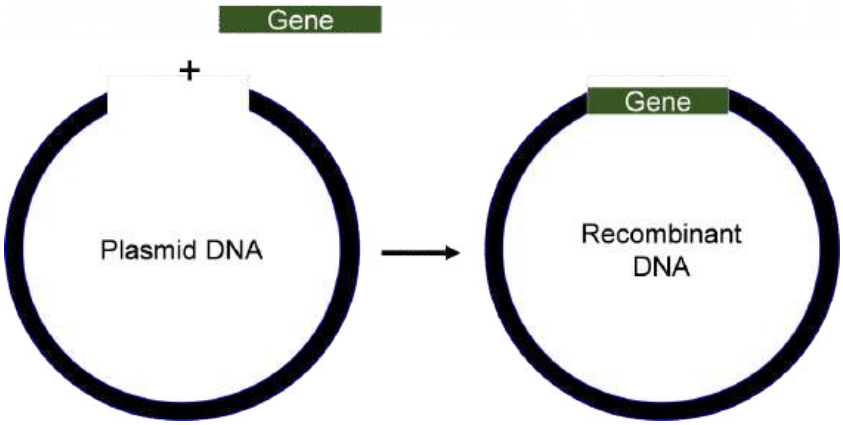


Figure 1. *Recombinant DNA is made from combining DNA from different sources. Image by Walter Suza.*

Advances in molecular biology in the early 1970s, including the success in creating, and transferring DNA molecules into cells, revolutionized both science and industry. The first genetically modified organisms were bacteria that made simple proteins of pharmaceutical interest, for example, insulin. As the technologies improved, other organisms including plants became amenable for improvement by rDNA technology. Table 1 provides important milestones in the development and application of rDNA technology.

Table 1.Events related to the development and application of recombinant DNA technology.

Event	Year
Mendel's experiments published	1866
DNA discovered in cell	1869
Mutation of genes by x-rays	1927
One gene-one enzyme hypothesis	1941
DNA is identified as the genetic material	1944
Structure of DNA determined	1953
Ribosomes synthesize protein	1954
Function of mRNA proposed	1961
Genetic code determined	1961-64
Isolation of a restriction enzyme	1970
Recombinant DNA techniques developed	Early 1970s
Isolation of a single copy gene from higher eukaryote	1977
Rapid method of DNA sequencing developed	1977
Plant transformation	1983
Field testing of transformed plants	ca. 1986
Release of engineered plants to general public in the US	1995-96

Transformation of cells with rDNA produces organisms called bioengineered or genetically modified organisms (GMOs). The GMOs contain new traits from another organism. The first GMOs were *Escherichia coli* cells that were transformed with genes from human to produce various proteins for pharmaceutical purposes.

Isolation of DNA, restriction digestions and its separation on an agarose gel:

Preparation of DNA

For recombinant DNA procedures to work, a pure DNA sample must be obtained. The challenge is that plant and animal cells produce numerous other compounds that often act as contaminants and may inhibit cloning or sequencing of the DNA. Also, tissues and organs from the same plant, or different plants often contain different composition of metabolites, for example, proteins, lipids, and carbohydrates. These compounds must be separated from the DNA during isolation. To achieve this, scientists take advantage of the chemical and physical properties of different molecules inside the cell. For example, DNA is negatively charged, making it soluble in aqueous solution. However, the polar sugar phosphate groups of the DNA are repelled by non-polar solutions. Therefore, the final step in many DNA purifications protocols involves precipitation using alcohol. Other compounds, for example, proteins can also be easily separated from DNA by altering the concentration of salt in the extraction buffer.

Digestion of DNA with restriction endonucleases

Restriction endonucleases are a group of enzymes derived (primarily) from bacteria. Although there are several different types of **restriction enzymes**, those most useful for rDNA technology recognize specific short sequences in DNA and cleave the DNA at that site to produce cohesive (sticky) or blunt-ended fragments (Figure 2).

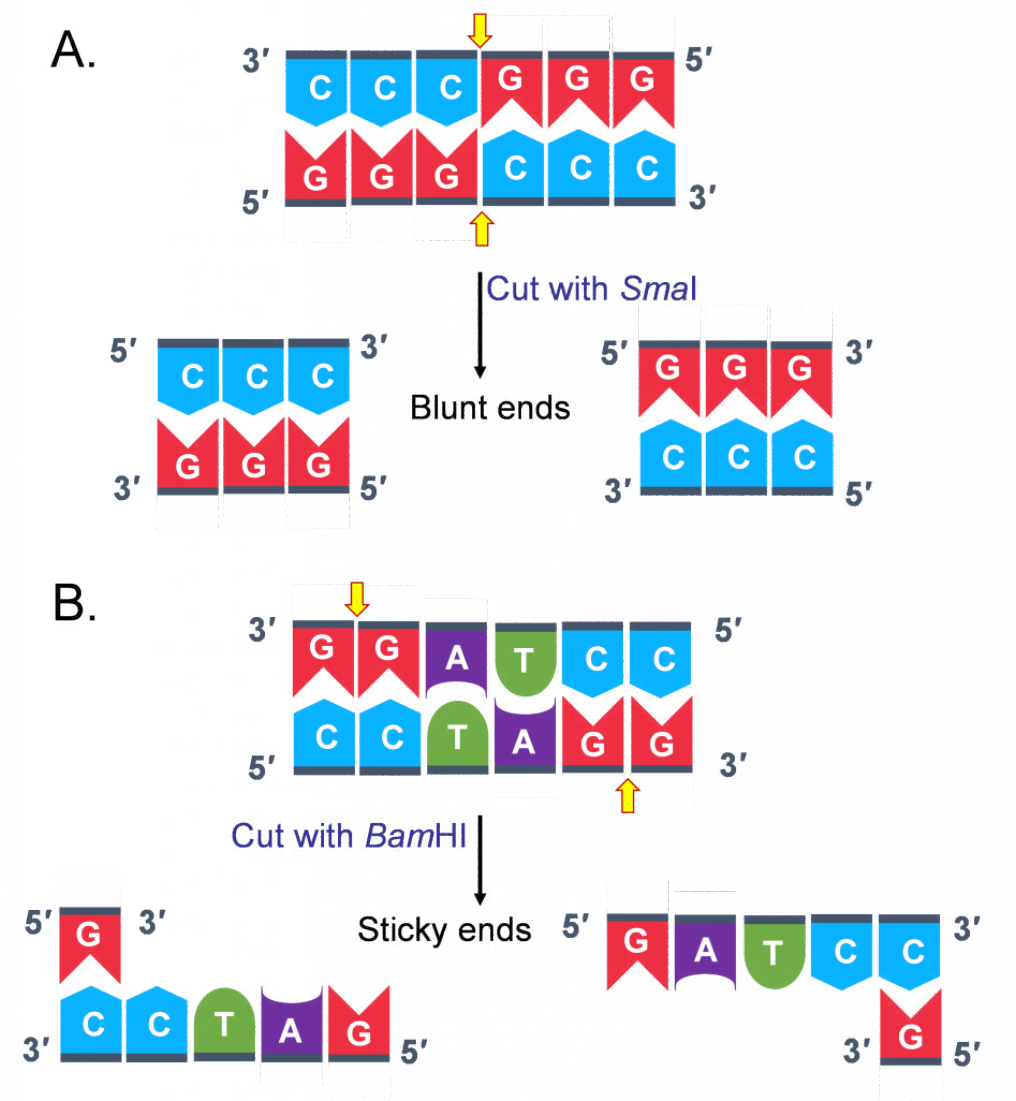


Figure 2. *Example of how restriction enzymes cut DNA. (A) Treating the DNA with SmaI results in fragments with blunt ends. (B) Whereas treatment with BamHI produces fragments with "cohesive" or "sticky" ends. Image by Walter Suza.*

More than 500 different restriction enzymes have been identified and can be purchased commercially. Thus, one may ask the question, how often does a restriction enzyme cut within a genomic sequence? It is not possible to give an exact answer for this question. However, let us assume that there are 4 bases on any strand of DNA. This means the probability of detecting an A (adenine) at a particular location is 1/4. Now, since most restriction enzymes recognize specific sequences of 6 bases long, the probability of finding such a site is (1/4)6 = 1 site in every 4,096 base pairs (bp). Assuming you have isolated genomic DNA from maize, and you want to digest it with a restriction enzyme that cuts every 4,096 (4,100) bp, how many fragments will you obtain? To answer this question, you need to have an idea of the size of the genome of the plant you are working with. The approximate size of the maize genome is 2,500,000,000 bp. Thus, using an enzyme that cuts every 4,100 nucleotides one would expect to obtain 2,500,000,000 bp/4,100 bp, approximately 610, 000 fragments. Note that nucleotide distribution is not always random and thus frequency may be different for given DNA. Also, methylation of specific bases in genomic DNA can prevent cleavage at some site.

Separation of digested genomic DNA on agarose gel

The only physical features of nucleic acid fragments that are routinely used for their characterization are their size and nucleotide sequence. Molecular weight of DNA is most conveniently evaluated by electrophoresis in agarose gels. Agarose forms a gel by hydrogen bonding when cooled from the melted state. This gel, interwoven network of agarose chains, interferes with the movement of DNA through the gel (see **Lesson on PCR and Gel Electrophoresis**). Pore size, which affects rate of movement of DNA fragments of a given size, depends on the concentration of agarose. The gel is submerged in electrolyte solution; sample is loaded into wells on one end and current is applied to facilitate movement of DNA fragments. Since DNA is negatively charged it will migrate in the electrical field. Fragments separate according to size. The distance of the migration in each time is proportional to 1/log MW. Following gel electrophoresis DNA can be visualized by staining with ethidium bromide (see **Lesson on PCR and Gel Electrophoresis**) or other DNA stains.

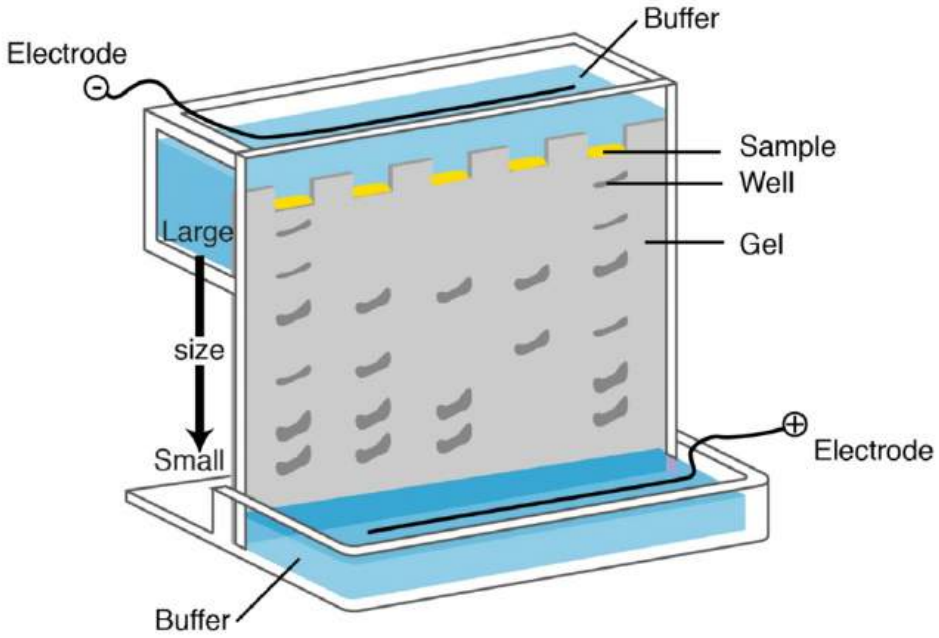


Figure 3. *Gel electrophoresis is used to separate nucleic acid (DNA and RNA) or protein molecules by their size. For DNA, the sample is placed in an electric field and the negatively charged DNA will migrate to the positive electrode. Migration speed depends on the size of the DNA fragment. Use of semisolid matrix such as agarose or polyacrylamide gel facilitates separation of the DNA molecules. Image from NIH-NHGRI.*

Tools for gene (DNA) cloning

Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a method by which millions of copies of a DNA fragment are produced in a test tube in a matter of minutes or a few hours. The basic steps in PCR reaction were discussed in the Lesson on PCR and Gel Electrophoresis.

Once the sequence of a particular gene is known, it becomes possible to use PCR to isolate that gene from any DNA sample. Recall in **Lesson on Gene Transcription** you learned that at the genomic level a gene is made up of regulatory

sequences, coding, and non-coding sequences. Thus, if the goal is to use PCR to isolate both regulatory and non-regulatory sequences of a gene, the approach would be to use DNA as the starting material.

In cloning genes by PCR, restriction enzyme sites are added at the 5′-end of the primers to facilitate cloning of PCR fragments. A few additional nucleotides (~6 nucleotides) added at the 5′-end of the restriction sites to facilitate restriction digestion of PCR products prior to cloning in a plasmid vector. Alternatively, PCR products may be cloned directly into a T-vector without restriction digestions in *E. coli* (read more about Promega cloning vector systems). After cloning into *E. coli*, the fragment is analyzed by sequencing and then sub-cloned into suitable vectors for expression studies.

Like all other biochemical processes, DNA synthesis by PCR is not a perfect process, and occasionally the polymerase enzyme will add an incorrect base to the growing DNA strand. In the context of DNA replication in a cell, the errors are corrected by the DNA polymerase, this is called "proofreading". Commercially available polymerases may or may not have proofreading capability.

Another important consideration in PCR analysis is contamination.

Minor contamination of the starting material can have serious consequences. Recall that minute amounts of starting DNA can be amplified to millions of copies through PCR. If one inadvertently (or carelessly) mixed DNA from two different sources, the results will be confounding making it impossible to distinguish lines and may cost a laboratory time and money. Ensure that proper procedures are followed in preparing PCR assays. One common source of contamination in plant biology is the products from previous amplification processes. A completed PCR reaction will contain millions of copies of amplified fragments so that even a minute droplet or aerosol from a pipette tip will contain an enormous number of amplifiable molecules. It is always essential to run negative controls which will reveal the presence of contaminating DNA in your PCR assays.

Cloning vector definition and requirements

A cloning vector is a specialized DNA sequence that can enter a living cell and provide means for detection of its presence to a researcher by conferring a selectable property on the host cell (e.g., resistance to antibiotics), and possess means for self-replication. A vector must also possess easily distinguishable physical traits, such as size, or shape, to allow purification away from the host cell's genome.

Ligase enzyme and gene cloning

Cutting and joining together of vector and DNA fragments from different origins results in rDNA. Recall that restriction endonucleases are used to cut DNA. To join DNA molecules together, an enzyme called DNA ligase is used. The enzyme DNA ligase is used to seal together restriction fragments by forming new phosphodiester bonds. The ligated vector and DNA fragment can now be transformed into a host cell for replication and expression.

