Plant Molecular Genetics and Bioengineering

Edited by
Walter P. Suza

IOWA STATE UNIVERSITY
Digital Press

Plant Molecular Genetics and Bioengineering

Plant Molecular Genetics and Bioengineering

Walter P. Suza; Faizo Kasule; Philip W. Becraft; Thomas Lübberstedt; and Madan K. Bhattacharyya

This book is published by

Iowa State University Digital Press

701 Morrill Rd, Ames, Iowa 50011, United States

www.iastatedigitalpress.com

DOI: https://doi.org/10.31274/isudp.2026.215

Iowa State University is located on the ancestral lands and territory of the Baxoje (bah-kho-dzhe), or Ioway Nation. The United States obtained the land from the Meskwaki and Sauk nations in the Treaty of 1842. We wish to recognize our obligations to this land and to the people who took care of it, as well as to the 17,000 Native people who live in Iowa today.

Contents

CHAPTER 1.

THE MOLECULAR NATURE OF A GENE

Walter P. Suza; Faizo Kasule; Philip W. Becraft; and Madan K. Bhattacharyya

Introduction

A gene is the fundamental unit of **heredity** that is inherited from a parent. The hereditary material of all living organisms is composed of DNA (deoxyribonucleic acid). The structure of DNA must allow it to store coded information that controls the biological functions of cells. DNA transmits hereditary information in a stable form by replicating accurately, ensuring that each new cell receives the same genetic instructions. Although DNA can undergo changes (we will discuss how this occurs in Chapter 3), the **replication** process is highly accurate, allowing reliable inheritance of genetic information. In a human cell, the total length of DNA from all **chromosomes** is more than 6.5 feet. DNA is tightly packaged into structures called chromosomes to fit inside the cell nucleus.

Most genes provide instructions for producing proteins involved in multiple cellular functions. The first step in the transmission of hereditary information in a gene is called **transcription**. Transcription involves copying the DNA nucleotide sequence into an intermediate nucleotide molecule called RNA (ribonucleic acid). For protein-coding genes, the initial RNA transcript, known as precursor mRNA (pre-mRNA), is processed into a mature messenger RNA (mRNA). The mRNA carries the information needed to synthesize a protein through a process called **translation**. Proteins are composed of 20 different amino acids, each with unique chemical and physical properties, linked together by **peptide bonds**. The order of amino acids is dictated by the sequence of nucleotide bases in the mRNA, and this amino acid sequence ultimately determines the protein's structure and function.

Chemical Structure of DNA

Nucleotides

DNA is a polymer made of nucleotide subunits. A **nucleotide** consists of three chemical groups: sugar, phosphate, and a nitrogen-containing base (Figure 1). The sugar in DNA is deoxyribose, a five-carbon sugar derived from ribose (a pentose), with the hydroxyl group at the 2' carbon replaced by a hydrogen atom. The bases (Figure 2) are of two types: purines (adenine and guanine) and pyrimidines (cytosine and thymine). When a base is attached to a sugar, a **nucleoside** is formed; the addition of a phosphate group to the **nucleoside** forms a **nucleotide**.

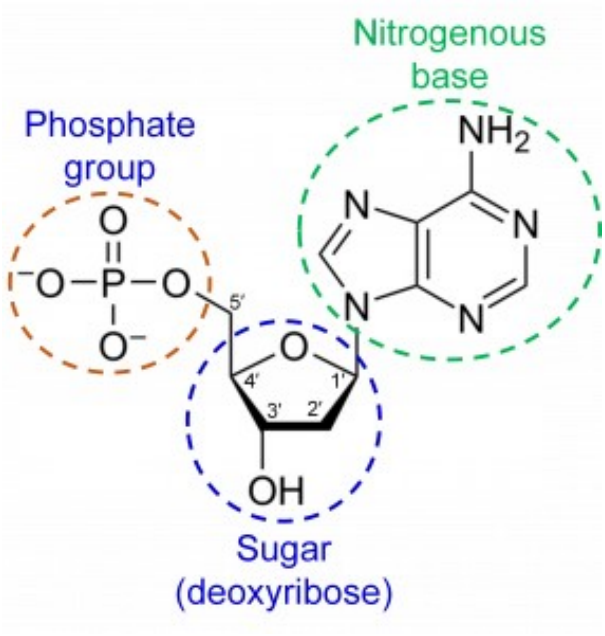


Figure 1 A nucleotide consists of three chemical groups. Image adapted from Hbf878, CC0, via Wikimedia Commons.

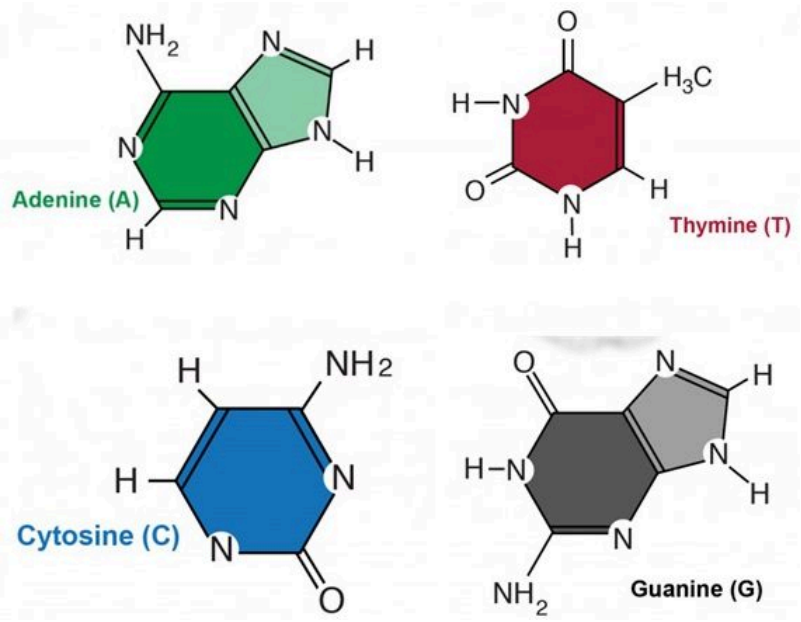


Figure 2 Adenine, Thymine, Cytosine, and Guanine are the nitrogenous bases of DNA. Adapted from NIH-NHGRI.

Sugar-Phosphate Bonds

Phosphate groups connect nucleotides to form a DNA strand. Phosphate groups create a linkage between the 5' carbon of the sugar in one nucleotide and the 3' carbon of the sugar in the next nucleotide. This type of linkage is known as **a phosphodiester bond**.

The carbon atoms on the deoxyribose sugar are numbered (Figure 1). 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 (Figure 3).