The transformation of *E. coli* takes several steps. First, a gene of interest is inserted into a plasmid that contains a selectable marker usually encoding for resistance to an antibiotic. Second, the plasmid construct containing the gene of interest is transformed into bacterial cells by briefly exposing the mixture of ligated plasmid-DNA fragment (rDNA molecule) and bacterial cells to cold (0^{o}C) and heat (37-42^{o}C). The next step is to grow the transformed cells on selection media containing an antibiotic. Only the cells that have been transformed with the plasmid containing the gene of interest and the marker for resistance to the antibiotic will survive. In addition to using an antibiotic, plasmid vector

systems that contain the *lacZ* gene encoding β-galactosidase allow for easier selection of positive colonies that may harbor the rDNA molecule of interest.

Types of cloning vectors

An example of a cloning vector is a **plasmid** (Figure 4), defined as an autonomously replicating extra chromosomal circular DNA which is faithfully passed on to progeny. Plasmids are double stranded circular DNA and range in size from about 1 kb – 200 kb. The most useful for cloning are 2 – 10 kb because smaller plasmids are easier to manipulate and usually produce higher copy numbers when grown in bacterial host cells.

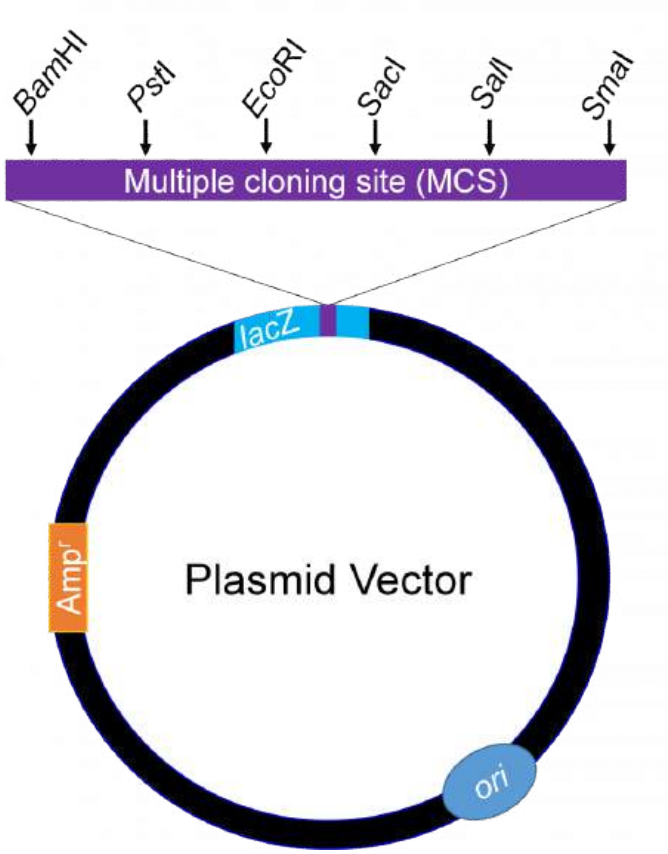


Figure 4. *Basic map of a plasmid vector. The MCS contains several restriction enzymes sites and few are shown here as example. Image by Walter Suza.*

Generally, no more than 10 kb is cloned into plasmids. For cloning large DNA fragments with high efficiency, a vector called **bacteriophage lambda** is used. Large chromosomal DNA fragments close to 23 kb are stable when introduced into a lambda phage vector.

mRNA as starting material for gene cloning

Recall in the **Lesson on Transcription** you learned about synthesis of RNA from DNA through the process of **transcription**. Transcription is an important step in gene expression. The mRNA produced can be isolated and "copied" back to DNA by a process called **reverse-transcription** (Figure 5). The first step of reverse transcription mimics

a strategy used by retroviruses (e.g., HIV) that have RNA genomes. As part of their gene-transmission package, retroviruses also contain an enzyme called RNA-dependent DNA Polymerases, commonly referred to as **reverse transcriptase**. After infecting a host cell the retrovirus uses its reverse transcriptase to copy its single stranded RNA genome into a strand of complementary DNA (cDNA). The reverse transcriptase then synthesizes the second DNA strand from the first strand to make a double stranded-DNA copy which integrates into the host genome.

If only a gene's coding sequence is required, isolating the gene from cDNAs would be the strategy. It is important that cDNAs are synthesized from tissues expressing the gene of interest. Thus, prior knowledge of where the gene is functional is important in constructing the cDNA molecules for cloning the gene of interest.

The cDNAs produced *in vitro* (Figure 4) can be used for PCR analysis, similar to chromosomal DNA. The combination of reverse-transcription and PCR (RT-PCR) is a valuable tool in gene cloning and quantification of mRNA.

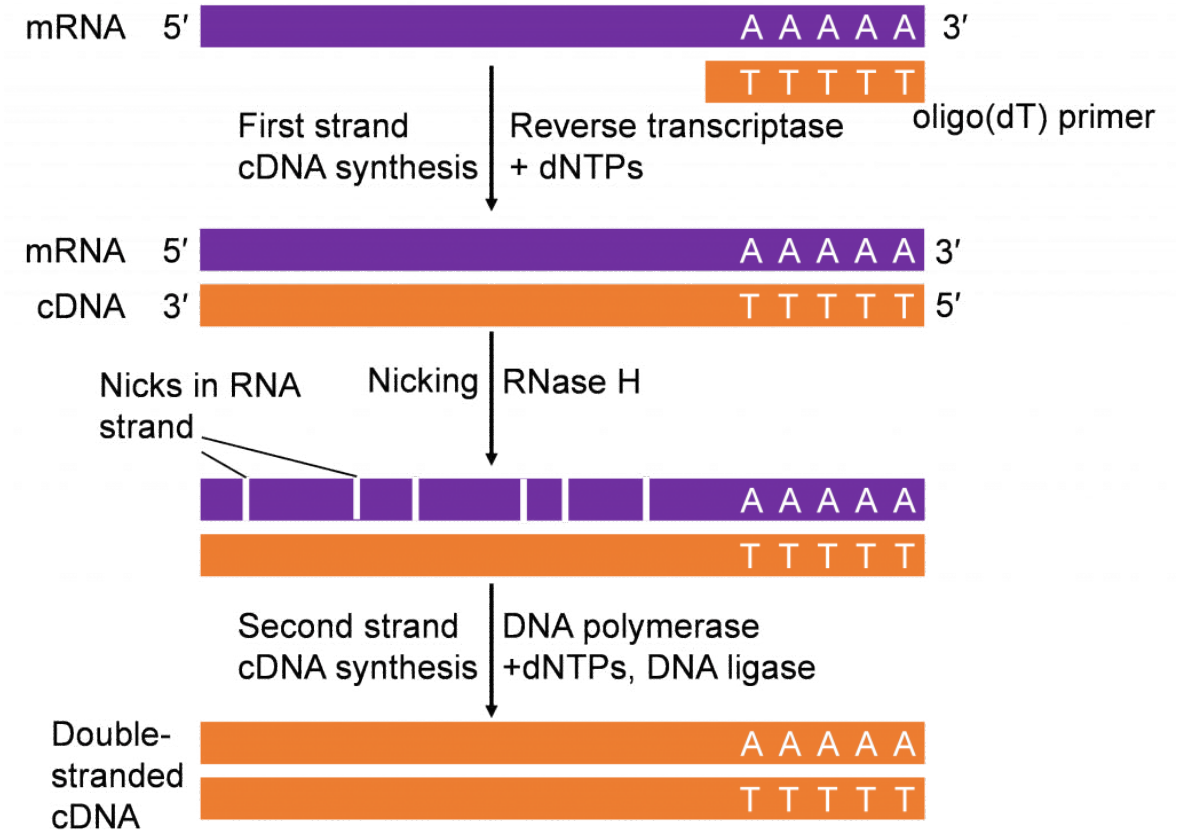


***Figure 5**. A population of mRNA isolated from plant tissues is combined with oligo(dT) primers that anneal to the poly(A) tail of the mRNAs to initiate the reverse transcription of cDNA from mRNA template with dNTPs. The result is hybrid molecules (mRNA-DNA). Treatment with RNAse H causes degradation of the mRNA leaving an intact single stranded DNA (first strand). DNA polymerase synthesizes the second strand by adding complementary dNTPs to the growing chain. The reagents for RT usually come in pre-mixed kits making such an assay easy to carry out routinely in most laboratories. Most mRNAs from cells in tissues analyzed are converted to cDNAs. Thus, using primers specific to a sequence, that double stranded cDNA can be amplified by PCR for various purposes for example, gene cloning. Image by Walter Suza.*

Lesson Summary

Recombinant DNA technology has contributed significantly to development of agricultural biotechnology. Transformation of cells with rDNA produces organisms called bioengineered or genetically modified organisms. The tools for plant rDNA technology include, vectors, restriction enzyme, ligation enzymes, bacterial hosts, methods to isolate and multiply nucleic acids, methods to quantify nucleic acids, Agrobacterium as a vector to insert foreign DNA into plants.

Lesson Activities

Activity 1

Below is a hypothetical DNA fragment containing restriction sites for EcoRI and BamHI.

1. How many fragments will be produced after cutting the DNA with BamHI? What size will these fragments be in bp?
2. How many fragments will be produced after cutting the DNA with both BamHI and EcoRI at the same time? What size will these fragments be in bp?
3. How many fragments will be produced after cutting the DNA with EcoRI? What size will these fragments be in bp?

Activity 2

1. You need to clone an EcoRI fragment in the EcoRI site of an expression plasmid vector in E coli. The fragment is 1,809 bp long. There is a BamHI site in the fragment at 1,204 bp. In the plasmid vector, there is a single BamHI site at the promoter region for expression of the gene in E. coli. The BamHI site is 100 bp upstream of EcoRI cloning site.

Following cloning the EcoRI fragment you will get two classes of clones, only one of which will produce the protein. Describe the approach you will take to identify the correct class of clones for expression of the gene in E. coli.

12. Genetic Engineering

WALTER SUZA; DONALD LEE; MARJORIE HANNEMAN; AND PATRICIA HAIN

Learning Objectives

- Define genetic engineering.
- List and briefly explain the five basic steps in genetic engineering. Describe why each is necessary.
- Identify the fundamental differences between genetically engineered crops and non-genetically engineered crops.
- Explain the limitations to traditional breeding that are overcome by genetic engineering.
- Identify the approximate length of time required to obtain a marketable transgenic crop line (complete the entire crop genetic engineering process).

Introduction

The production of genetically engineered plants became possible after Bob Fraley and others succeeded to use Agrobacterium tumefaciens to transform plant cells with **recombinant DNA** in the early 1980s (Vasil, 2008a). Since this breakthrough in plant biotechnology, GM crops are now routinely developed and grown in many parts of the globe. Current statistics on adoption of genetically engineered crops in the U.S. can be found on the USDA Economic Research Service's website.

Genetic engineering has been used successfully to develop novel genes of economic importance that can be used to improve the genetics of crop plants. Genetic engineering is the targeted addition of a foreign gene or genes into the genome of an organism. The genes may be isolated from one organism and transferred to another or may be genes of one species that are modified and reinserted into the same species. The new genes, commonly referred to as transgenes, are inserted into a plant by a process called transformation. The inserted gene holds information that will give the organism a trait (Figure 1).

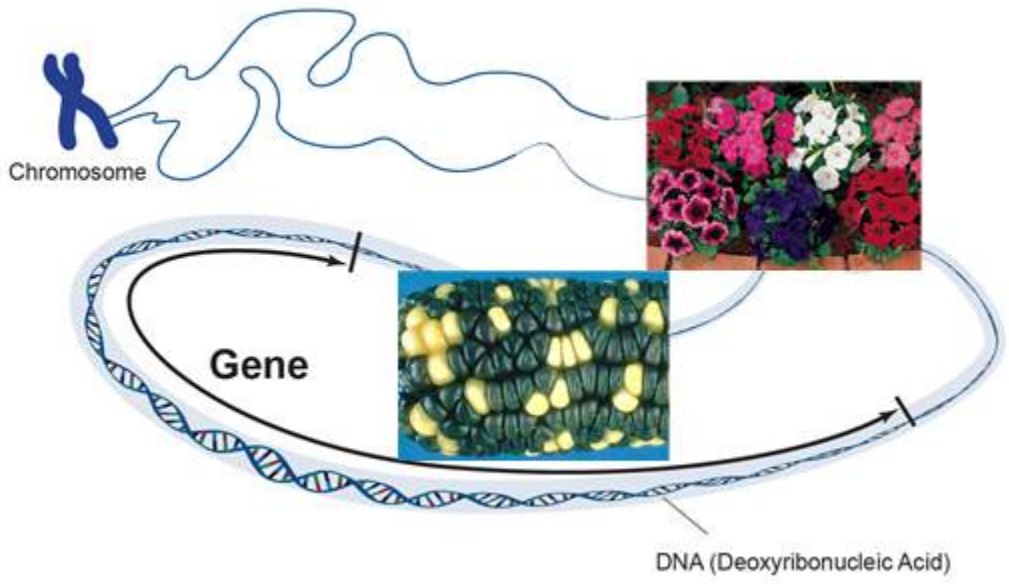


Figure 1. *Traits, such as flower and seed color, are controlled by DNA. Adapted from NIH-NGHRI.*

Crop genetic improvement (plant breeding) is an important tool but has limitations. First, in conventional terms, genetic improvement can only be done between two plants that can sexually mate with each other. This limits the new traits that can be added to those that already exist in that species. Second, when plants are mated, (crossed), many traits are transferred along with the trait of interest including traits with undesirable effects on yield potential.

Genetic engineering, on the other hand, is not bound by these limitations. It physically removes the DNA from one organism and transfers the gene(s) for one or a few traits into another. Since crossing is not necessary, the 'sexual' barrier between species is overcome. Therefore, traits from any living organism can be transferred into a plant. This method is more specific in that a single trait can be added to a plant.

The overall process of genetic engineering. A basic explanation of the five steps for genetically engineering a crop is provided. The five steps are:

1. Locating an organism with a specific trait and extracting its DNA.
2. Cloning a gene that controls the trait.
3. Designing a gene to express in a specific way.
4. Transformation, inserting the gene into the cells of a crop plant.
5. Cross the transgene into an elite background.

Step 1: DNA Extraction

The process of genetic engineering requires the successful completion of a series of five steps and discoveries. To better understand each of these, the development of Bt maize will be used as an example.

Before the genetic engineering process can begin, a living organism that exhibits the desired trait must be discovered. The trait for Bt maize (resistance to European corn borer) was discovered around 100 years ago. Silkworm farmers in the Orient had noticed that populations of silkworms were dying. Scientists discovered that a naturally occurring soil

bacteria was causing the silkworm deaths. These soil bacteria, called Bacillus thuringiensis, or Bt for short, produced a protein that was toxic to silkworms, the Bt protein.

Although the scientists did not know it, they had made one of the first discoveries necessary in the process of making Bt corn. The same Bt protein found to be toxic to silkworms is also toxic to European corn borer because both insects belong to the Lepidoptera order. The production of the Bt protein in the bacteria is controlled by the bacteria's genes.

To be able to work with the gene responsible for making the Bt toxin, scientists must extract DNA from the Bt bacteria (Figure 2). This is accomplished by taking a sample of bacteria containing the gene of interest and taking it through a series of steps that separate the DNA from the other parts of a cell.

Step 2: Gene Cloning

The second step of the genetic engineering process is gene cloning. During DNA extraction, all the DNA from the organism is extracted at once. This means the sample of DNA extracted from the *Bacillus thuringiensis* bacteria will contain the gene for the Bt protein, but also all the other bacterium's genes. Scientists use **gene cloning** to separate the single gene of interest from the rest of the DNA extracted (Figure 2).

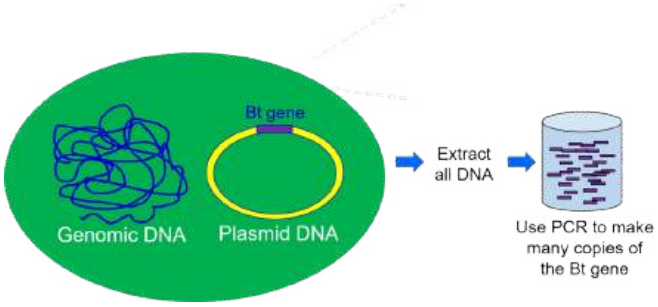


Figure 2. *Using DNA from B. thuringiensis to clone the Bt gene. Image by Walter Suza.*

The next stages of **genetic engineering** will involve further study and experimentation with this gene. To do that, a scientist needs to have thousands of exact copies of it. This copying is also done during the gene cloning step.

Step 3: Gene Design

Gene design relies upon another major discovery. This was the 'One gene One enzyme' Theory first proposed by George W. Beadle and Edward L. Tatum in the 1940's. Discoveries made during their research laid the groundwork for the theory that a single gene stores the information that directs the cell in how to produce a single enzyme **(protein**). Therefore, there is a single gene that controls the production of the Bt protein. It is called the Bt gene.

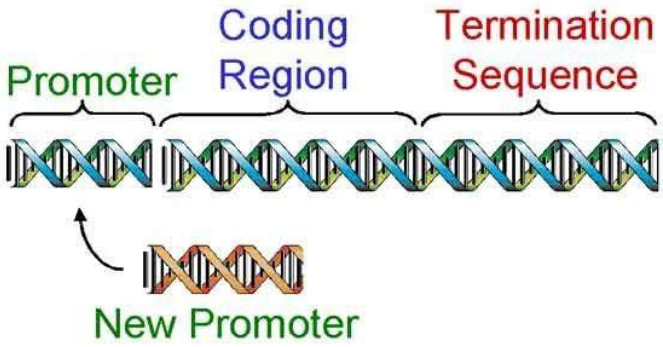


Figure 3. *Replacing existing promoter with new promoter. Image by Patty Hain.*

Once a gene has been cloned (Figure 2), genetic engineers begin the third step, designing the gene to work once inside a different organism. This is done in a test tube by cutting the gene apart with **restriction enzymes** and replacing certain regions (Figure 3).

Scientists replaced the bacterial gene promoter with promoters turn on the Bt gene in selected parts of the plant or promoters that can always turn on the **Bt gene** in all tissues. As a result, the first Bt gene released was designed to produce a level of **Bt protein** lethal to European corn borer and to only produce

the Bt protein in green tissues of the corn plant, (stems, leaves, etc.). Later, Bt genes were designed to produce the lethal level of protein in all tissues of a corn plant, (leaves, stems, tassel, ear, roots, etc.).

Plant transformation and tissue culture

The process of transformation involves the insertion of the desired **transgene construct** (Figure 5) into cells of the recipient plant species. In this process, scientists isolate tissue or cells from the cultivar they wish to transform and use one of several methods to insert the transgene into the tissue or cells. The transgene construct contains the following key features.

a. A *promoter that acts to turn the gene on and off in the cell.* The CaMV 35s promoter from the cauliflower mosaic virus (CaMV) is commonly used in genetic engineering. Other types of promoters, such as, the nopaline synthase promoter (NOS-Pro) also may be used to express transgenes in plant tissues.
b. A *selectable marker that is used to select cells that successfully obtained the construct during the transformation process.* In figure 4, the selectable marker in the construct is NPT II (Kanr) that controls resistance to the antibiotic kanamycin. The cells of the plant used for transformation will be grown on a media containing the antibiotic. Other selectable markers that have been used successfully in plants include genes controlling herbicide resistance.
c. A *terminator sequence*, such as the nopaline synthase (NOS) is included to mark the end of the transgene sequence for proper expression in plant cells.

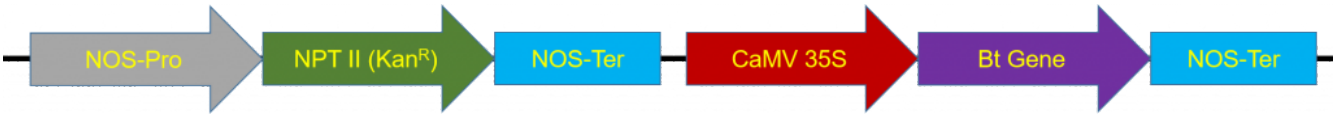


Figure 4. *The basic elements of a transgene construct. Image by Walter Suza.*

Two commonly used transformation methods include Agrobacterium tumefaciens-mediated transformation and biolistics transformation (aka **gene gun**), commonly referred to as particle bombardment (Figure 5). The biolistics method involves the use of high pressure to propel tungsten or gold beads coated with DNA of the gene construct into plant cells.

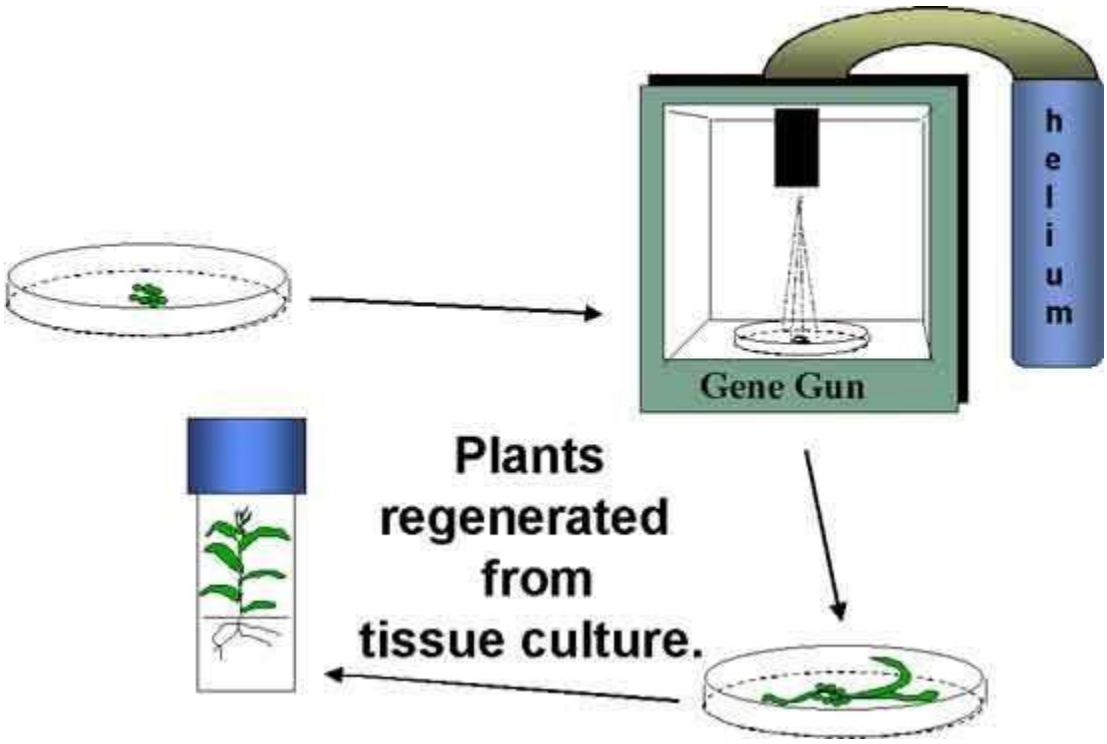


Figure 5. *Using the gene gun method to transform plant cells. Image by Patty Hain.*

Agrobacterium-mediated plant transformation

Crown galls are tumors of plants that arise at the site of infection by some species of the Agrobacterium. *Agrobacteria* do not enter the plant cells but transfer a DNA segment called T-DNA from their circular extra chromosomal *tumor-inducing* (Ti) plasmid into the genome of the host cells. Ti plasmids are maintained in *Agrobacteria* because a part of their T-DNA contains genes that encode unusual amino acids used by Agrobacterium. The T-DNA also encodes genes that affect host plant hormone physiology resulting in induced growth of the infected cells and tumor formation. Scientists took advantage of Agrobacterium's ability to stably integrate its T-DNA into the plant genome for introducing rDNA into plant cells. They first removed the genes that cause tumor or crown gall disease in plants from the T-DNA and engineered the plasmid for replication in both *Escherichia coli* and Agrobacterium cells. The initial replication of the construct in *E. coli* is useful for verifying the presence of the cloned gene and increasing the quantity of construct DNA for subsequent uses, including sequencing and transformation into Agrobacterium.

The steps in Agrobacterium-mediated transformation of plants are described in Figure 6.

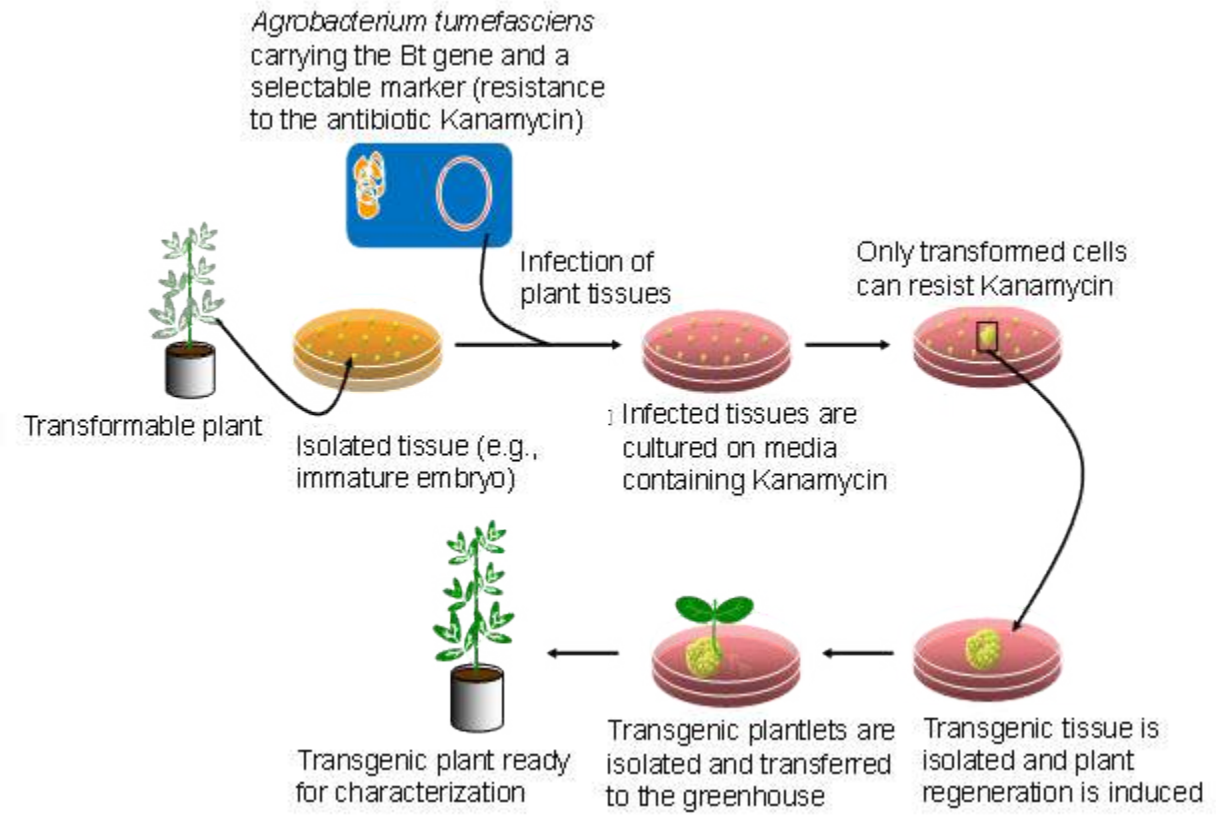


***Figure 6**. Transferring genes into plant cells by the Agrobacterium method. After infection of select tissue with the bacterium carrying a transgene construct with an antibiotic resistance gene as the selectable marker, the tissue is grown in a medium that contains the antibiotic that will kill all untransformed tissues or cells. Therefore, only tissues whose cells have been transformed with the transgene construct survive in the presence of the antibiotic. The surviving tissue is removed from the antibiotic and allowed to regenerate into whole plants. Normally, transgenic plants will be monitored in a controlled environment such as a growth chamber or greenhouse before they are grown in the field. Image by Kan Wang.*

At present, very few host cells receive the construct during the transformation process. Each random insertion of the construct into the genome of plant cells is referred as an event. Useful events are rare because of the random nature of the transformation process. Selectable markers are very important because they allow the identification of the rare events (Figure 7). Scientists must screen many potential transformants to identify events that are useful for breeding.

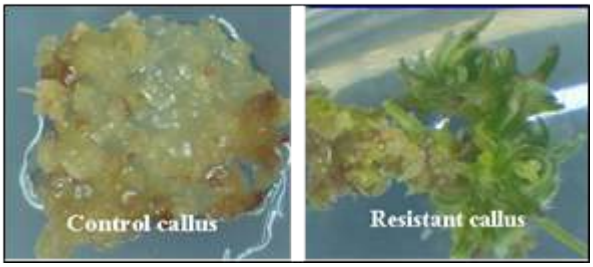


***Figure 7**. Selection for herbicide tolerance in buffalograss transformed with the gene that makes plants tolerant to glyphosate. The control calli lacking the glyphosate tolerance gene are killed by the herbicide that is part of the media on which the cells are grown (left panel). The calli in the right panel grew from a cell that received the glyphosate resistance gene during transformation and survived on the media. Image by Shui-Zhang Fei.*

From there, the new DNA may or may not be successfully inserted into a chromosome. The cells that do receive the new gene are called **transgenic** and are selected from those that are not transgenic (Figure 7). Many types of plant cells are totipotent meaning a single plant cell can develop into an entire plant. Therefore, each transgenic cell can then develop into an entire plant which has the transgene in every cell. The transgenic plants are grown to maturity in greenhouses and the seed they produce, which has inherited the transgene, is collected. The genetic engineer's job is now complete. He/she will hand the transgenic seeds over to a plant breeder who is responsible for the final step.

Inheritance of a transgene in plants

Transformation is successful when a transgene is incorporated into one of the chromosomes. The cells that have only one copy of the transgene in their genomes are said to be hemizygous (hemi = half, zygous = zygote). Because the segregation in the progeny of a hemizygous plant is the same as for a heterozygous plant, the term heterozygous will be used in this course when referring to a plant that is not homozygous for the transgene. The trait will segregate in the progeny in the same manner as any other gene in the plant as illustrated below (Figure 8).