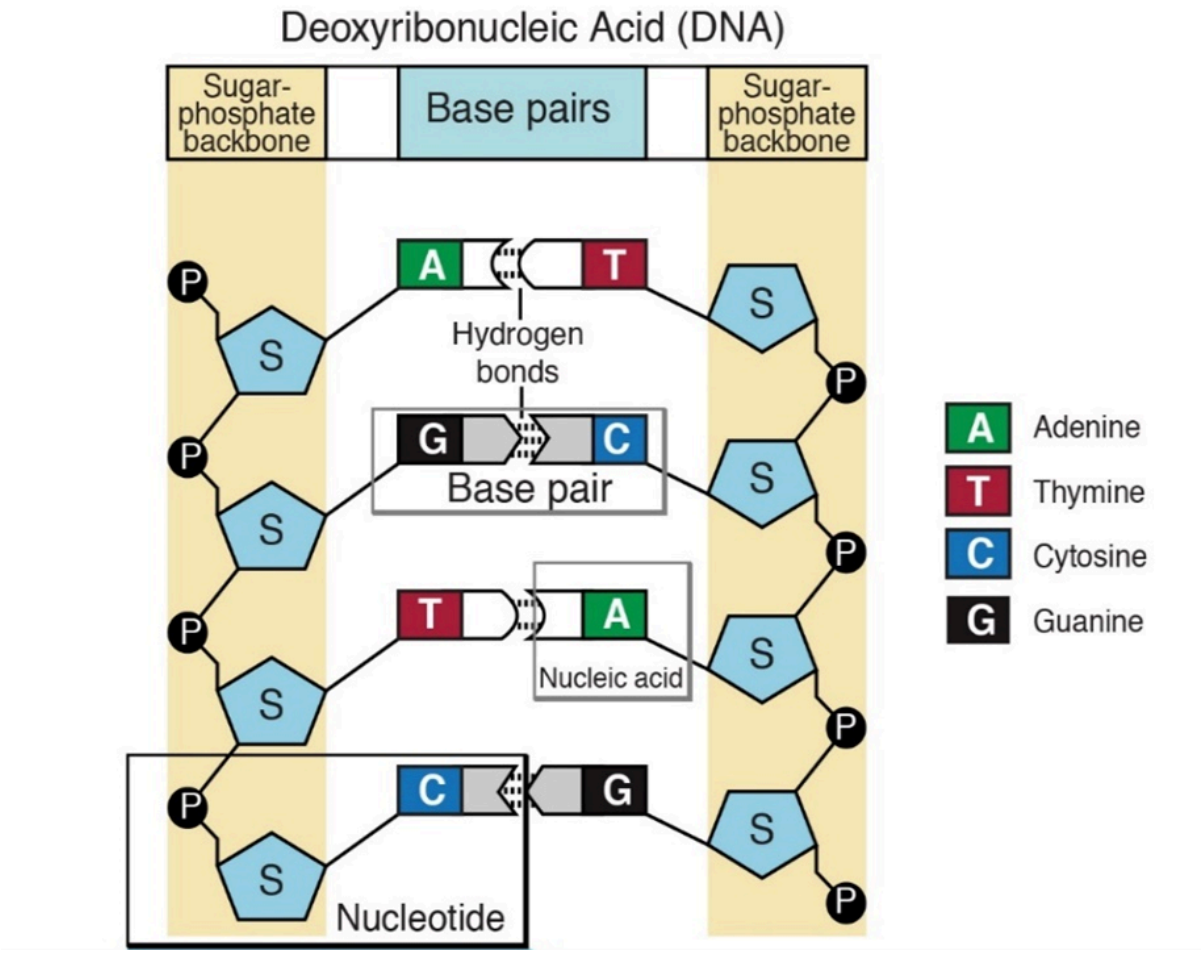


Figure 3 DNA is a double-stranded molecule. Illustration by NIH-NHGRI.

DNA Strands

DNA exists as a double-stranded molecule. The bases on one strand form hydrogen bonds with the bases on the other (Figure 3). This is called base pairing, which is highly specific with A, only pairing with T, and C only pairing with G. Thus, the sequences on the two DNA strands are **complementary.** Whenever A appears on a strand, T appears in the corresponding position of the complementary strand.

In addition to containing complementary nucleotide sequences, the DNA strands are in opposite orientations (Figure 3). 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. The sugar-phosphate backbone forms the outer framework of the double helix, made of alternating deoxyribose sugars and phosphate groups (Figure 3). At the same time, the nitrogenous bases are located inside the double helix, where they pair with bases of the opposite strand.

This helical twist allows DNA to have a compact and the most stable configuration, which accommodates all the molecular units and chemical bonds that make up the DNA, while still being accessible for processes like replication and transcription.

DNA Packaging

Inside the cell nucleus, DNA is assembled, or packaged, into a material composed of DNA and protein called chromatin. Chromatin protects and regulates DNA and efficiently stores the very long DNA molecules that make up chromosomes (Figure 4) within the limited space of the nucleus.

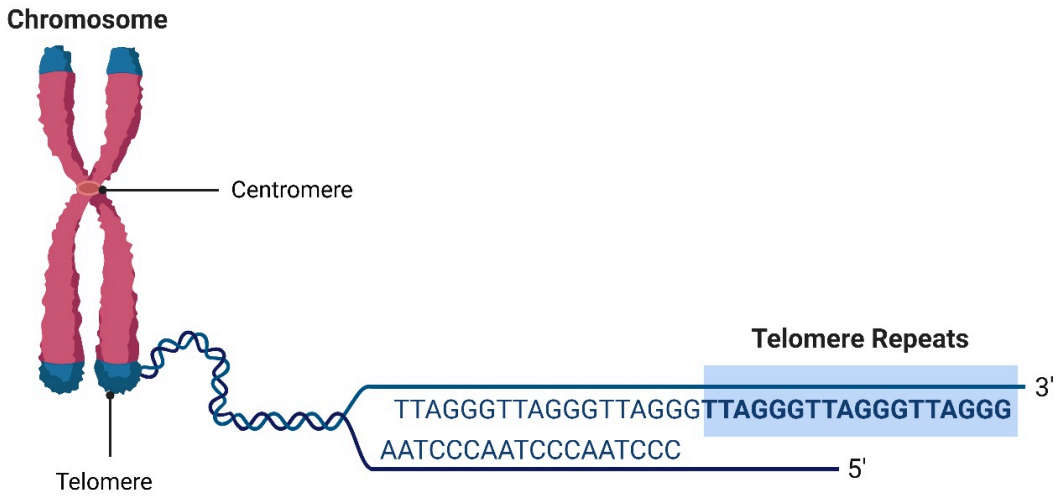


Figure 4 Schematic of a chromosome showing the centromere and the telomere. The expanded view of the telomere illustrates telomeric DNA repeats. Image modified by Faizo Kasule in BioRender, adapted from Calado et al. (2009).

Centromeres

The centromere (Figure 4) is a chromosomal region that controls chromosome segregation at **mitosis** and **meiosis**. Centromeres connect to microtubules of the spindle apparatus, which directs their movement to opposite poles (daughter cell nuclei) during cell division. Scientists successfully isolated these

centromeric sequences from yeast and engineered brewer's yeast (*Saccharomyces cerevisiae*) plasmids that could replicate like the chromosomes. Through these efforts, scientists were able to pinpoint centromeric function to a DNA stretch of about 120 bp that was resistant to DNase and bound to a single microtubule.

In the plant model species *Arabidopsis thaliana*, centromeres contain 178 bp satellite repeats occurring in tandem arrays ranging from 0.4–1.4 Mb. In contrast, in rice and maize, the amount of satellite DNA in centromeres is more variable and can range from 60 kb to 2 Mb on different chromosomes.

Recent studies have shown that minichromosomes, small chromosomes containing telomeres, replications, and a centromere with minimal extra genetic material, can be stably engineered to act as vectors for stacking genes, offering a powerful tool to study complex trait development in plant biotechnology and through Genetic manipulation of centromere-specific genes, e.g., CENH3 from Arabidopsis, facilitated the generation of haploid plants. This discovery may benefit plant breeders in the future by helping to speed up the generation of homozygous individuals.

Telomeres

The telomere (Figure 4) lies at the end of the chromosome and confers stability by "sealing" the end of a chromosome. Telomeres consist of a long series of short, repeated DNA sequences that occur 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 profoundly impacted molecular genetics. For example, yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) have proven useful in the physical mapping of plant genomes, requiring cloning and multiplying of large (>100 kb) DNA fragments in yeast or bacterial cells.

The rise in the use of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) systems by inducing double-strand breaks (DSBs) at any chromosome position has led to the creation of artificial chromosomes and, through this, spontaneous chromosome rearrangements, leading to genome evolution and crop domestication. Construction of artificial chromosomes for important crop species such as maize may provide novel options to improve food crops in the future.

For more information read: Construction and behavior of engineered minichromosomes in maize

Nucleosomes and Histones

DNA from higher eukaryotic organisms is about 10^9 to 10^{10} base pairs (bp). Human and maize chromosomal DNA is about 10^9 bp, which is equivalent to 1.8 m if all chromosomal DNA were stretched end to end in a linear manner. Yet the diameter of the nucleus is only about 4–6 μm, which makes it difficult to fit chromosomal DNA inside the cell nucleus. Fitting DNA inside the nucleus is accomplished by the ability of the DNA to assume a very condensed structure. Therefore, the organization of eukaryotic DNA into chromatin is an important aspect of DNA packaging.

The ordered coiling of chromosomal DNA around a histone protein core forms the chromatin (Figure 5). Chromatin consists of nucleosomes, representing chromosomal DNA's association with histone proteins. A nucleosome consists 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, with two molecules of each core histone to approximately 200 bp of DNA.