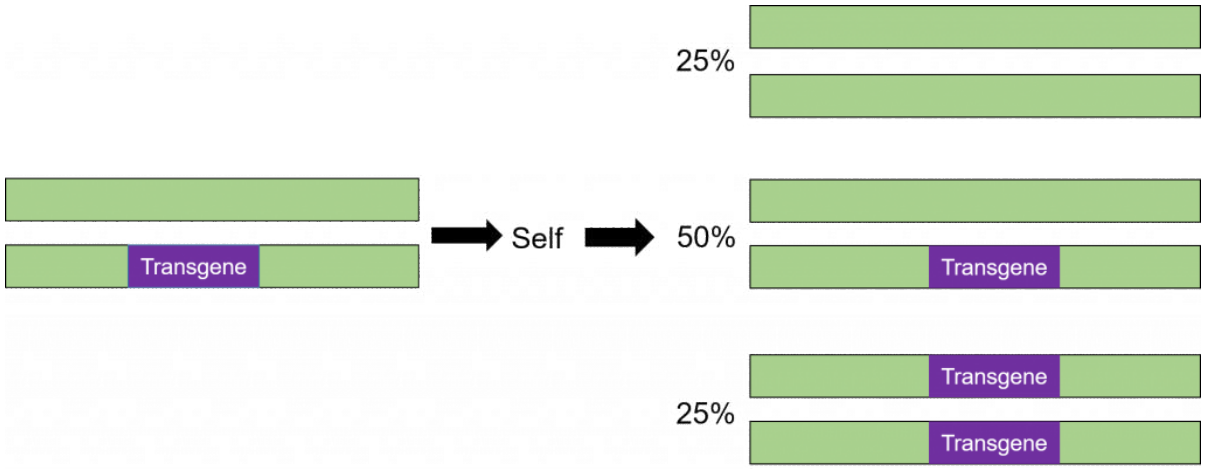


***Figure 8**. The ratio of possible offspring when a hemizygous diploid plant is self-pollinated. Image by Walter Suza.*

Step 5: Backcross Breeding

The fifth and final part of producing a genetically engineered crop is backcross breeding (Figure 9). Transgenic plants are crossed with elite breeding lines using traditional plant breeding methods to combine the desired traits of **elite** parents and the transgene into a single line. The offspring are repeatedly crossed back to the elite **line** to obtain a high-yielding **transgenic** line. The result will be a plant with a yield potential close to current hybrids that expresses the trait encoded by the new **transgene**.

***Figure 9**. Using the backcross breeding method. Image by Patty Hain.*

The Process of Plant Genetic Engineering

The entire **genetic engineering** process is basically the same for any plant. The length of time required to complete all five steps from start to finish varies depending upon the gene, crop species, and available resources. It can take anywhere from 6-15+ years before a new **transgenic** hybrid is ready for release to be grown in production fields.

The tissue culture process of regenerating transgenic plants from callus may result in genetic variation that is not associated with the transgene. Also, the parent line used for transformation commonly is selected for the frequency with which useful events can be obtained and not its agronomic performance. Therefore, transgenes are incorporated into commercial cultivars by conventional breeding procedures, such as backcrossing.

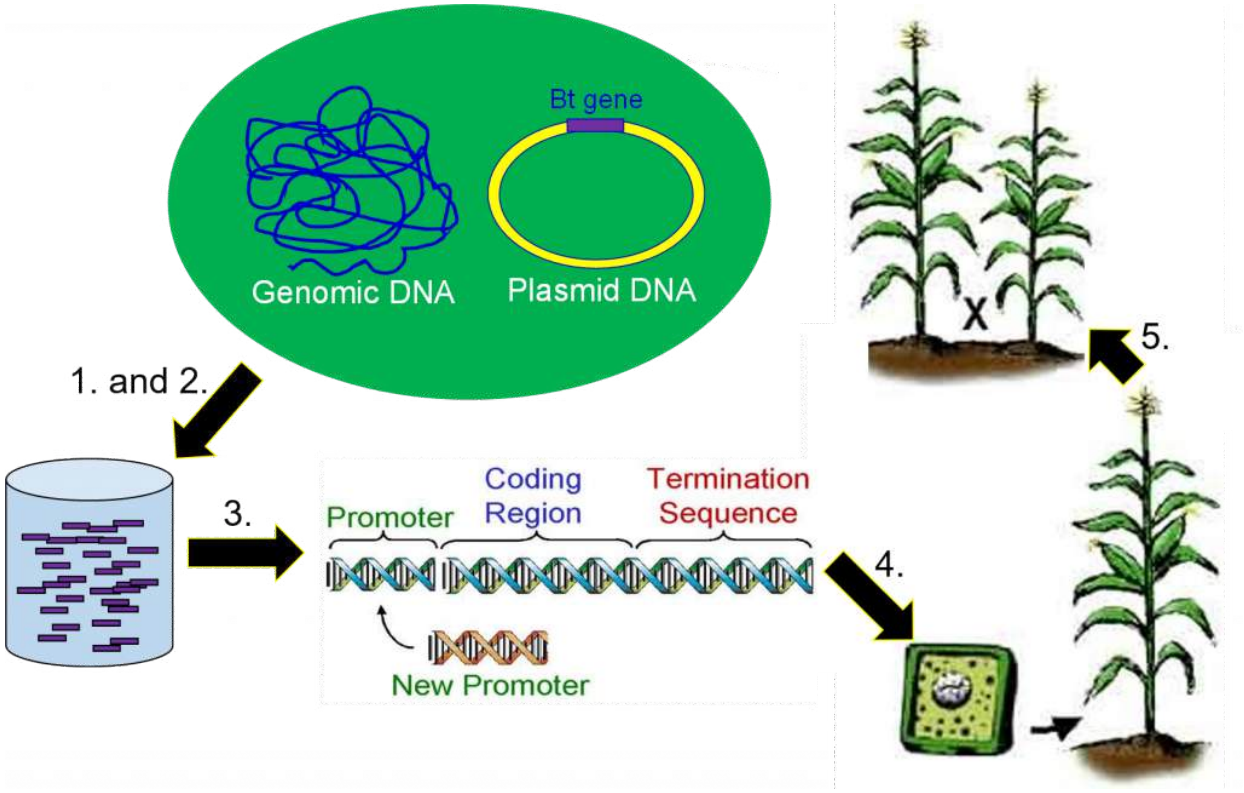


***Figure 10**. Crop genetic engineering includes: 1) DNA isolation 2) gene cloning 3) gene design 4) transformation, and 5) plant breeding. Image by Walter Suza and Patty Hain.*

Key Takeaways

Genetic engineering is the directed addition of foreign DNA (genes) into an organism.

Five basic steps in crop genetic engineering:

1. DNA extraction – DNA is extracted from an organism known to have the desired trait.
2. Gene cloning – The gene of interest is located and copied.
3. Gene modification – The gene is modified to express in a desired way by altering and replacing gene regions.
4. Transformation – The gene(s) are delivered into tissue culture cells, using one of several methods, where hopefully they will land in the nucleus and insert into a chromosome.
5. Backcross breeding – Transgenic lines are crossed with elite lines to make highyielding transgenic lines.

References

Vasil, I. K. (2008) A short history of plant biotechnology. Phytochem 7: 387-394.

Vasil, I. K. (2008) A history of plant biotechnology: from the Cell Theory of Shleiden and Schwann to biotech crops. Plant Cell Rep 27: 1423-1440.

13. Introduction to Mendelian Genetics

DONALD LEE; WALTER SUZA; AMY KOHMETSCHER; AND MARJORIE HANNEMAN

Learning Objectives

1. Outline the experimental approach Mendel used to propose the idea that genes exist, control traits, and are inherited in predictable ways.
2. Compare the methods used by Mendel and Punnett to predict trait inheritance.

Introduction

In plant and animal genetics research, the decisions a scientist will make are based on a high level of confidence in the predictable inheritance of the genes that control the trait being studied. This confidence comes from a past discovery by a biologist named Gregor Mendel, who explained the inheritance of trait variation using the idea of monogenic traits.

Monogenic characters are controlled by the following biological principles:

- Living things have genes in their cells that encode the information to control a single trait. These genes are stable and passed on from cell to cell without changing.
- The genes are in pairs in somatic cells. When these cells divide to form gametes, the pair of genes is divided. One gene from the pair goes into a gamete.
- Male gametes (pollen) combine with female gametes (eggs) in the wheat flower pistil and fuse to form the next generation (zygote). Gamete union is random.
- The zygote, again, has two copies of each gene. As the zygote grows into a multicellular seed and the seed grows into a plant, the same two gene copies are found in every cell.

Let's take a short genetics history lesson to understand their confidence.

Mendel's Peas

***Figure 1**. Gregor Mendel was born Johann Mendel.*

In the mid 1800's, an Austrian monk named Gregor Mendel (Figure 1) decided he should try to understand how inherited traits are controlled. He needed a model organism he could work with in his research facility, a small garden in the monastery, and a research plan. His plan was designed to test a hypothesis for the inheritance of trait variation.

Since Mendel could obtain different varieties of peas that differed in easy to observe traits such as flower color, seed color and seed shape, and he could grow these peas in his garden, he chose peas as the model organism for conducting his inheritance control study. A model is easy to work with and often what you learn from the model you can apply to other organisms.

The Hypothesis

While many biologists were interested in trait inheritance, at the time Mendel conducted his experiments none of the biologists had published evidence that inheritance could be predicted. Mendel made this bold statement. His hypothesis was that he could observe "mathematical" regularities in the appearance of a trait that was passed on from parents to their offspring. Mendel had the idea that mathematical regularities could be observed and could be used to explain the biology of inheritance!

The Plan

Mendel's experimental plan was designed to test the hypothesis. He identified true breeding lines of peas by allowing them to self pollinate (which we will refer to as "selfing") and examining their offspring. Pea plants have flowers that contain both male and female reproductive parts; if a pea flower is left undisturbed, the male and female gametes from the same flower will combine to produce seeds, the next generation. If the pea always made offspring like itself, Mendel had his true breeding line. He then made planned crosses between lines that differed by just one trait (monohybrid crosses). The controlled monohybrid cross was the first step in his experiment that allowed him to look for mathematical regularities in the data for three generations. Table 1 below shows the data from a series of these monohybrid cross experiments.

The Analysis

By summarizing his data in a single table, Mendel could look for those hypothesized math regularities. A regularity is a repeated observation.

Table 1. Mendel result from crossing peas.

Character	Cross and Phenotypes	F_1	F_2	F_2 Ratio
Seed form	Round X Wrinkled	All round	5474 Round 1850 Wrinkled	2.96 to 1
Cotyledon color	Yellow X Green	All yellow	6022 Yellow 2001 Green	3.01 to 1
Seed coat color*	Gray X White	All gray	705 Gray 224 White	3.15 to 1
Pod form	Inflated X Constricted	All inflated	882 Inflated 299 Constricted	2.95 to 1
Pod color	Green X Yellow	All green	428 Green 152 Yellow	2.82 to 1
Flower position	Axial X Terminal	All axial	651 Axial 207 Terminal	3.14 to 1
Stem length	Tall X Short	All tall	787 Tall 277 Short	2.84 to 1

*Gray seed coat also had purple flowers; White seed coat had white flowers.

Table 1 demonstrates that Mendel was serious about the math. He generated large numbers of offspring that allowed him to observe mathematical ratios. From his table of data, we can see mathematical patterns appear with every monohybrid cross he made.

- F_1: All the plants had the same phenotype as one of the parents.
- F_2: Both phenotypes are present, the phenotype that was not expressed in the F_1 appears again in the F_2 but is always the least frequently produced. The average ratio is about **3:1** for the two phenotypes.

What was striking to Mendel was that every character in his study exhibited the same kind of mathematical pattern. This suggested that the same fundamental processes inside the plant's reproductive cells were at work controlling the inheritance of each trait.

Now Mendel had the task of providing a description of the fundamental biology process controlling each of these traits. He needed to come up with ideas that no one had yet proposed to explain biology.

New Idea #1:

The traits expressed in the pea plant were controlled by some kind of particle. These hereditary particles are stable and passed on intact from parent to offspring through the sex cells. (NOTE: Sex cells or gametes were not a new idea, Mendel was aware that biologists knew sexually reproducing plants and animals needed to make gametes.) We now call these particulate factors genes and will use that term in the rest of this reading.

New Idea #2:

Genes are stable, and genes can have alternative versions (alleles).

New Idea #3:

Genes are in pairs in somatic cells and these paired genes separate during gamete formation. Each gamete will have one gene from the pair of genes. The segregating of the paired genes from the somatic cells of the parent into gametes is random. Because segregation is random, a parent that has two different alleles for a gene pair will make two kinds of gametes and makes these gametes at equal frequencies.

From Mendel's ideas, we can see that in a situation in which there was a normal version of a gene (we can call it the R gene) and an alternate version (r), the plant could produce gametes with just the R gene or just the r gene.

New Idea #4:

Plant flowers are designed to allow male gametes (pollen) to combine randomly with the female gametes (egg). When the gametes randomly come together, they bring the genes they carry to the same zygote. This means plants could have the genotype **RR, Rr**, or **rr** in families that have both the **R** and **r** alleles.

New Idea #5:

Mendel proposed that the genes controlling a trait not only paired in somatic cells, they also interacted in controlling the traits of the plants. For the traits in his experiment, he proposed that one allele interacted with the other in a dominant fashion. That means a plant that is the genotype **RR** would have the same phenotype as an **Rr** plant. The **R** allele is dominant to the **r** allele.

Ideas and Data advance science

Those were Mendel's new ideas; he used them to make sense of his experiment data and observations. Let's think like Mendel and apply those ideas.

All the F1 were the same

Mendel's new ideas could explain this observation. Since his parents were true breeding, he was always making a cross between homozygous parents. Homo means the same, so the parents had two copies of the same version of the gene.

Crossing RR X rr plants to produce Rr

Since the **R** is dominant to **r**, then the **Rr** offspring (named the F_1) look the same (have the same phenotype) as the **RR** parent. Therefore, only one phenotype is observed in the F_1. But the F_1 genotype is different from either parent. It is heterozygous (two different alleles).