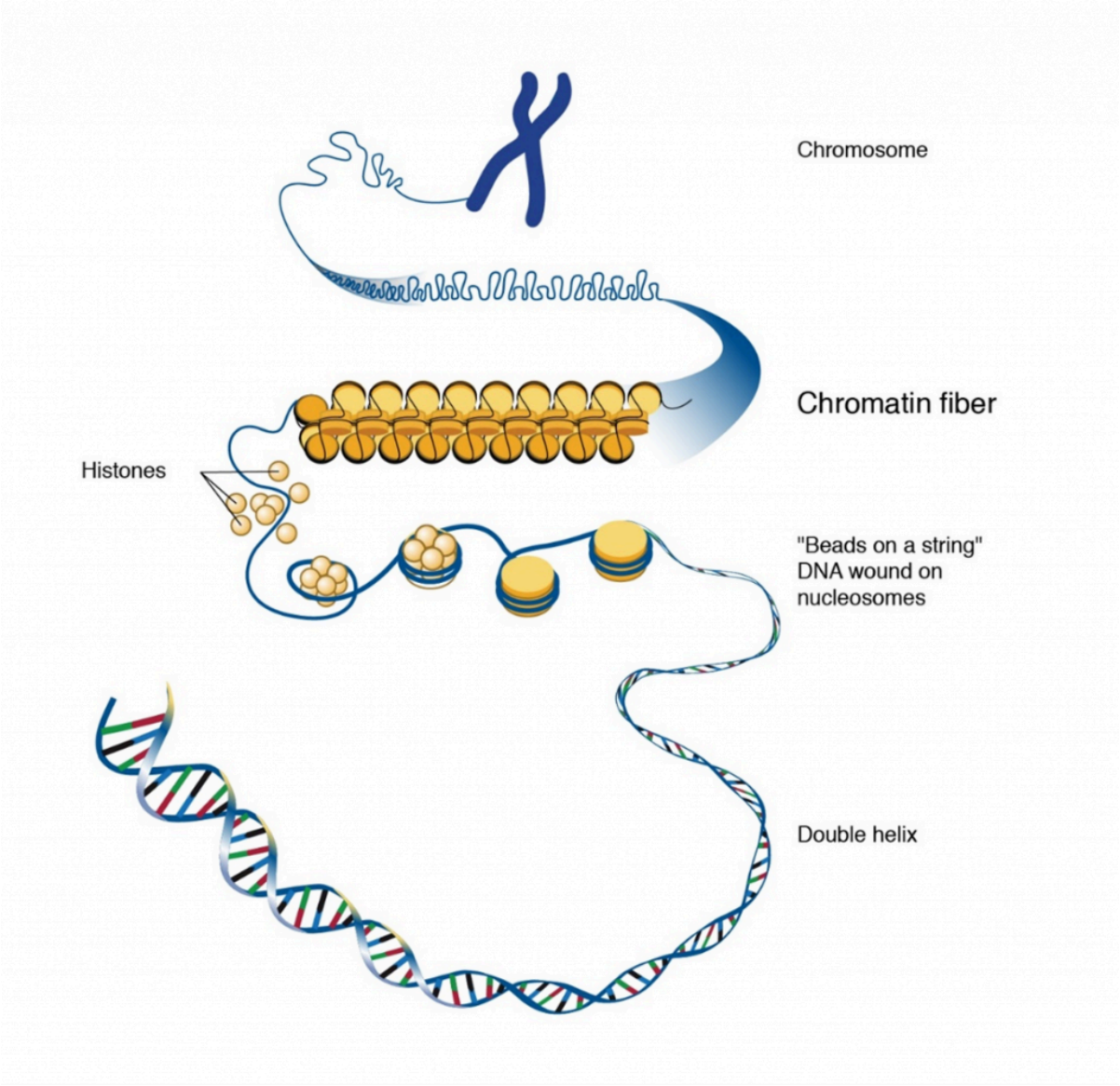


Figure 5 DNA is packed into chromatin and chromosomes. Illustration by NIH-NHGRI.

Higher Order Structures

If the higher-order chromatin structure is disrupted, electron microscopy reveals the appearance of "beads on a string" (Figure 5) 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.

Modifications of Chromatin Structure/Packaging

The tight packaging of DNA in chromatin must be modified to allow DNA replication and transcription.

The regulation of chromatin structure is complex and not completely understood, but is partly accomplished through the chemical modification (methylation and acetylation) of histone proteins within nucleosomes (Figure 6). It is important to know that the Polycomb Repressive Complex 2 (PRC2) mediates gene silencing in plants through the trimethylation of histone H3 at Lysine 27 (H3K27me3), leading to condensation of chromatin structure to form the heterochromatin. This restricts RNA polymerase II from accessing and transcribing the genes. Histone acetylation by histone acetyltransferases (HATs) using acetyl-CoA as an acetyl donor promotes open chromatin (euchromatin) and transcription activation (Figure 6).

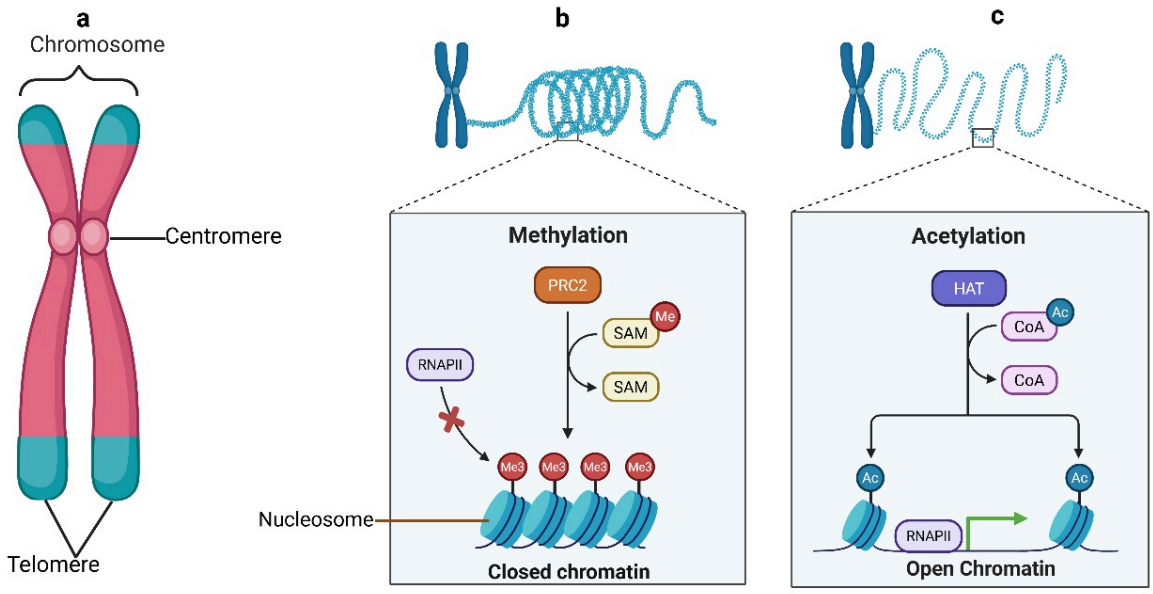


Figure 6 a) DNA methylation leads to closed chromatin (heterochromatin) and gene silencing mediated by methyltransferases and Polycomb Repressive Complex 2 (PRC2), which methylates histone H3 at lysine 27 (H3K27), a modification associated with gene silencing. S-adenosylmethionine is a methyl group donor that provides the methyl group that DNA methyltransferases add to DNA, and that PRC2 adds to lysine 27 on histone H3. **b).** Histone acetylation by histone acetyltransferases (HATs) using acetyl-CoA as an acetyl donor to promote open chromatin (euchromatin) and transcription activation. Created in BioRender by Faizo Kasule, adapted from Bannister & Kouzarides (2011).

The enzyme DNase has been used to compare the chromatin structure of DNA regions undergoing active transcription with those less transcriptionally active. Such DNAse assays (Figure 7) may be conducted on DNA isolated from two different tissues or the same tissue at different stages of development. The idea is that the more sensitive the DNA is to DNase, the more active the transcription in the tissue from which the DNA was isolated, and vice versa.