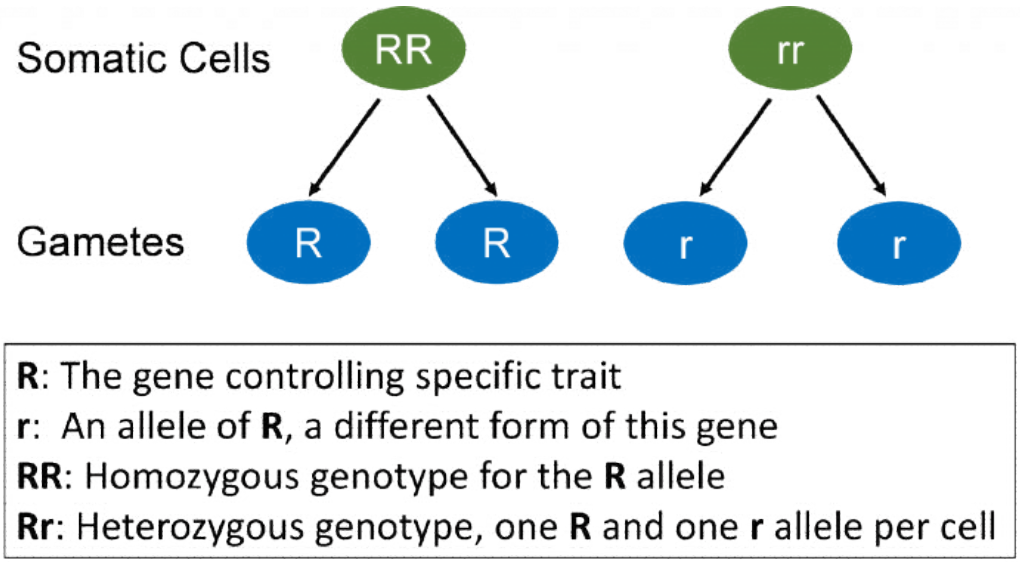


Figure 2. *Mendel used letters to represent the genes (he called them particulate factors) that he proposed were found in pairs in the cells of the pea plants. These could vary in form (alleles) and segregate when gametes (sex cells) were formed as the plants prepared to sexually reproduce. Image by Walter Suza.*

The F_2: both traits appear in about a 3:1 ratio

Mendel could explain the reappearance of the recessive trait and the ratio by combining the idea of genes with the idea of random segregation. Mendel used simple algebra to explain this result.

First, he wrote out a mathematical expression to account for the gametes made in the male part of the F_1 flower or in the female part.

½ R + ½ r = all the gametes made (Figure 2).

Next, he reasoned that if pollen randomly united with the egg to combine the genes in the gametes, then algebra could be used to predict the result by multiplying the gamete expressions.

(½ R + ½ r) X (½ R + ½ r) = all the F_2 offspring made.

If we do the multiplication above, we get …

¼ RR + ¼ Rr + ¼ Rr + ¼ rr = ¼ RR + ½ Rr + ¼ rr = predicted fractions of F2 genotypes.

WAIT!

If this math is causing your brain to lose focus, you might be experiencing what Mendel's contemporaries experienced when they read his published research paper. While many biologists were motivated to understand how the variation among animals and plants was controlled and inherited, it took biologists 30 years to recognize that Mendel's new ideas to explain inheritance of traits in peas could be applied to inheritance of traits in other living organisms.

One possible explanation for this 30-year delay in appreciation is that it was difficult for biologists to understand how math could explain biology. One biologist that did understand what Mendel was describing was Punnett. Punnett

decided to convert Mendel's algebra into a more graphic representation of the process of gamete segregation and random union.

The Punnett Square

Math: (¼ RR + ½ Rr + ¼ rr).

Punnett designated the gametes made in the male and female parents with single letters (Figure 3). The diagram shows that when the gametes combine, the offspring (inside the squares) again have the genes in pairs in their cells. Accounting for the random union of gametes is accomplished with the four squares in the diagram. Two squares give the same **Rr** result, one the **RR** genotype and one **rr**. Both the algebra and diagram approaches provide the same prediction. Crossing an **Rr** with an **Rr** will produce three genotypes, **RR, Rr** and **rr**. They will be produced in a ratio based on the principle of segregation.

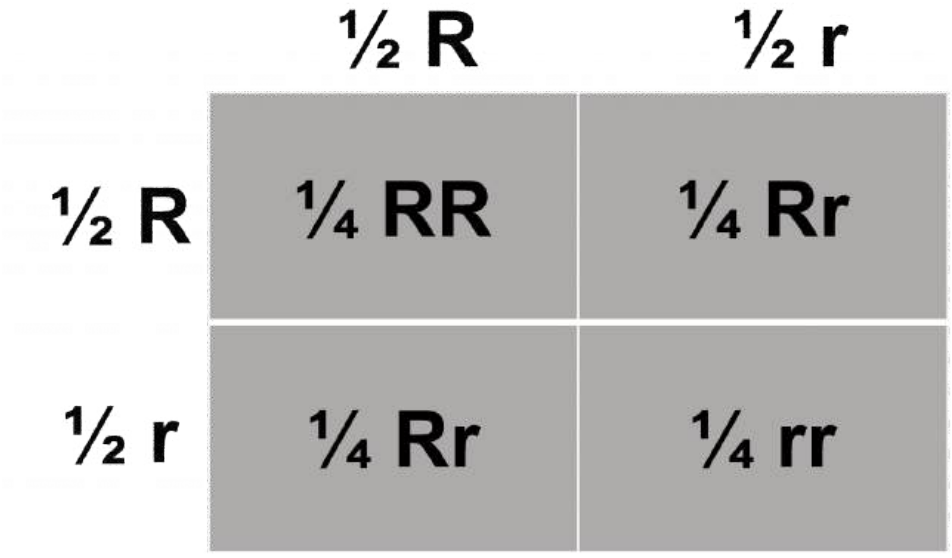


Figure 3. *The square diagram accounts for all the gene inheritance possibilities from selfing the Rr (F_1) parents. If inheritance is random, math can be applied, and the diagram used to predict both possible outcomes and the fractions expected of each.*

The genes controlling the monogenic traits behaved in predictable ways

Punnett's diagram clarified for many biologists what Mendel was telling them in his published article. This was a challenging idea to understand because he was asking biologists to use something they could not see (genes) and explain something they could see (traits in peas or some other living organism).

Because Mendel recognized he was proposing a very different idea with the segregation principle, he was likely motivated to share the most convincing evidence possible. Mendel conducted additional experiments. One experiment was to test the hypothesis that there were two different kinds of F_2 which expressed the dominant trait, and these two types were being made by the F_1 in predictable fractions. How would Mendel show that F_2 which had the same phenotype did not always have the same genotype?

Mendel tested the breeding behavior of the F_2. Mendel harvested all the selfed seed produced by his F_2 and grew progeny

rows of F_3. His segregation principle predicted that of the dominant F_2, there should be two that are heterozygous for every one homozygote made (on average). The results of this experiment are summarized in **Table 2**. Did Mendel's data support the hypothesis?

Table 2. Selfing dominant F_2 to produce F_3

F_2 type	Mixed rows	True breeding	Ratio
Round seed	372	193	1.93 to 1
Yellow cotyledon	353	166	2.13 to 1
Gray seed coat	64	36	1.78 to 1
Inflated pod	71	29	2.45 to 1
Green pod	60	40	1.50 to 1
Axial flower	67	33	2.03 to 1
Tall plant	72	28	2.57 to 1

Average ratio heterozygote F_2 to homozygote F2 was 2.06 to 1.

The data show that, if we select a sample of F_2 with the dominant trait (Round seed or Yellow cotyledon), the principle of segregation predicts that there should be 2 heterozygotes for every 1 homozygotes.

Mendel's data from rows of F_3 that all came from F_2 with the dominant trait supported his hypothesis. There were always two kinds of rows (true breeding and mixed) and the rows were in a 2:1 ratio. This fits with the **principle of segregation**.

By publishing these results in a scientific journal, Mendel allowed other scientists to learn from his work. This story reveals the real power of publishing research in the "permanent" scientific literature. The power of publication does not mean you were right with your science. The real power is that other scientists can find your paper, read it, think about your ideas, and then test them. In Mendel's case, he was already dead when his fellow biologists discovered that his new ideas to explain the biology of peas were not only correct, but universal in their application.

Mendel's Dihybrid Cross Experiments

Proper credit must be given to the idea of independent assortment. Gregor Mendel was the first to put this idea down on paper based on what he observed with his pea experiments. Furthermore, Mendel performed additional experiments to back up his ideas. Let's examine his experiments with peas from the late 1800's.

The outline below describes Mendel's dihybrid cross experiments. The pattern observed in the results should look familiar!

The Experiment

- **Parents:** round seeds, yellow seeds (RRYY) x wrinkled seeds, green seeds (rryy).
- **F_1:** All round and yellow seeds (RrYy).
- **Selfing:** F_1 (RrYy x RrYy):

Mendel explained his results as follows:

The F_1 plants have the genotype RrYy and can make four kinds of gametes RY, Ry, rY and ry.

Table 3. Gametes are in equal frequencies in the male and female parts of the plant.

	Possible gametes
Male gametes	¼ RY + ¼ Ry + ¼ rY + ¼ ry
Female gametes	¼ RY + ¼ Ry + ¼ rY + ¼ ry

Table 4. Punnett square demonstrates the possible genotypes produced when the RrYy dihybrid is selfed.

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

Note that with both the Mendel algebra and Punnett square, the **RRYY** genotype occurs one time and the **RrYy** genotype occurs four times (Table 4). Mendel's algebra and Punnett's squares can be summarized to give the same results.

Selfing the F_2 to produce F_3

Table 5. Phenotype classes and their fractions in F_2.

Phenotype	Number	Fraction
Round, yellow	315	9/16
Round, green	108	3/16
Wrinkled, yellow	101	3/16
Wrinkled, green	32	1/16

The easiest experiment to perform was to let the plants self-pollinate and then keep good records. After scoring his 556 F_2 seeds (Table 5) he took the 315 that were round and yellow and planted them in one part of his garden. The plants that grew were allowed to self-pollinate. Of the 315 round and yellow seeds planted, 301 plants matured and produced seed. The seed produced was the F_3 generation. At harvest, Mendel needed to exercise the utmost care. Each F_2 plant was handled separately. The seeds from the plant were harvested and Mendel then scored the F_3 seeds that came from the same F_2 plant. This can be referred to as $F_{2:3}$ data and the table below summarizes his complete experiment using all of the F_2 phenotypes.

Mendel's F_2 data supported his principle of independent assortment. There were four different types of round yellow F_2 based on the kinds of progeny they could produce or their breeding behaviors. Based on the F_3 progeny produced, the F_2 genotype was deduced. For example, if a round, yellow seed gave all round progeny it must have the genotype **RR**. If it gave both round and wrinkled it was **Rr**.

Furthermore, the numbers of F_2 plants with each breeding behavior were in agreement with what was expected with independent assortment. There were four times as many round and yellow F_2 that gave all four phenotypes of F_3 seeds (138) compared to the round and yellow F_2 that were true breeding (38). Overall, there were nine types of breeding behaviors demonstrated in the F_2 demonstrating that there were nine F_2 genotypes. In all cases, the fractions observed in the F_2 agreed to the principle of independent assortment. Mendel's well-planned experiment provided a convincing demonstration that genes behaved in this predictable manner.

The only thing better than performing an experiment that shows you were right about a new hypothesis is performing two experiments that show that you were right. That is what Gregor Mendel did! In his second experiment he crossed dihybrid F_1 plants with homozygous recessive plants in a test cross. This type of cross is named because the geneticist wants to perform a cross that will test or reveal the genotype of an organism. Therefore, a test cross is usually made between an organism with a dominant trait and a partner with a recessive version of this trait. Mendel performed the **RrYy x rryy** testcross and the expected progeny are shown in the Punnett square below:

RrYy gametes: **RY, Ry, rY, ry**

rryy gametes: all **ry**

	RY	**Ry**	**rY**	**ry**
ry	RrYy	Rryy	rrYy	rryy

The observed result closely matched the expected. The testcross experiment provides additional support for the

principle of independent assortment.

Mendel established a rigorous precedent for using carefully planned multi-generation experiments to reveal the principles that governed trait inheritance. The beauty of Mendel's accomplishments is that both the principles and his experimental approach can be applied to understanding the genetic control and inheritance of traits in many kinds of organisms still today.

Lessson Summary

Mendel's principles of segregation and independent assortment are valid explanations for genetic variation observed in many organisms. Alleles of a gene pair may interact in a dominant vs. recessive manner or show a lack of dominance. Even so, these principles can be used to predict the future...at least the potential outcome of specific crosses.

Video Example

Watch this video about Punnett Squares for more information

14. Deviations from Mendelian Genetics: Linkage (Part 1)

DONALD LEE; WALTER SUZA; AND MARJORIE HANNEMAN

Learning Objectives

At the completion of this lesson, you will be able to:

1. Contrast the inheritance of traits that are controlled by independent genes, linked genes and pleiotropic genes.
2. Demonstrate the physical basis of linkage by drawing the key events in meiosis.
3. Calculate map distances from 2-point test cross and F2 data.
4. Assemble linkage maps from map distance information.

Maize Genetics

Maize (*Zea mays, corn*) is not only a commercially important crop; it has also been a valuable organism used in basic genetic studies to understand genetic principles. Geneticists such as Dr. John Osterman, in the School of Biological Sciences at the University of Nebraska, use corn in basic genetic studies. Let's examine one experiment.

Many seed traits in corn have been studied by geneticists. Color in the outside (aleurone) or inside (endosperm) of the seed can be important for end uses such as what color of corn chips you prefer. Starch development traits in the kernel influence what the corn seed can be used for (popcorn, sweet corn, waxy corn).

A line of corn that was true breeding for a red kernel color was crossed to a true breeding line with shrunken kernels but no red color (white or yellow). Even though the two parents differed by two traits, kernel color and kernel shape, the F_1 offspring produced from the cross were all red with plump or normal kernels.

Parents: Red, plump x white, shrunken

F_1 offspring: All red, plump

Which phenotypes would you say are dominant based upon this single result? Going a step further, what is your hypothesis for the genotypes of the parents and F_1?

The next year these red plump seeds were planted in the crossing nursery near the white shrunken line. Dr. Osterman made a **backcross** or **testcross** between the F_1 progeny and white, shrunken plants.

Thinking back on the genetics principles we have covered so far in this course, what phenotypes would you expect to observe in the testcross progeny?

If we use the principle of independent assortment and the information we already have, we would predict that there should be four types of seeds in testcross progeny.

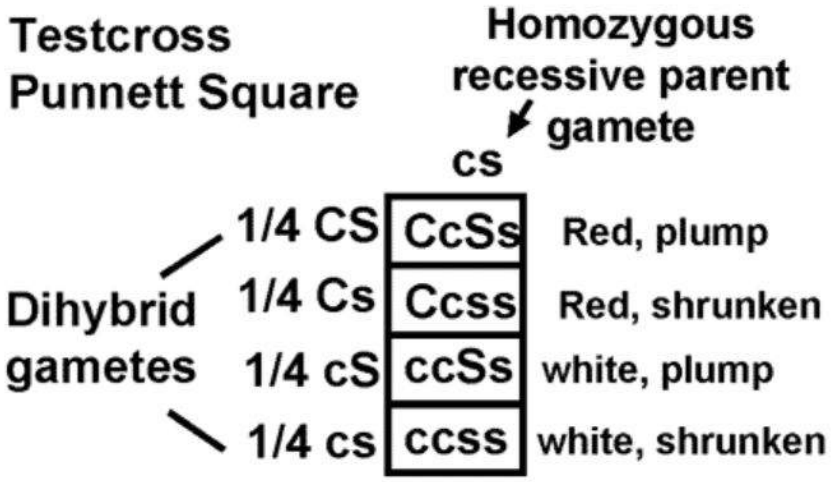


Figure 1. *Seed trait inheritance in test cross progeny assuming independent assortment. Image by Donald Lee.*

We should get both parental combinations of red, plump and white, shrunken. We should also get two new combinations; red, shrunken and white, plump. Furthermore, the principle of independent assortment predicts the four combinations should be in a 1:1:1:1 ratio. (Make sure you understand why!)

What did Dr. Osterman actually observe? When he harvested the ear from the first plant, peeled back the husks, and examined the kernels, all he saw were red, plump and white, shrunken seeds. Only the parental combination of traits was found among the several hundred seeds on that ear!

Half the seeds had the dominant red and plump combination while the other half had the recessive traits. While this result does not fit what we expect based on the idea of independent traits, Dr. Osterman was not surprised. Obviously, another genetic hypothesis can explain this result. Let's examine two alternative explanations.

Figure 2*. Testcross progeny showing red, plump and white, shrunken phenotype. Image by Donald Lee.*

Pleiotropy

Sometimes a single gene pair will control more than one trait. For example, barley that has good malting characteristics for brewing also tends to sprout prior to harvest. High sucrose lines of soybean have a sweeter flavor in soymilk and cause less gas production in the consumer compared to lines with normal sucrose levels. Perhaps a seed color gene in corn also controls whether the kernel is plump or shrunken. White kernels would always be shrunken, red would always be plump. When genes control more than one trait, these traits will always be inherited together. This explanation is consistent with what Dr. Osterman observed on the seeds of the testcross ear.

Linkage

The second explanation involves two genes but looks closely at their chromosome locations. Cytogeneticists have observed ten pairs of chromosomes in the somatic cells of corn. Corn has countless traits controlled by tens of thousands of genes. Since genes are on chromosomes, it makes logical sense that each chromosome is made of thousands of genes. When genes are on the same chromosome they will travel together as the chromosome is passed on to the gametes (Figure 3). If the gene for red vs. white kernel color is on the same corn chromosome as the gene for plump vs. shrunken, those genes will travel together during gamete formation. This could also explain why the red and plump gene alleles are inherited together. The same would be true of the white and shrunken alleles on the homologous chromosome in the F_1. Therefore, we could explain the result by the hypothesis of linkage; a separate gene pair controls each trait but these genes are located near each other on the same chromosome.

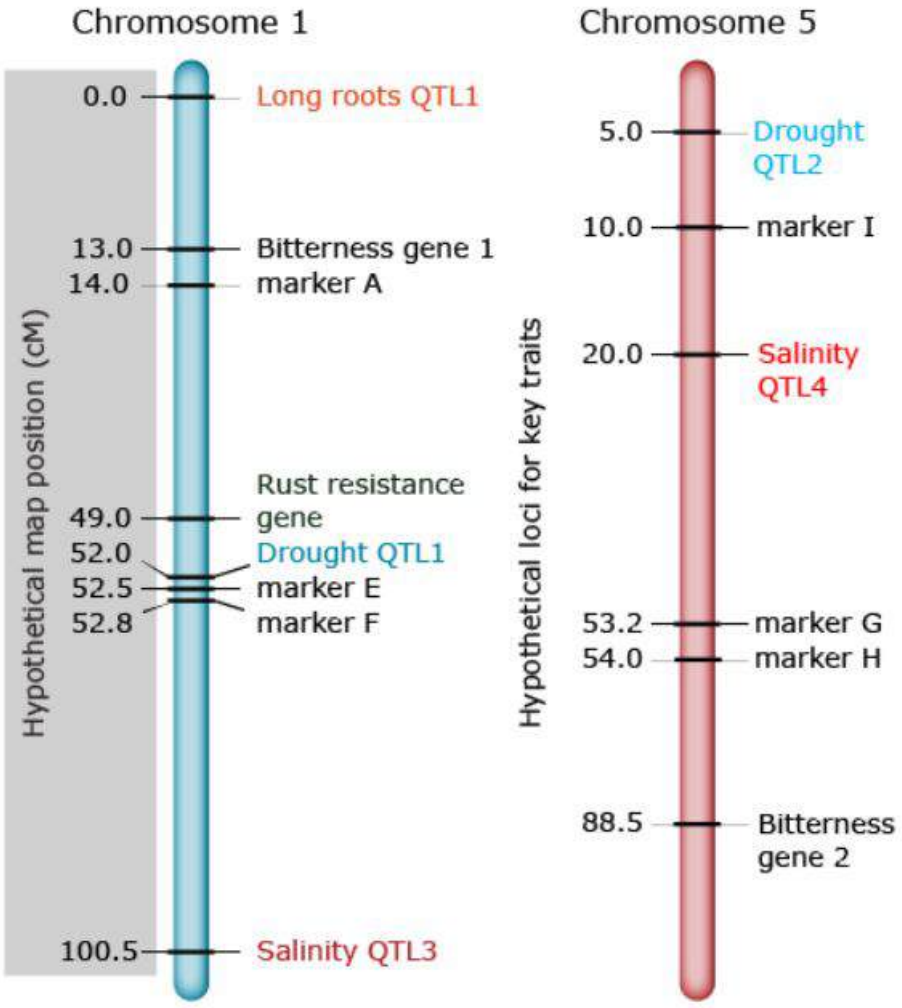


***Figure 3**. Genes located on the same chromosome are genetically linked. Genetic linkage analysis can be used to determine the order of genes on chromosomes. Closely linked genes are not segregating independently, like genes located on different chromosomes. Linked genes can be used as genetic markers, which have become an important tool in plant breeding. Image by Walter Suza.*

More Data

Two genetic hypotheses can explain the testcross data. How can we determine which hypothesis is correct? Like all good geneticists, Dr. Osterman made an effort to generate large numbers of offspring in his studies. He had made several test crosses that summer and each cross generated ears with several hundred seeds. Upon careful examination of the seeds on all the ears, Dr. Osterman did observe a few red shrunken and white plump kernels. The new combinations were rare but their appearance in the testcross progeny allows us to eliminate one of the genetic hypotheses. Which hypothesis can be ruled out?

Linked Genes Tend to Stay Together

Because the red trait and plump trait are not always found together, one gene cannot be responsible for both traits, the occurrence of the recombinant red, shrunken and white, plump testcross offspring eliminates this as a reasonable hypothesis. Can we still use linkage as an explanation though? Yes, if we think through chromosome behavior during meiosis, we can see why genes on the chromosome tend to stay together but will not always be passed on together. Let's think about meiosis and crossing over to understand these results.

Crossing Over

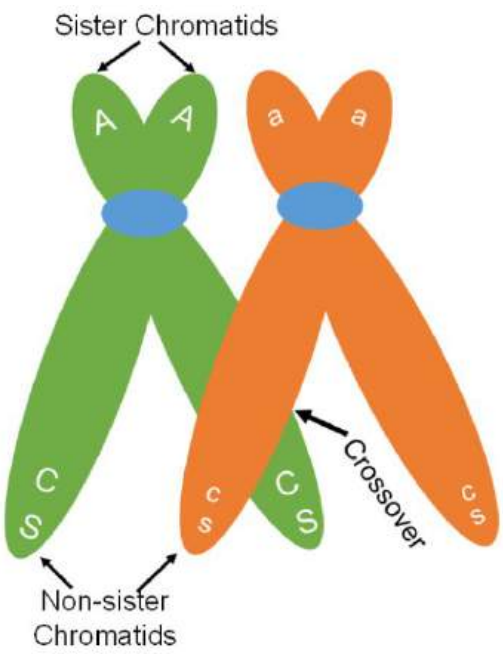


***Figure 4**. Crossover not between C,c and S,s. Only CS and cs. Parental gametes are made. Image by Marjorie Hanneman.*

One of the key differences between meiosis and mitosis is the synapsis of homologous chromosomes during prophase I of meiosis. Synapsis is the process in which homologous chromosomes carefully pair. The pairing allows for an orderly first division to send one chromosome from each pair to separate cells. The close association of the homologous chromosomes also allows for crossing over between non-sister chromatids (Figure 4). During this process sections of the chromosomes break off and are exchanged between non-sister chromatids. When non-sister chromatids crossover, chromatids can be made that have a new combination of genes compared to the original combination on the chromosome. The original combination was inherited from the organism's parents and is called the parental combination of genes. The new combination made is called the recombinant combination. In figure 4, a crossover occurs but the original or parental combination of **CS** (red and plump) and **cs** (white and shrunken) will stay together. Crossing over can cause new gene combinations to occur on a chromosome if the crossover occurs between the linked genes.

When a crossover occurs between genes, chromatids with both the parental combination and chromatids with a new combination will be made. We can see this in figure 5. Two of the chromatids are not involved in the crossing over. These chromatids will maintain the parental combination and when meiosis is complete, the two gametes made that have these chromosomes will be called parental gametes. The gametes made that have the other two chromosomes, those that went through crossing over and have the new gene combination, are called recombinant gametes (Figure 6).

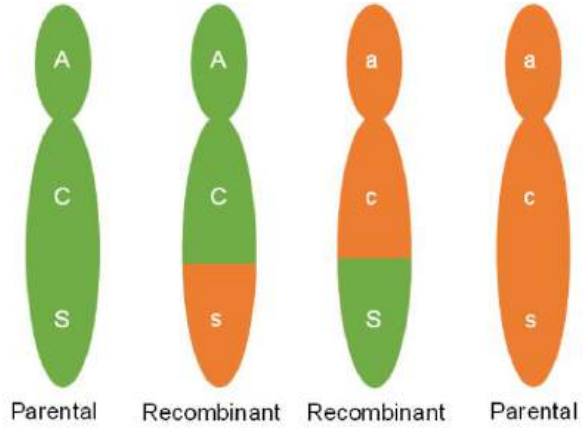


***Figure 6**. Parental and recombinant gametes from crossovers between C,c and S,s. Image by Marjorie Hanneman.*

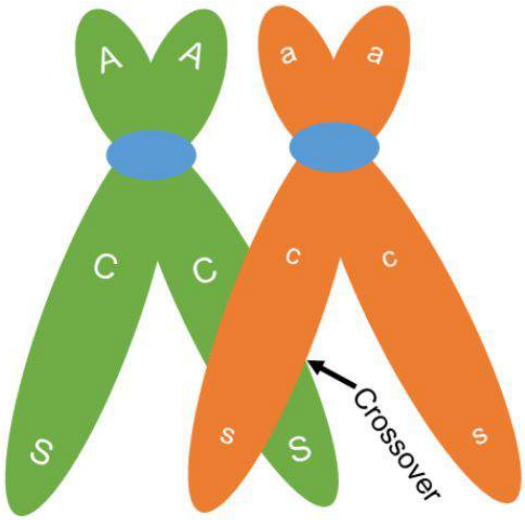


***Figure 5**. Crossover between the C,c and S,s gives Cs and cS recombinant plus CS and cs parental. Image by Marjorie Hanneman.*

When crossing over occurs between two non-sister chromatids, cells will make equal numbers of recombinant and parental gametes. In looking back at Dr. Osterman's data though the number of plants that would have inherited recombinant type gametes was far below 50%. Why are parental combinations of linked genes made more frequently than new combinations? Understanding this requires us to imagine many cells going through gamete formation.

Making Lots of Gametes

Sexually reproducing organisms tend to make a large number of gametes to ensure reproductive success. Therefore, we need to think about crossing over events happening in many cells. Geneticists do not understand all the details of crossing over but in general, crossing over is a random process that occurs during **prophase I**. We also know that the occurrence of one crossover along a chromosome can reduce the chance of a second crossing over. As you look at figure 5, you can see that crossing over could occur at any point between where the **C,c** genes (the **C,c** locus) and the **S,s** locus reside on the chromosome causing recombinant gamete formation. If a crossover does not occur between the **C,c** and **S,s** loci (plural for locus), then only the parental gamete combination of **CS** and cs will be made. Considering the size of the chromosome and the relative positions of the **C,c** and **S,s** loci; how often would cells have crossovers occur between the genes compared to somewhere else along the chromosome? If crossing over is random, we would not expect a cross over to occur in very many cells between the **C,c** and **S,s** loci. Genes that are on the same chromosome tend to stay together at a rate that is proportional to how far apart they are on the chromosome. Since the **C,c** and **S,s** loci are close to each other, crossovers are rare. The **'C'** gene tends to stay with the **'S'** gene as it is passed from the red, plump line to the F1 and then to the testcross generation. The same is true of the **'c'** and **'s'** parental combination.

How does this discussion of crossing over help us understand what we see from our test cross result? The parental

combinations of red, plump, and white shrunken were almost always made. The new combinations of red, shrunken and white, plump were rarely observed. It would take a recombinant gamete to make one of these rare combinations. While we never see the genes in the gametes or directly observe the genes on the chromosomes, the outcome suggests that these loci are linked close together on the chromosome. Making the connection between the results of crossing studies such as these has allowed geneticists to map genes or determine their relative positions on a chromosome. Let's go through a mapping study to see how it is done.

Two-Point Testcross Mapping

The results of a testcross study different from Dr. Osterman's is given below

- Red, Shrunken **CCss** X White, Plump **ccSS**
- F_1: all Red, Plump (**CcSs**)
- F_1 (**CcSs**) crossed with White Shrunken (**ccss**)

Table 1. Testcross progeny data

Phenotypes	**# of progeny**
Red, plump	120
Red, shrunken	3420
White, plump	3334
White, shrunken	126
Total	**7000**

Obviously, the geneticist who did this experiment made a lot of test crosses to generate these numbers. How can we use the numbers to map these genes? With this information we can answer one question. 'What is the distance between the **C,c** and **S,s** loci?' We cannot get out a fancy microscopic ruler and physically measure the distance on the corn chromosome. With this information all we can do is estimate the map unit distance. Map units are a measure of the tendency for crossovers to occur between two loci. Because genes that are farther apart will have a higher likelihood of crossovers, the higher the crossover frequency, the farther apart the genes are on the chromosome. Let's apply this idea to our test cross data.

Test cross data allows us to indirectly measure the frequency of gametes made by an individual. All of the testcross progeny inherited a gamete with the recessive '**c**' and '**s**' alleles from the white, shrunken parent. Therefore, the alleles that the F_1, dihybrid parent has passed on determine the traits in the seed (Table 1). We need to be able to measure how often crossovers occurred between the **C,c** and **S,s** loci when these dihybrids made gametes. From the data we have the following:

- Gametes that were passed to the F_1 from the parent lines were **Cs** and **cS**.
- Gametes made from crossing over in the F_1 (recombinant gametes) were **CS** and **cs**.
- Gametes with the original parent combination (parental gametes) were **Cs** and **cS**.

Therefore, the frequency of recombinant gametes was 246 (120 + 126) divided by the total gametes we have information on (7000) or 246/7000 = **3.5%**

Map unit distance between the **C,c** and **S,s** loci = 3.5 Map units.

One map unit is equal to 1% recombinant gametes. Again, this is not a physical measurement. It is a relative measure of how often crossovers occurred between these loci. How reliable is this measurement? Let's look at another data set (Table 2).