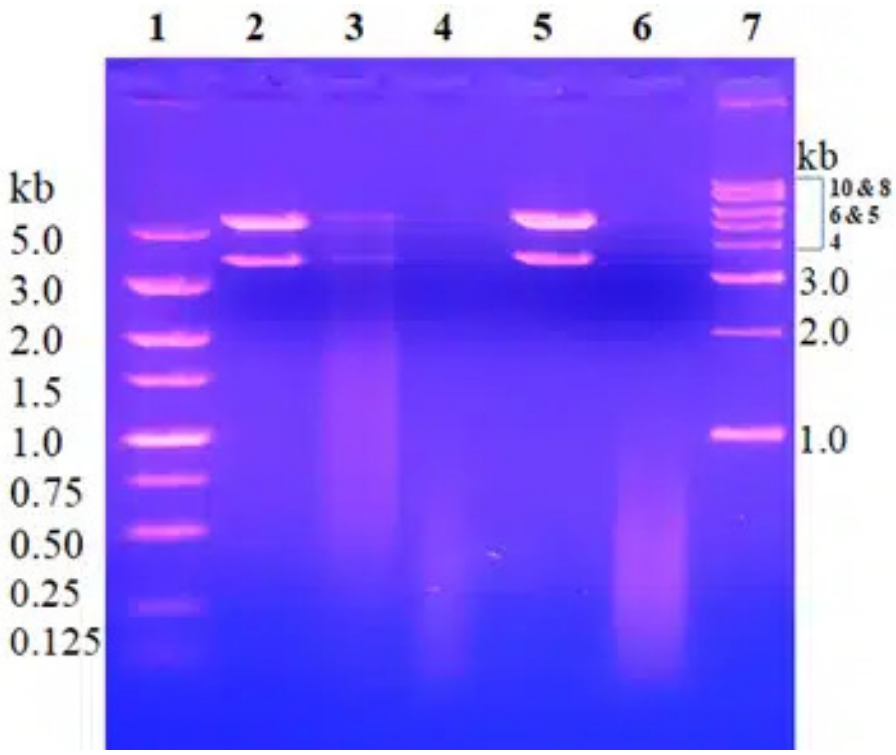


Figure 7 Result from a DNAse assay. DNA samples were incubated at 37 °C for 1 hour and analyzed on a 1.0% agarose gel using TAE buffer. Using reference lanes 1 and 7, and DNase-free lanes 2 and 5, the DNase-affected DNA in lanes 3, 4, and 6 can be examined. Agarose-gel electrophoresis was widely used for decades, prior to technological advancements such as colorimetric assays and PCR-based methods in DNA testing. Image from Boston BioProducts.

Central Dogma in Molecular Biology

Francis Crick (Figure 8), who co-discovered the structure of DNA with James Watson, proposed this fundamental theory in molecular biology in the late 1950s. Crick proposed that genetic information in any biological system flows from DNA to RNA to proteins. This ensures that the information stored in DNA can be accurately transcribed into RNA and then translated into proteins to perform cellular functions in the cell.

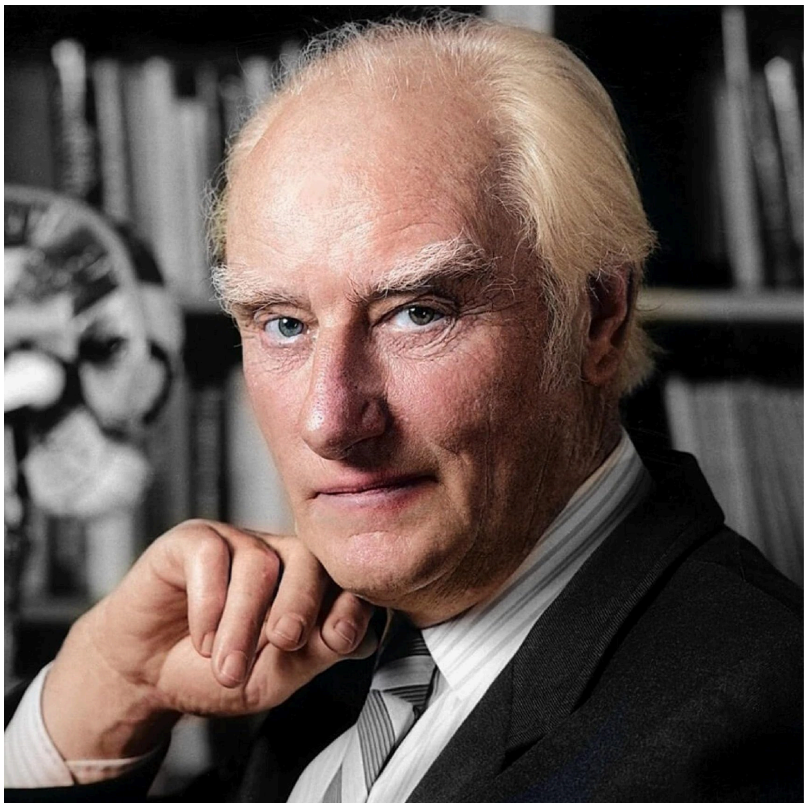


Figure 8 Francis Harry Compton Crick shared the 1962 Nobel Prize in physiology or medicine with James Watson and Maurice Wilkins for discovering the structure of DNA. Photo: Marc Lieberman, CC BY 2.5, via Wikimedia Commons.

Watch: This video about Central Dogma.

This is the process by which genetic information is accurately copied, providing the foundation for cell division and the development of organisms.

Semi-Conservative

In the replication process, the two DNA strands separate, each serving as a template for synthesizing **complementary** daughter strands (Figure 9). Therefore, the sequence of the daughter strand is determined by the sequence of the template strand. After replication, each DNA molecule is a duplex consisting of one parental (template) strand and one newly synthesized daughter (copy) strand. For this reason, it is referred to as **semi-conservative** replication and was first demonstrated by Matthew Meselson and Franklin Stahl in 1958. Thus, DNA replication is characterized by sequence conservation from one generation to another of the two individual strands on the parental duplex, a property that ensures the accuracy of genetic material reproduction.

Figure 9 Matthew Meselson and Franklin Stahl, standing at the site where they met at Woods Hole 42 years earlier. Photograph taken by F. L. Holmes.

In the experiment conducted in the 1950s, Mathew Meselson and Franklin Stahl showed that DNA replication is semi-conservative. They showed that by growing bacteria in a medium containing heavy nitrogen (15N), a stable isotope of 14N, which resulted in denser DNA. The bacteria were transferred to a medium containing light 14N, and their DNA was extracted and separated using a density gradient. After replication, each DNA molecule consisted of one original (heavy) strand and one newly synthesized (light) strand. This experiment confirmed the semiconservative model replication.

For more information read: The Meselson-Stahl Experiment (1957–1958), by Matthew Meselson and Franklin Stahl

Replication in Eukaryotes

DNA replication occurs in the S-phase of the cell cycle (Figure 10) to synthesize new genetic material for

cell division. In Eukaryotic organisms such as maize, DNA replication occurs in the nucleus. The process is complex and involves several enzymes and proteins in three main stages: initiation, elongation, and termination (Figure 11).

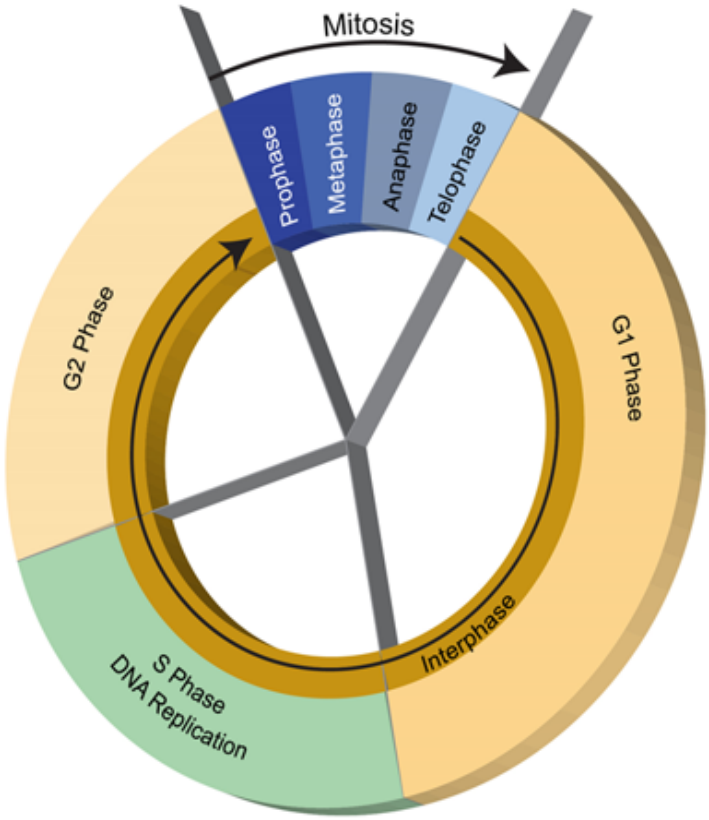


Figure 10 The cell cycle depicts the stages in the life of a cell. Image by NIH-NHGRI.

During initiation, DNA bound to histones in nucleosomes is made accessible to proteins and enzymes involved in replication (Figure 11). The process begins when an enzyme called DNA helicase unwinds the double helix by breaking hydrogen bonds between the two strands. The unwinding of DNA leads to the formation of two replication forks that extend in both directions to result in what is referred to as a replication bubble.

As DNA unwinds, it is forced to rotate and supercoil. The torsional strain created by the action of DNA helicase is relieved by topoisomerases. Topoisomerases introduce temporary breaks in the strands of DNA, and this allows replication to proceed without impediment. Type I topoisomerases break one strand, while type II topoisomerases cut both strands.

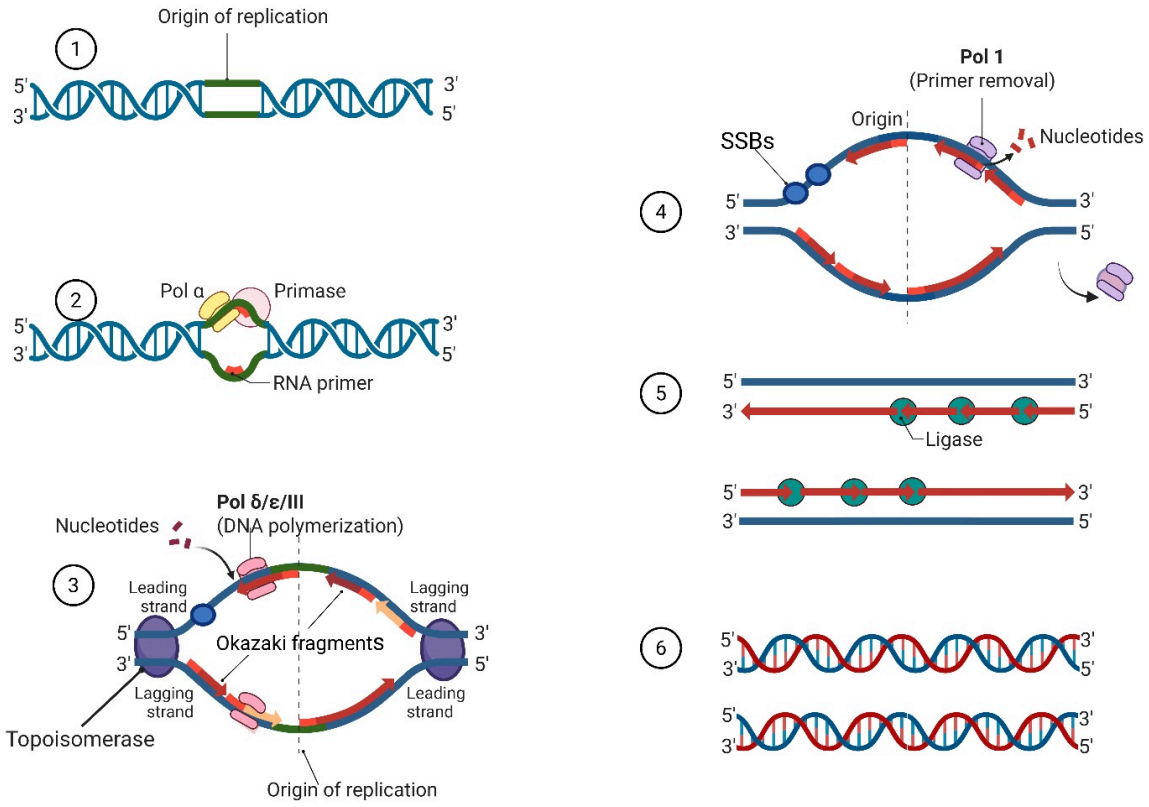


Figure 11 Semi-conservative DNA replication. Replication begins at the origin of replication (1), where the DNA strands separate to form a replication bubble (2). DNA polymerase III synthesizes new strands using each parental strand as a template by adding nucleotides to the 3' end (3). RNA primers are removed and replaced by DNA polymerase I (4). DNA ligase joins the Okazaki fragments on the lagging strand by forming phosphodiester bonds between them (5). At the end of this semi-conservative DNA replication, two DNA molecules, each composed of one parental and one newly synthesized daughter strand. Image Created by Faizo Kasule in BioRender, adapted from Lubischer (2007).

In elongation, the DNA polymerase synthesizes both leading and lagging strands by adding nucleotides in the 5' to 3' direction. The antiparallel DNA template strands are read in the 3' to 5' direction. However, DNA polymerase cannot initiate synthesis on its own. It requires a short RNA primer synthesized by the primase enzyme. DNA synthesis proceeds continuously on the leading strand in the direction of the replication fork. In contrast, the lagging-strand synthesis is discontinuous, resulting in short fragments referred to as Okazaki fragments, each starting with a new primer (Figure 11).

As mentioned above, DNA replication must be accurate to ensure the genetic information is correctly passed from parent cells to their daughter cells. Therefore, DNA polymerases δ and ε proofread each base added to the nascent strand to correct errors that may occur during elongation.

DNA polymerase halts when it encounters another replication fork and replicates the template, and drops off. After elongation, the RNase H enzyme removes RNA primers, and DNA polymerase replaces the gaps with DNA. DNA ligase then joins fragments by forming phosphodiester bonds between them, com-

pleting the backbone of the new strand. Replication is terminated when two replication forks meet. However, a small portion of the lagging strand may remain unreplicated, which could lead to shortening of chromosomes. To counter this, the telomerase enzyme extends chromosome ends by adding repetitive DNA nucleotides, ensuring that essential genes are not lost during cell division.

The Polymerase Chain Reaction (PCR)

Knowledge about DNA replication in living cells led to the innovative development of the Polymerase Chain Reaction (PCR), which is a versatile technique used routinely in the laboratory to amplify specific regions of the DNA template. A detailed description of the PCR technique will be provided in Chapter 6.

Essentially, PCR requires template DNA, primers, nucleotides, and cofactors, and is conducted in a thermal cycler. The enzyme DNA polymerase uses single-stranded DNA as a template for the synthesis of new strands of DNA complementary to the template. PCR relies on short synthetic DNA fragments called primers to define the segment of the genome to be copied. Through multiple cycles of DNA synthesis (denaturation, primer annealing, and strand extension), the selected region is amplified, producing many copies of the target sequence. This process produces many copies of the sample DNA, which can be used for analyses such as screening for a particular nucleotide base sequence associated with a trait. There are various types of PCR, including real-time PCR, which allows rapid and real-time detection using fluorescent probes, and reverse transcription PCR (RT-PCR), which converts RNA into DNA for amplification, and has been crucial in gene expression studies. After performing PCR, the results can be visualized via gel electrophoresis.

Transcription and Gene Structure

The DNA sequence contains information to control all biological functions, including the manifestation of traits important to agriculture (yield, drought tolerance, disease resistance, etc.). How is the information contained in DNA sequences converted into the cellular activities necessary for plants and other organisms to function? DNA sequences direct the synthesis of other molecules that perform these cellular functions through an RNA intermediate. The process of producing RNA molecules from DNA templates is called **transcription**. Some RNA molecules perform cellular functions directly, while many others (messenger RNAs) are used to direct the synthesis of proteins in **translation**.

Gene Structure

Only specific regions of chromosomal DNA undergo transcription. A **gene** can be defined as the region of a chromosome that is transcribed, together with associated DNA sequences that regulate transcription (Figure 12). Gene regulation is discussed further in Chapter 2.

Figure 12 Simplified gene structure with key functional regions.

Promoter and Terminator

The instructions for starting and stopping transcription are located within DNA sequences. Specific nucleotide segments called promoters are recognized by RNA polymerase for RNA synthesis. The promoter is located around the first nucleotide (known as +1) that is transcribed into RNA. After the transcription of the full-length RNA strand is completed, a second segment of DNA called a terminator invokes termination of RNA synthesis and the detachment of RNA polymerases from the DNA template.