- Red, plump **CCSS** X White, shrunken ccss
- F_1: all Red, Plump (**CcSs** or **CS / cs**)
- F_1 (**CS / cs**) crossed with White Shrunken (**ccss**)

Table 2. Testcross progeny data from another data set

Phenotypes	**# of Progeny**
Red, plump	192
Red, shrunken	5
White, plump	3
White, shrunken	200
Total	**400**

Recombinant gamete frequency: 5 + 3 / 400 = 2%; 2 map units from **C,c** to **S,s**

Based on this result is the map unit distance measurement reliable? Yes, it is, to a certain degree. It is important to recognize the differences between the two test cross experiments.

Cis Versus Trans

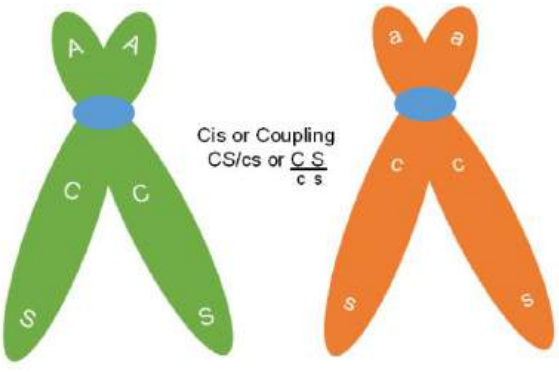


***Figure 7**. Dihybrid in Cis. Image by Marjorie Hanneman.*

The first difference to note is that the parental gametes are not always the same allele combinations, but they are always the most frequently produced gametes. In the smaller experiment above, the parents had either both dominant or both recessive traits together. Therefore, the F_1 parent was making **Cs** and **cS** recombinant gametes. This dihybrid parent was in **cis** or **coupling phase** with respect to these genes (Figure 7).

In the larger experiment with 7000 testcross progeny the dihybrid F_1 was in trans or repulsion phase for the **C,c** and **S,s** genes. In this case the **Cs** and **cS** are the parental gametes (Figure 8). Therefore, when genes are linked, they can be arranged in two ways in a dihybrid. This means that a **CcSs** plant can have its genotype in the cis arrangement with both dominant genes on one chromosome and both recessives on the other (written **CS / cs**) or can also be in trans (**Cs / cS**).

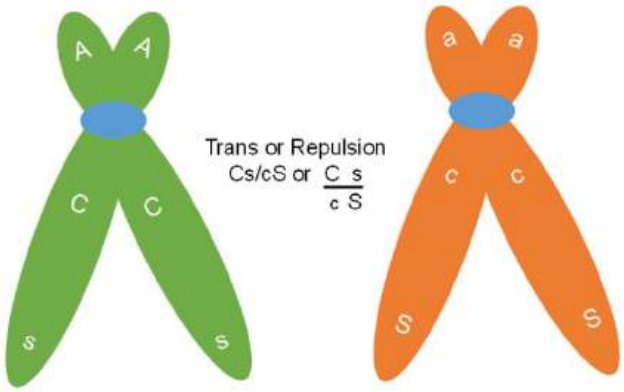


***Figure 8**. Dihybrid in Trans. Image by Marjorie Hanneman.*

Map Distance Measurement

The first experiment gave us 3.5 map units and the latter gave 2 map units.

Why do our two experiments give us two different map unit distances? In both cases, we can see that the parental gene combinations have a strong tendency to stay together. Because the experiments were done with different population sizes, with different plants, and the difference in our map unit estimates is small, the difference could be attributed to chance. Environment may also influence the rate of crossing over to some degree. For practical purposes this gene mapping information is reliable enough to tell us that these two genes reside close to each other on the same chromosome. Once we have information on distances between other genes, we can improve our map.

Mapping Another Seed Trait Gene Pair

A third seed trait in corn is normal vs. waxy starch. Waxy starch has a different chemistry than normal starch and can be scored by staining with iodine. This gene is commercially important because some waxy corn is produced for specialty markets. Let's look at the following test cross data (Table 3):

Parent: Red, Normal (CCWW) X White, waxy (ccww)

F_1: Red, Normal (CcWw) X White, Waxy (ccww)

Table 3. Testcross data for another seed trait

Phenotypes	**# of testcross progeny**
Red, normal	2781
Red, waxy	759
White, normal	749
White, waxy:	2711
Total	**7000**

- Parental gametes made by the F_1: CW and cw
- Recombinant gametes made by the F_1: Cw and cW = 759 + 749 = 1508
- **Map unit distance between the C,c and the W,w loci:** 1508/7000 = 21.5 Map Units

It is clear from the testcross data (Table 3) that the genes at the **C,c** and **W,w** loci do not have as great a tendency to be inherited together compared to the genes at the **C,c** and **S,s** loci. This would be because the loci are farther apart, and our 21.5 map units is an indicator of the relative distance.

From the two distances calculated, how far apart are the **W,w** and **S,s** loci? Here is where we can try to use map unit distances to do gene mapping the same way we use miles to do road mapping. We can pencil our two possible maps (Figure 9).

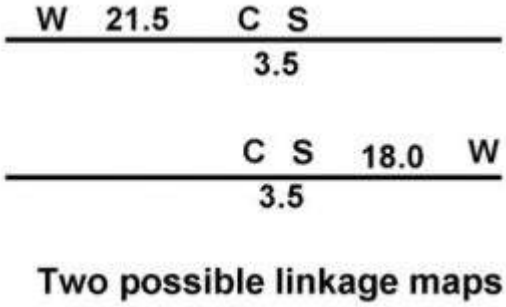


***Figure 9**. Two possible linkage maps from two-point data. Image by Donald Lee.*

In one map the **C,c** locus is in between the **W,w** and **S,s** loci. Using the 21.5 plus 3.5 map units we have calculated, the best estimate will be that 25 map units separate the **W,w** and **S,s** loci. The other map that fits our data is having the S,s locus in the middle with the **W,w** and **S,s** loci 18 map units apart (21.5 – 3.5). If the genes occupy a fixed position on the chromosome, only one map will be correct. How can we determine which is the correct map?

The Third Map Distance

A third map distance needs to be estimated in order to determine the linkage map of these three loci. We need to make a cross that gives us the opportunity to measure crossing over frequency between the **S,s** and **W,w** loci. One such cross is shown below.

- **Parent**: ssWW (shrunken, normal) X SSww (plump, waxy)
- **F_1**: sW / Sw X sW / Sw

Test cross offspring:	
Plump, normal:	630
Plump, waxy:	2824
Shrunken, normal:	2900
Shrunken, waxy:	646
Total:	**7000**

- **Recombinant gametes**: 646 + 630 = 1276
- **Map distance S,s to W,w**: 1276/7000 = 18/2 map units

Based on the third map distance our best map for all three loci is the second alternative (Figure 9). To map all three loci using this two-point test cross data, the geneticist needed to generate three different crosses. This is how gene maps in organisms such as corn have been generated. As new genes controlling traits are discovered, crosses can be made to test if these genes are independent or linked to other genes. Once linkage is determined, additional crosses can be analyzed to position the gene on the linkage map.

Map Distances from F_2 Data

In our last test cross, the F_1 dihybrid **sW / Sw** was crossed to a **sw/sw**, shrunken waxy plant. If the shrunken waxy line was not available, the geneticist could still obtain information to determine map distance by selfing the F_1 to produce F_2 offspring. Here is the procedure.

Results of Selfing the sW/Sw F_1:

F_2 Offspring	**Number**
Plump, normal	254
Plump, waxy	122
Shrunken, normal	120
Shrunken, waxy	4
Total F2s:	**500**

Step one: Focus on the double recessive (sw/sw).

How do we obtain recombinant gamete frequencies from this F_2 phenotype information? The first thing to remember is that this is a self-cross rather than a testcross and an additional Punnett square is needed to depict how gametes produced these F_2 (Figure 10). From the Punnett square we can see that four kinds of gametes are made in both the male and female in this cross. Therefore, the plump, normal seeds can be made nine different ways in our diagram and will consist of five different genotypes (**SW/SW, SW/sW, sW/SW, sW/Sw** and **SW/sw**). We cannot tell by observing the seed phenotype though, which of the genotypes it has. The plump, waxy and shrunken, normal F_2 seeds will each consist of two different genotypes. The only F_2 phenotypes that consist of a single genotype are the shrunken, waxy. They are all **ssww** (**sw/sw**). By focusing on how this phenotype was produced in the F_2, we can estimate the frequency of recombinant gametes made by the selfed parent.

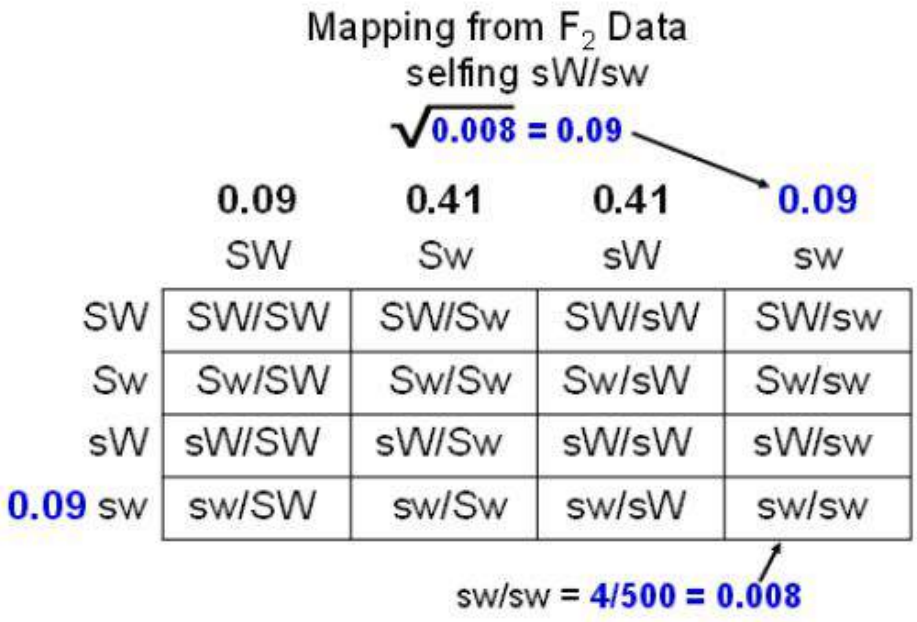


Figure 10. *Mapping from F_2 data. Focus with the method described is on the double recessive, ssww or sw/sw, F_2. Image by Donald Lee.*

Step two: Go from phenotype frequency to gamete frequency.

Look at the Punnett square and think about how the shrunken waxy seeds were made. First, a **sw** gamete in the male part of the plant and a **sw** gamete in the female needed to be made. The parent had the genotype **sW/Sw** that means that these would be recombinant gametes in both the male and female. Once these gametes were made, they needed to come together at random during pollination to produce the seed. Mathematically, we can say that the frequency of the shrunken, waxy seeds is the frequency of the **sw** gamete made in the male times the frequency of the **sw** gamete made in the female. If we assume that crossovers occur in the pollen forming cells at the same frequency as the egg forming cells, then the frequency of the shrunken waxy seeds is the square of the frequency of **sw**. Make sure this last paragraph makes sense to you by walking through figure 10.

Step three: Take the square root.

Now we can use our seed phenotype frequency to estimate gamete frequency in the selfed parent. Four of the 500 seeds produced were shrunken waxy. That would give our square in the lower right a frequency of 0.008 (4/500). This frequency should be the **sw** gamete frequency squared (**sw2**). If we want to know the frequency of **sw**, we can take the square root of 0.008 and get about 0.09. Now we have the frequency of one gamete estimated. This would be the same for both the male and female gametes, where 9% of all gametes produced by the **sW/Sw** parent will be **sw**. As it turns out, we can calculate the frequency of the other three gametes from this information if we just think about how the gametes were made in meiosis.

Step four: Double the frequency, decide if parental or recombinant.

In our continuing example the **sw** gamete was made when a crossover occurred in the **sW/Sw** parent. Crossovers are a reciprocal exchange between non-sister chromatids so every time a **sw** gamete was made, a **SW** gamete was also made. Therefore, if the **sw** frequency is 0.09 then the **SW** frequency is also 0.09. These are our recombinant gametes, so the frequency of recombinant gametes is 2 X 0.09 or 0.18. We can tell from this frequency that these are the recombinant gametes, they are less frequently made than the parental gametes. The parental gametes would total 82% of the gametes, each parental gamete making up half (41%) of that total. Thus, 18 map units between the **S,s** and **W,w** loci is our map distance estimate from the data. Because we use only a subset of the information in this procedure (only the double recessive numbers) this estimate is more subject to chance than the test cross data. We can use more complex math to estimate gamete frequencies from all the F_2 data but this square root method is a quick way to detect linkage and estimate map distance. In cases where geneticists cannot perform a testcross working with F_2 data is a necessary part of gene mapping. Can you think of some instances where this would be true?

15. Deviations from Mendelian Genetics: Linkage (Part 2)

DONALD LEE; WALTER SUZA; AND MARJORIE HANNEMAN

Learning Objectives

In this lesson you will learn to make predictions about inheritance using map unit distances and genetic markers, assemble maps from multiple-point linkage data, define the relationship between linkage maps, linkage groups and genome maps, and describe how DNA or molecular markers are observed and used in gene mapping.

At the completion of this lesson, you will be able to:

1. Make predictions about inheritance using map unit distances and genetic markers.
2. Assemble maps from multiple-point linkage data.
3. Define the relationship between linkage maps, linkage groups and genome maps.
4. Describe how DNA or molecular markers are observed and used in gene mapping.

Introduction

Why do geneticists map genes? This is a legitimate question after all this discussion. Gene mapping helps geneticists do two things, make better predictions and isolate genes. The following is an example of how gene mapping can help breeders make better selections.

Gene Maps and Selection

This example shows how gene mapping has a practical application. Soybeans have perfect flowers that self-pollinate. Tedious flower manipulation is required to make a hybrid seed. Soybean breeders would like to find ways to make hybrid seeds in commercial quantities. They have discovered a gene that causes a lack of pollen formation and makes plants male-sterile. Therefore, when bees visit these plants, they will make a cross-pollination and produce hybrid seed if they carry pollen from a nearby male-fertile plant. A hybrid seed production field could be set up if they could have all male-sterile plants in one row and male-fertile in the next. The genetic control of this trait is shown below:

Male-fertile plants: **MM or Mm**

Male-sterile: **mm**

Can you see the problem one would have in generating seed that is all male-sterile (genotype **mm**)? These plants will

never be true breeding because they cannot self-pollinate (they have working female parts but no pollen). The only plants that can be self-pollinated to produce mm offspring are **Mm** plants. Progeny from selfed **Mm** plants will segregate 3 fertile to 1 sterile, making it impossible to obtain a pure collection of mm seeds to plant in a 'female' row for hybrid seed production. If the mm genotype could be picked out in seeds prior to planting, this would solve the problem. Unfortunately, the male sterile trait is impossible to select for in the seed but breeders discovered a trick they could use that took advantage of their knowledge of gene maps. A second trait in soybean that is easy to select for in seeds is controlled by a gene that was closely linked to the **M,m** male sterile gene. This trait is green vs. yellow seed coat. The seed coat trait is controlled as follows:

GG or Gg: green

gg: yellow

The **G,g** locus and **M,m** locus have been mapped and are about two map units apart. Therefore, we can take advantage of this linkage to select seeds from a cross that will tend to also be male sterile. This is how the process would work. The following cross is made:

GGMM (green, male-fertile) X **ggmm** (yellow, male sterile)

The cis dihybrid (**GM/gm**) will be male-fertile and can be self-pollinated. Because of linkage, the **GM** and **gm** parental gametes are made 98% of the time. Therefore, seeds that are **gg** and yellow are almost always going to be **mm** and male sterile. If the breeder sorts out the F_2 seeds based on color, about one fourth of the seeds will be yellow. Based on 2-map unit distance, about 0.24 would be the expected frequency of **ggmm** out of all the F_2 but 96% of the yellow F_2 will be male sterile (0.24 out of 0.25). Therefore 96 out of 100 seeds planted in the 'female' row of selected yellow seeds will be male-sterile (mm). Harvesting seeds from these rows will provide a high percentage of hybrid seeds. Thus, the breeder took advantage of linkage to select for a trait that was easy to detect (yellow seeds) and obtain individuals with a trait that was impossible to select (male-sterile). The power of this prediction potential has driven much of the recent efforts in gene mapping in crop plants, livestock, and humans.