The CaMV 35S promoter from the cauliflower mosaic virus (CaMV) is commonly used in genetic engineering because it is a strong promoter that directs high transcription levels. A terminator sequence, such as the nopaline synthase (NOS), is also commonly used to mark the end of the transgene sequence for proper transcription and expression in plant cells.

Transcribed Region

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

Introns and Exons

As mentioned above, and described in detail below, **introns** are sequences 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.

Protein Coding Region

The protein-coding region of a gene is composed of the sequence of nucleotides that code for **amino acids** (Figure 13). 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 consist of **untranslated regions**.

Figure 13 Protein-coding region of a gene.

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; therefore, sequence variation is relatively rare. UTRs and introns, on the other hand, are under less stringent selection and are the sources of sequence variants that can be used to develop genetic markers.

Transcription

Genetic information of DNA is transferred to an intermediate molecule called RNA, which is often translated into amino acid sequences used to build proteins. RNA is a nucleic acid, like DNA, but with some important differences. RNA contains ribose sugar, which has an OH group at the 2' position, while deoxyribose is found in DNA, has a hydrogen at the same position. RNA molecules are primarily 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.

A DNA template directs RNA synthesis in a process called **transcription**. A protein complex containing the enzyme **RNA polymerase** synthesizes an RNA molecule by adding nucleotides 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 DNA template strand, also referred to as “antisense” or “negative” strand. The RNA is identical in sequence (except U replaces T) to the other strand, which is called the “sense” or “positive” strand. Because transcription produces RNA molecules, they are often referred to as **transcripts**.

Transcription in eukaryotes occurs in the nucleus and is carried out by three polymerase enzymes (RNA

polymerase I, II, and III). These enzymes share subunits and structural features but are known to transcribe different types of genes. RNA polymerase I transcribes genes for ribosomal RNA (rRNA), RNA polymerase II transcribes most genes, including protein-coding genes, and we shall focus on this polymerase for this section, while RNA polymerase III transcribes transfer RNA (tRNA) and other small RNAs.

Like replication, transcription in eukaryotes occurs in three stages: initiation, elongation, and termination. Eukaryotic RNA polymerase II requires multiple proteins called general transcription factors, collectively known as TFII (Transcription Factor RNA polymerase II), to assemble at the promoter region for the polymerase to start transcription during initiation. Nucleosomes are also unpacked during this stage to make DNA accessible to RNA polymerase II. With the help of several transcription factors, RNA polymerase binds to a region upstream of the gene called referred to as the promoter, often containing a TATA box, usually 23-35 base pairs before the transcription start site, with the help of the TATA-binding protein (TBP) subunit. This distorts the DNA by creating a bend, and this facilitates the binding of TFIID to the TATA box via TBP. TFIIA binds to stabilize the TFIID-TATA box interaction. Other TFIIs are recruited like TFIIA, TFIIB, TFIIF, TFIIE, TFII, and TFIIH (Figure 14).

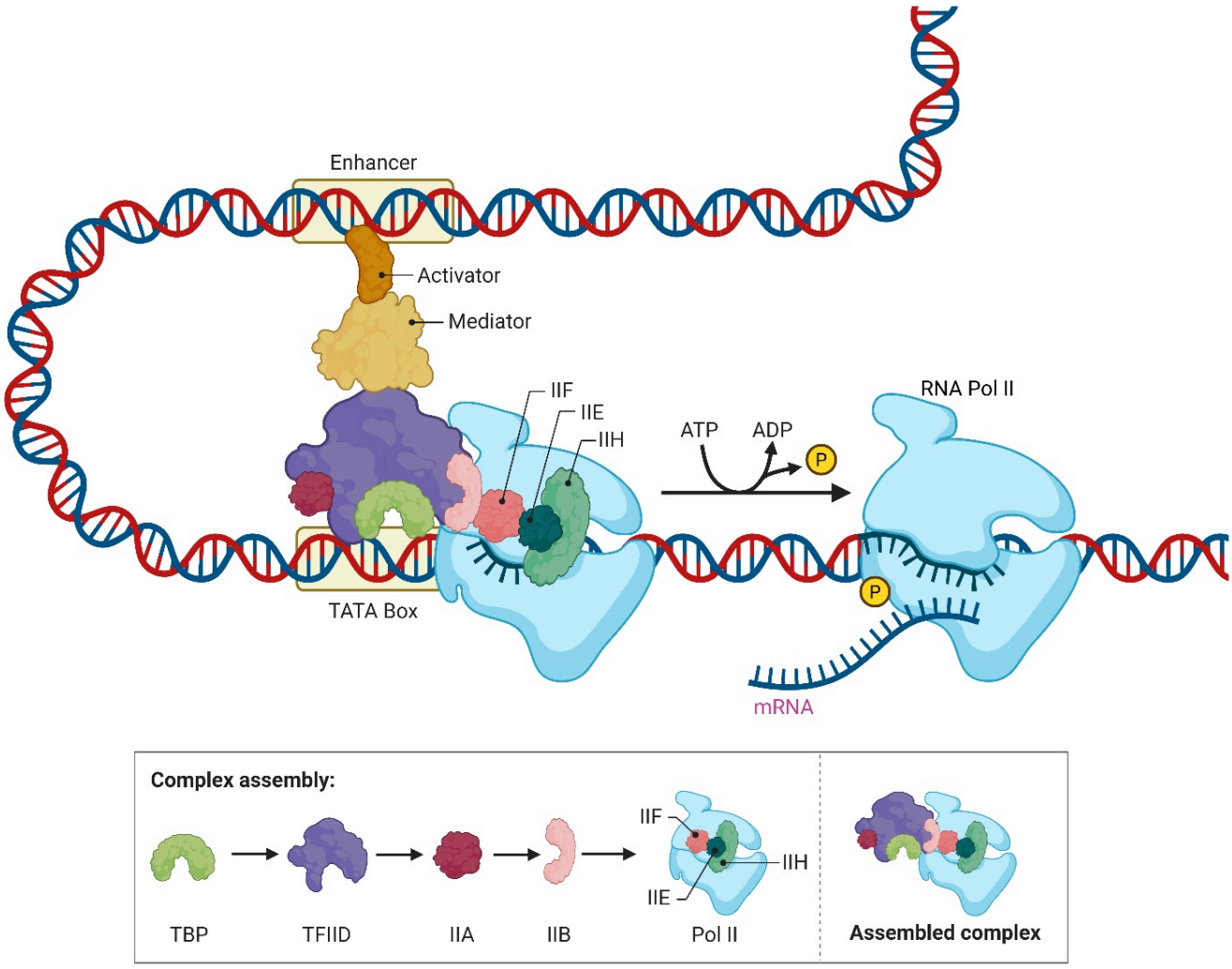


Figure 14 The preinitiation complex (PIC) formation: general transcription factors bind to a promoter region on DNA to initiate transcription. The preinitiation complex normally contains RNA polymerase II and six general transcription factors: TFIID, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH. The process starts with TFIID binding to the promoter with the help of the TATA-binding protein (TBP) subunit. This bends the DNA with the help of TBP-associated factors. The binding of *TFIID is followed by TFIIA and then TFIIB.TFIIA interacts with the DNA and TFIID, and TFIIB binds TBP and the RNA polymerase II. TFIIF binds tightly to RNA polymerase II, and finally, TFIIE and TFIIH join to complete the complex. An enhancer region in DNA recruits transcriptional activators that stabilize promoter-binding through DNA looping and interactions with a mediator. Unphosphorylated C-terminal domain (CTD) of RNA Polymerase II binds to mediators, whereas phosphorylation of CTD releases mediators and stimulates promoter release. Image modified by Faizo Kasule from a BioRender template created by Shadma Nafis, adapted from* Cooper (2000).

Unlike DNA polymerase, which requires a primer to start replication, RNA polymerase does not need a primer to initiate RNA synthesis (Figure 15). During the elongation phase, once bound to the DNA, TFIIH acts as a DNA helicase to unwind the double helix to expose the template strand. This also has a kinase activity that phosphorylates the C-terminal domain (CTD) of RNA polymerase II, leading to a conformational change that signals RNA polymerase II to start elongation. Once this process begins, most TFIIs are released to make them available for new transcription rounds with a new RNA polymerase.

The RNA polymerase II moves along the DNA in the 3' to 5' direction, synthesizes complementary RNA in the 5' to 3' direction through nucleophilic attack of the 3'-OH group to the growing RNA chain on the α-phosphate of the incoming ribonucleotide triphosphate, and rewinds the DNA behind it. This produces a primary transcript called pre-messenger RNA (pre-mRNA) (Figure 15). The RNA polymerase II transcription termination signal is perceived when the enzyme transcribes 1000-2000s nucleotides beyond the gene's end. Termination is triggered, causing RNA to be cleaved between the AAUAAA sequence and a downstream GU-rich region.

Unlike DNA replication, RNA polymerase lacks proofreading activity; however, the cell compensates through RNA processing, transcript surveillance, and various regulatory systems like selective degradation of defective RNA that ensure transcript fidelity and proper function.

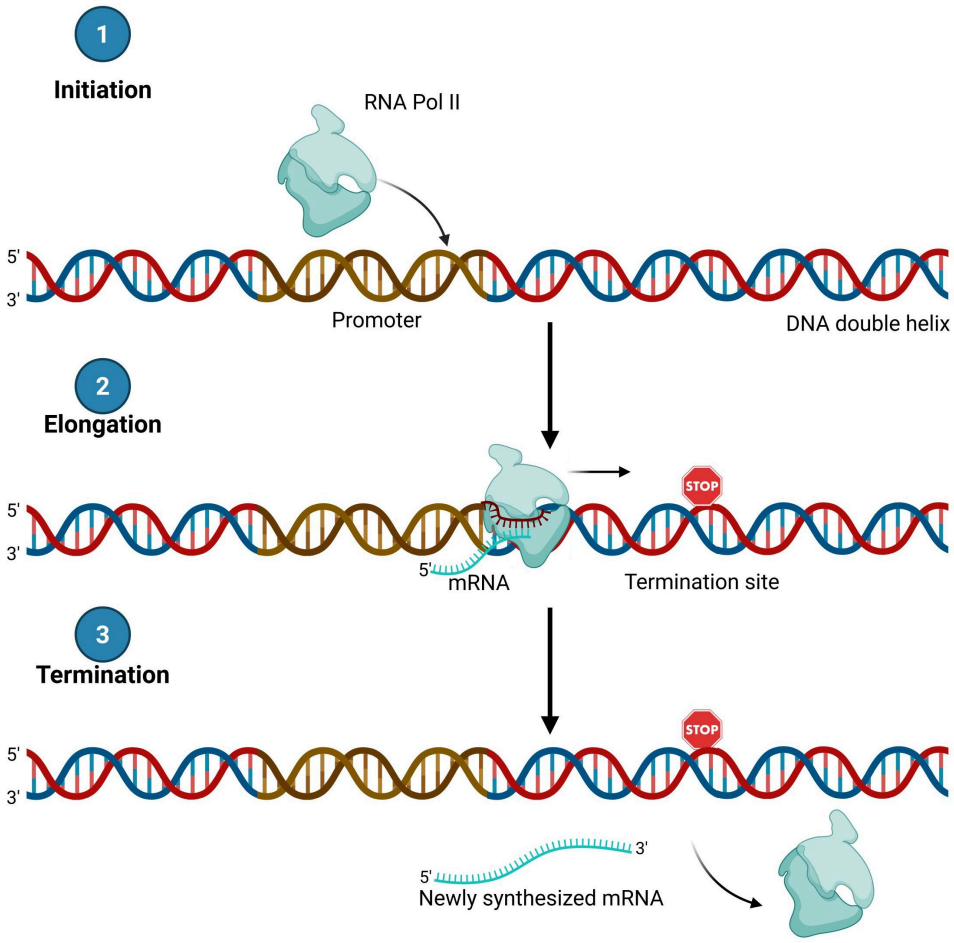


Figure 15 The three stages of transcription: initiation (1), elongation (2), and termination (3). During **initiation,** general transcription factors help RNA polymerase II recognize and bind to core promoter elements like the TATA box, upstream of the gene. This recruitment positions the RNA polymerase II at the transcription site, and the initiation complex unwinds the DNA to expose the template strand. **Elongation**, the RNA polymerase II synthesizes nascent mRNA in the 5' to 3' direction by moving in the 3' to 5' direction along the DNA template. **Termination**, RNA polymerase II counters polyadenylation signals and downstream terminator sequences, which trigger protein factors associated with the polymerase to release the newly synthesized RNA, and the RNA polymerase II dissociates from the DNA template, completing transcription. *Image created in BioRender by Faizo Kasule, adapted from Kuehner et al. (2011).*

RNA Processing

As described above, exons are protein-coding (translated) sequences, while introns are intervening sequences that interrupt exons. The removal of introns, along with the addition of a 5' cap and a 3' poly-A tail to the mRNA, is called **RNA processing**. A simplified illustration of transcription and RNA processing is shown in Figure 16. The detailed mechanisms of these processes are discussed below.

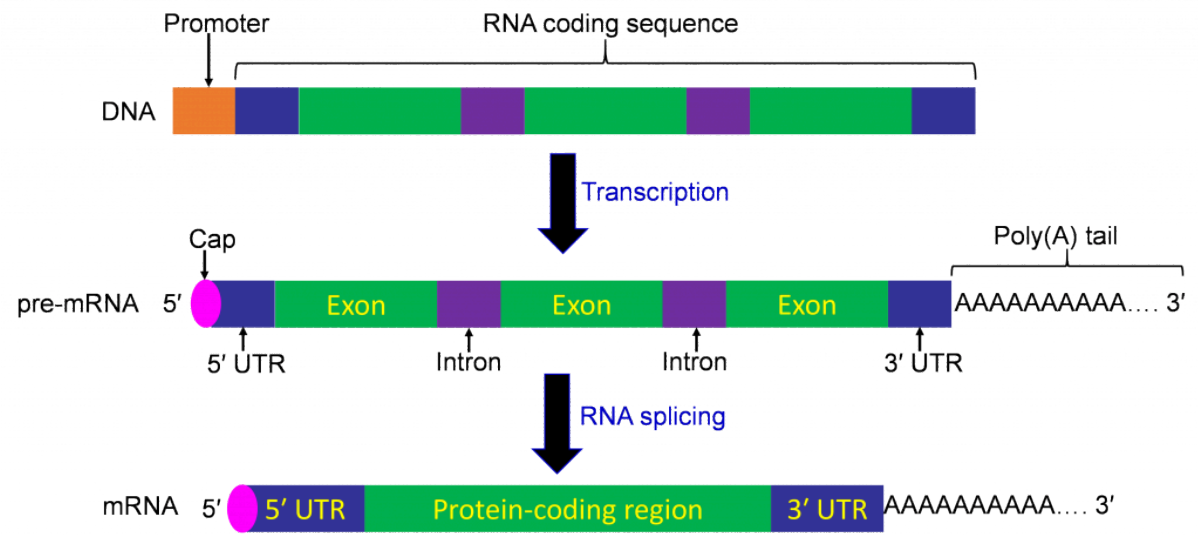


Figure 16 Simplified illustration of transcription and RNA processing.

Intron Splicing

The process of transcription produces pre-mRNA that contains both introns and exons. The process of splicing involves the removal of introns from pre-mRNA and the joining together of the exons (Figure 17). A spliceosome, a large ribozyme complex made up of a group of RNA and proteins performs the splicing reaction on the C-terminal domain (CTD) of RNA polymerase II. In yeast, the spliceosome has about 90 proteins, while in humans 170 proteins. This complex functions with the help of small nuclear ribonucleoproteins (snRNPs), i.e., U1, U2, U4, U5, and U6, which consist of about 150 nucleotides and other secondary structures and a core of the seven Sm proteins, which form a ring and bind the conserved AAUUUGUGG Sm site.