Table 1. Predicting the percentage of male-sterile F_2 among yellow F_2 seeds using a chart. Chart is in text form below for additional accessibility options.

	0.49 big G, big M	**0.01 big G, little m**	**0.01 little g, big M**	**0.49 little g, little m**
0.49 big G, big M				
0.01 big G, little M				
0.01 big G, little m				
0.01 little g, big M			0.0001	0.0049
0.49 little g, little m			0.0049	0.2401 (selected)

Three Point Test Cross: Multiple Point Gene Mapping

Gene mappers are motivated to map all of the tens of thousands of genes found on the chromosomes of plant or animals. Analyzing data from crosses to determine map distances for two genes at a time makes the process time-consuming

and tedious. Therefore, geneticists will often attempt to map as many genes as possible from a single set of progeny. We will go through the simplest multiple point mapping example, a three-point testcross, to demonstrate this process.

We will use our three corn seed trait loci again and use data from one cross to map these three loci. The first step would be to obtain a trihybrid individual that is heterozygous at all three loci and then perform a testcross with this trihybrid.

Parents: **CCssWW (CsW/CsW) X ccSSww (cSw/cSw)**

Trihybrid: **CcSsWs (CsW/cSw) X ccssww (csw/csw)**

Table 2. Test cross offspring

Seed trait	Gamete from trihybrid	Number
Red, shrunken, normal	**CsW**	**2777**
White, plump, waxy	**cSw**	**2708**
Red, plump, waxy	CSw	116
White, shrunken, normal	csW	123
Red, shrunken, waxy	Csw	643
White, plump, normal	cSW	626
Red, plump, normal	**CSW**	**4**
White, shrunken, waxy	**csw**	**3**
Total number of progeny:		**7000**

We observe eight phenotypes of seeds in the testcross progeny because the trihybrid can make eight kinds of gametes. As we knew, these three genes are linked and so the uneven ratio of phenotypes reflects the combinations of two parental and six recombinant gametes. The parental gametes are still the most frequently produced type, and the six recombinant gametes are made when crossing over occurs between the loci. We can systematically account for these crossovers if we follow the following four steps. First, we will explain the steps and then we will show how to use them to map the three genes.

Step 1: Identify the parental gametes.

These are **CsW** and **cSw**, that combination which came from the parent generation and the combination made by the trihybrid at the highest frequency.

Step 2: Classify the recombinants.

If we observe the gene combination in each of the six recombinant gametes, we can ask ourselves if the gamete has

a new combination for each pair of genes. For example, **CSw** has a new combination for the **CS** and the **Cw** compared to the parental gametes but the same combination for **Sw** as the **cSw** parental gamete. Therefore, these 116 gametes represent crossovers between **C,c** and **S,s** as well as **C,c** and **W,s**. This information is recorded (see below).

Step 3: Determine recombinant gamete frequency.

Once all six recombinant gametes have been classified, the total number of crossovers between the three loci can be added up and crossover percentage determined between each pair of loci. This number will reflect gene distance but one more step is needed to complete the process.

Step 4: Add in the double crossover gametes.

Two of the six recombinant gametes were made as a result of double crossovers between the two loci that are furthest apart. These crossovers have been added to the map distances between the middle locus and the two outside loci. Now we need to add these double crossovers to the outside loci distance.

Applying these four steps to our three-point test cross data would work as follows:

1. CsW and cSw
2. **C,c** to **S,s**: 116 (CSw) + 123 (csW) + 4(CSW) + 3(csw) = 246
   C,c to **W,w**: 116 (CSw) + 123 (csW) + 643 (Csw) + 626 (cSW) = 1508
   S,s to **W,w**: 643 (Csw) + 626 (cSW) + 4 (CSW) + 3 (csw) = 1276
3. 246 / 7000 = 3.5% recombinant gametes **C,c** to **S,s** 1508 / 7000 = 21.5% recombinant gametes.
   C,c to **W,w** 1276 / 7000 = 18.2% recombinant gametes **S,s** to **W,w**.
4. From the information in step 3 we can see that the **C,c** and **W,w** loci are the farthest apart so the S,s locus must be between them. We can also see that we have underestimated the distance between the outside loci (3.5 **C,c** to **S,s** + 18.2 **S,s** to **W,w** = 21.7 map units not 21.5). While this difference is small, we can rectify this and double check our work by adding in the double crossovers. The **CSW** and **csw** gametes are made very rarely. That is because is takes two crossovers in the trihybrid's chromosome to make them, not just one.

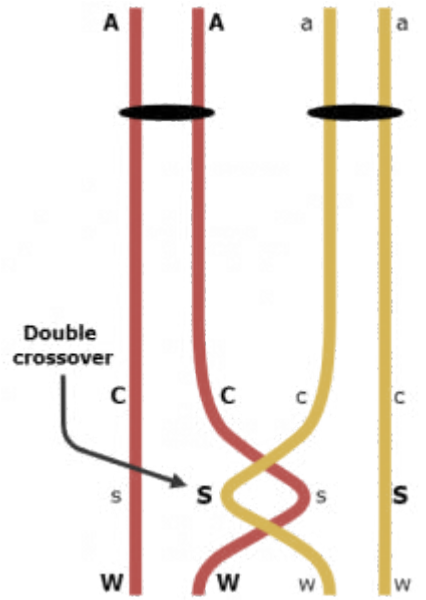


***Figure 1**. Double crossovers need to occur to produce all dominant allele (CSW). Image adapted from double crossover by Donald Lee, recreated in color by Abbey Elder*

When two crossovers occur between the **C,c** and **W,w** loci the parental combinations of **CW** and **cw** are restored. Therefore, we did not count the **CSW** and **csw** gametes as representing crossovers when they actually represent two crossovers each (Figure 1). Thus, we should add up the double crossovers (3 + 4 = 7), multiply times two and add these 14 crossovers to the **C,c** to **W,w** distance (1508 + 14 = 1522 / 7000 = 21.7 map units). Now we have double checked our map distances and have our three-point map complete.

It can be intimidating to see all the data generated from a three-point test cross. This information, however, can be used systematically to save the geneticists time and map three genes in one experiment. The three point data also provides a more accurate measure of map distances compared to two-point data when genes are farther apart on a chromosome. This is because a third gene in between the more distant loci can account for double crossovers that would not be detectable in a two-point analysis. For this reason, map distances tend to be underestimated when genes are further apart as we saw to a modest extent in this example (21.5 two-point vs. 21.7 three-point).

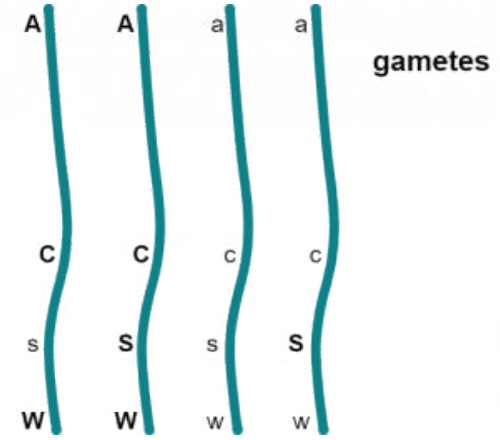


***Figure 2**. All recessive allele (csw) gametes. Image adapted from recessive gametes by Donald Lee, recreated in color by Abbey Elder*

How would a geneticist work with a four-point test cross? There would be sixteen phenotypes in the progeny as a result of the hybrid parent making two parental gametes and fourteen recombinant gametes. These recombinant gametes would be a result of single, double, and even triple crossovers that occurred in prophase I. While the data would be tedious to work with, one data set could be analyzed to reveal the map of four genes. It is not surprising that geneticists now have written computer programs which will perform these types of calculations, save time, and reduce the chance of calculator error. Geneticist's mapping genes in economically important plants and animals will also generate mapping populations that are a result of crossing parents that have different alleles at hundreds or thousands of loci. Even though the development of computers and programs has become a part of modern gene mapping the process of indirectly measuring crossover frequency by observing the inheritance of trait combinations is the basis of generating these gene maps.

Linkage Maps, Linkage Groups

The organism that gene mapping was first performed in was fruit flies. Working with fruit flies gave the early gene mappers many advantages. First, the fruit fly geneticists observed a great of deal of genetic variation among fruit flies for body part traits (wing size, eye shape and color, leg bristle types etc.) that were easy to see if they had a low powered microscope. The phenotype variants arose naturally or could be induced by chemical or radiation mutagenesis. Additionally, short generation time, large offspring numbers and low rearing costs allowed them to complete informative linkage experiments in a short time. Finally, cytogeneticists knew that fruit flies had four pairs of chromosomes so all the genes would map into four linkage groups. One chromosome was very small, and few genes mapped to this chromosome. The **X-chromosome** and the largest autosome have many genes and more than 100 map units separate genes on opposite ends of the chromosome.

Let's stop and think about how they determined this large map unit distance. If the **A,a** and **W,w** loci are 106 map units apart, how frequently will an **AW / aw** dihybrid make the gametes **Aw** and **aW**? Since these are the recombinant gametes and map distance is the frequency of recombinant gametes, we are tempted to say 106% but that is impossible. The parental gametes for any two loci will never be made less than 50% of the time, the percentage we observe when two loci are independently assorting. In fact, 50% is the correct answer here. Anytime genes that are on the same linkage group are 50 map units or more apart, they will behave as if they are independent of one another. That is because the genes are far enough apart to always allow at least one crossover to occur between them during prophase I. So how can we ever determine that **A,a** and **W,w** are 106 map units apart? Simply by determining the distances between these genes and other genes in between them that are less than 50 map units apart (Figure 3). Figure 3 shows that distances can be found between **A,a** and **B,b**, **B,b** and **C,c**, then **C,c** and **W,w**. Each of these distances is added to arrive at the 106 map units between **A,a** and **W,w**. Therefore, if the geneticist combines the mapping information from many linkage experiments or performs a multiple point linkage analysis on a progeny group segregating for many linked genes, they can deduce the larger map unit distances.

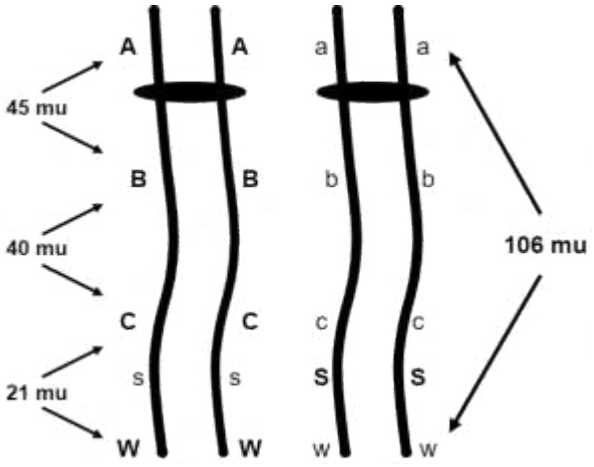


***Figure 3**. Genes from large linkage groups can be over 100 map units apart. Image adapted from linkage groups by Donald Lee, recreated in draw.io by Abbey Elder.*

Limited Traits, Limited Linkage Maps

Many gene mappers do not enjoy the advantages that fruit fly geneticists have in mapping genes. Cattle for example have thirty linkage groups (2n = 60), they often have one offspring per cross, and it is difficult of observe hundreds of phenotype differences controlled by single genes among cattle. Soybean have 20 linkage groups but for many decades, soybean geneticists had compiled about 30 different linkage maps and had no idea as to which maps were really part of the same linkage group. Soybean geneticists needed to discover more genes that had clear phenotype effects and then perform crosses that compared the inheritance of these new genes with the genes that had already been mapped. In the 1970's and 1980's, a new type of genetic trait, molecular markers, was discovered that greatly accelerated the gene mapping process in all organisms.

Appendix 1: Codon Table

Codon Table

First Position		Second Position: U code	U amino acid	C code	C amino acid	A code	A amino acid	G code	G amino acid		Third Position
	U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U	
		UUC		UCC		UAC		UGC		C	
		UUC	leu	UCA		UAA	STOP	UGA	STOP	A	
		UUG		UCG		UAG	STOP	UGG	trp	G	
	C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U	
		CUC		CCC		CAC		CGC		C	
		CUA		CCA		CAA	gln	CGA		A	
		CUG		CCG		CAG		CGG		G	
	A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U	
		AUC		ACC		AAC		AGC		C	
		AUA		ACA		AAA	lys	AGA	arg	A	
		AUG	met START	ACG		AAG		AGG		G	
	G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U	
		GUC		GCC		GAC		GGC		C	
		GUA		GCA		GAA	glu	GGA		A	
		GUG		GCG		GAG		GGG		G	

Each three-letter sequence of mRNA nucleotides corresponds to a specific amino acid, or to a stop codon. UGA, UAA, and UAG are stop codons. AUG is the codon for methionine, and is also the start codon.

To see how the codon table works, let's walk through an example. Suppose that we are interested in the codon CAG and want to know which amino acid it specifies.

1. First, we look at the left side of the table. The axis on the left side refers to the first letter of the codon, so we find C along the left axis. This tells us the (broad) row of the table in which our codon will be found.

2. Next, we look at the top of the table. The upper axis refers to the second letter of the codon, so we find A along the upper axis. This tells us the column of the table in which our codon will be found.

The row and column from steps 1 and 2 intersect in a set of boxes in the codon table, one half containing four codons and the other half containing the mapped amino acid(s). It's often easiest to simply look at these four codons and see which one is the one you're looking for.

If you want to use the structure of the table to the maximum, however, you can use the third axis (on the right side of the table) corresponding to the intersect box. By finding the third nucleotide of the codon on this axis, you can identify the exact row within the box where your codon is found. For instance, if we look for G on this axis in our example above, we find that CAG encodes the amino acid glutamine (gln).[1]

1. This description of a codon table is from "The genetic code" on Khan Academy, available under a Creative Commons Attribution NonCommercial ShareAlike 4.0 License