For splicing to occur, the spliceosome recognizes three splicing signals, i.e., the 5' splice donor site, the branch point sequence, and the 3' splice acceptor site. The process involves several steps, and it starts in complex E, where U1 binds to the 5' splice site, which is the commitment step that defines the intron-exon boundaries. This is followed by the binding of U2 to the branch point in complex A, which displaces the Adenosine nucleotide for nucleophilic attack. Then complex B will form when U6 replaces U1 at the 5' splice site, which causes U4 to base pair with U6, as a regulator. This then causes U5 to align the exons. Upon U4 dissociation, this allows U6 to base pair with U2, and this allows catalysis by binding Mg^{2+} ions; these stabilize the negative charge on the phosphate backbone and activate the nucleophile, i.e., the OH group, and create a catalytic center with U2 and the branch point. Two transesterification reactions follow this: First, the 2' OH group of the branch point Adenosine nucleophilic attacks on the 5' splice site. This cleaves the 5' exon from the intron, forming a lariat structure where the intron will loop and connect to the branch point Adenosine via a 2'-5' phosphodiester bond. Second, the newly freed 3'OH of the 5' exon attacks the phosphate at the 3' splice site. This process joins the two exons together and releases the intron.

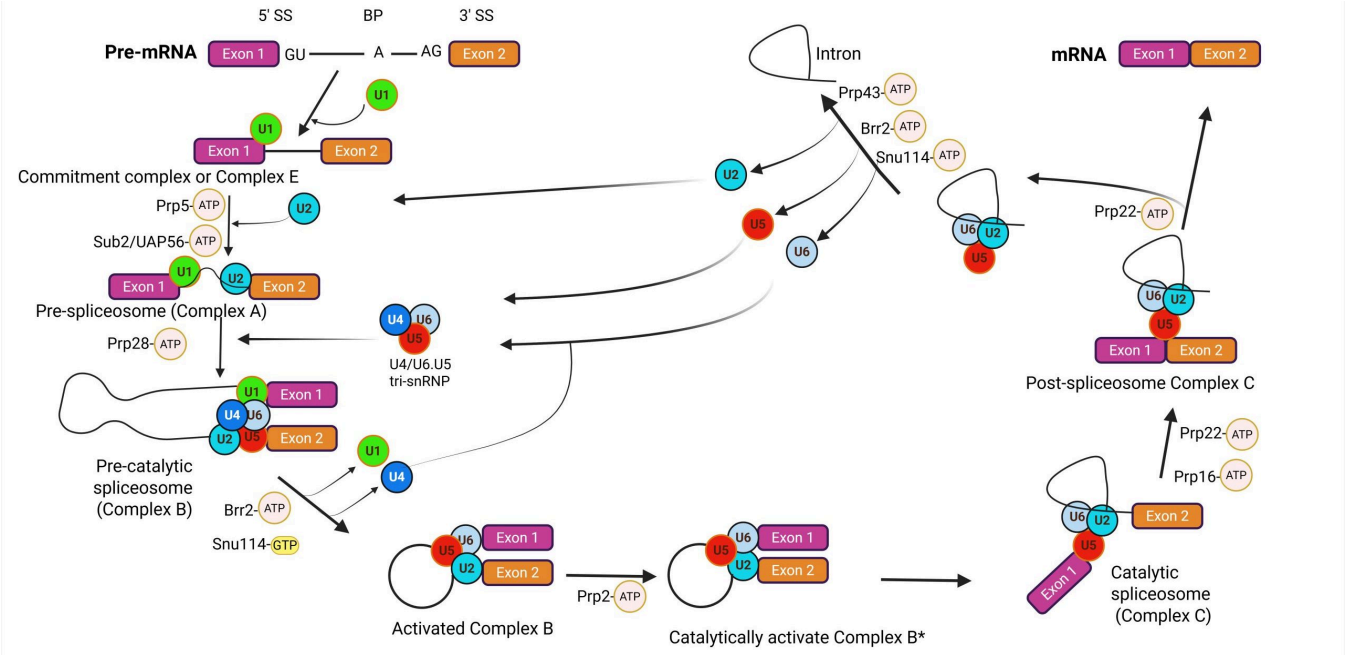


Figure 17 Pre-mRNA splicing. **E Complex formation,** which involves U1 snRNP binding the 5' splice site (5 ' ss); SF1/mBBP recognizes the branch point sequence (BPS); U2AF binds the polypyrimidine tract and 3 ' splice site (3 ' ss). **A Complex formation** involving Prp5 promotes U2 snRNP recruitment to the BPS, displacing SF1 and forming the A (pre-spliceosomal) complex. **B Complex formation,** which involves the U4/U6·U5 tri-snRNP, is recruited. Prp28 facilitates the release of U1 from the 5'ss, allowing U6 to bind. **Spliceosome activation,** Brr2 unwinds the U4/U6 duplex, releasing U4. U6 pairs with U2 to form the catalytic core. The spliceosome becomes activated. First Catalytic Step (B Complex) involves Prp2 and its cofactor Spp2 remodels the spliceosome to the B* complex. The branch point adenosine attacks the 5'ss, forming a lariat and freeing the 5' exon. The **Second Catalytic Step (C Complex)** involves Prp16 facilitates transition to the second step. The 3'-OH of the 5' exon attacks the 3'ss, ligating the exons and releasing the lariat intron. **mRNA Release involves** Prp22 promotes the release of the mature mRNA from the spliceosome. **Spliceosome Disassembly** involves Prp43, assisted by Ntr1 and Ntr2, disassembling the spliceosome, recycling snRNPs, and releasing the intron lariat. *Image created in Biorender by Faizo Kasule, adapted from Will & Lührmann, 2011.*

Introns sometimes serve as boundaries for sequences encoding functional protein domains, leading to the possibility of new and variant proteins by exon shuffling. Also, introns can produce variant RNA forms through alternate splicing, allowing more than one gene product from a single gene. This is because the spliceosome commitment depends on whether exon or intron recognition sequences are disrupted, often leading to exon skipping or intron retention. It is important to note that alternative splicing is very common in eukaryotes, with 75% of human transcripts undergoing this process. This, therefore, affects protein localization, activity, and function. Some introns result from the insertion of transposable elements and may be spliced perfectly or imperfectly, offering more possibilities for new genetic diversity.

For more information read: Alternative Splicing within and between Drosophila Species, Sexes, Tissues, and Developmental Stages

For more information read: Differential isoform expression and alternative splicing in sex determination in mice

5' Capping

Once RNA polymerase has synthesized about 25 nucleotides of RNA, the 5’ end of the new RNA is modified by the addition of a “cap” (Figure 18). The 5' capping adds a 7-methyl guanidine to the first nucleotide of the mRNA molecule, usually an adenine or guanine. The phosphodiester linkage between 7-methyl guanidine and the target nucleotide is 5'-5' instead of 5'-3', and 3 phosphates rather than one are retained in the linkage. The capping process involves three enzymes acting in order: a phosphatase removes a phosphate from the 5 ' end, a guanyl transferase adds guanosine monophosphate (GMP) in orientation, and a methyl transferase methylates the added guanosine. These enzymes are bound to the phosphorylated RNA polymerase tail, enabling rapid capping as nascent RNA is produced. The cap stabilizes the 5' end of the mRNA, protecting it from degradation in the cytosol, and plays a role in translation initiation.

Poly Adenylation

The transcription of a gene may proceed beyond what ends up as the 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, resulting in the poly(A) tail (Figure 18). The poly(A) tail is necessary to process and transport mRNA to the cytoplasm properly. The poly(A) tail is also important for mRNA stability by protecting against exonucleolytic degradation and initiating translation in eukaryotic organisms by helping Ribosomes recognize mature mRNA.

In eukaryotes, cleavage and polyadenylation are two important processes that control the formation of the mRNA 3' end in the nucleus. RNA polymerase II will transcribe beyond the location where the mature mRNA will terminate, about 20 nucleotides upstream of the cleavage site. This is usually a conserved polyadenylation site, often AAUAAA, followed by a pre-mRNA cleavage point downstream. Usually, a GU-rich region is downstream of this site to improve the efficiency of cleavage, and following cleavage, an enzyme known as poly(A) polymerase (PAP) adds the poly(A) tail of about 250nts to the 3' end. The cleavage and polyadenylation involve many protein complexes which coordinate these processes; cleavage and polyadenylation specificity factor (CPSF) binds the AAUAAA signal, cleavage stimulation factor (CstF) that the GU-rich region, and cleavage factors I and II (CFI and CFII) participate in cleavage. Another protein, Poly(A) Binding Protein II (PABII), binds to the Poly(A) tail to promote tail elongation and stability. All these processes are coordinated on the C-terminal domain (CTD) of RNA polymerase II. The Poly(A) tail helps in the nuclear export of mature mRNA transcripts

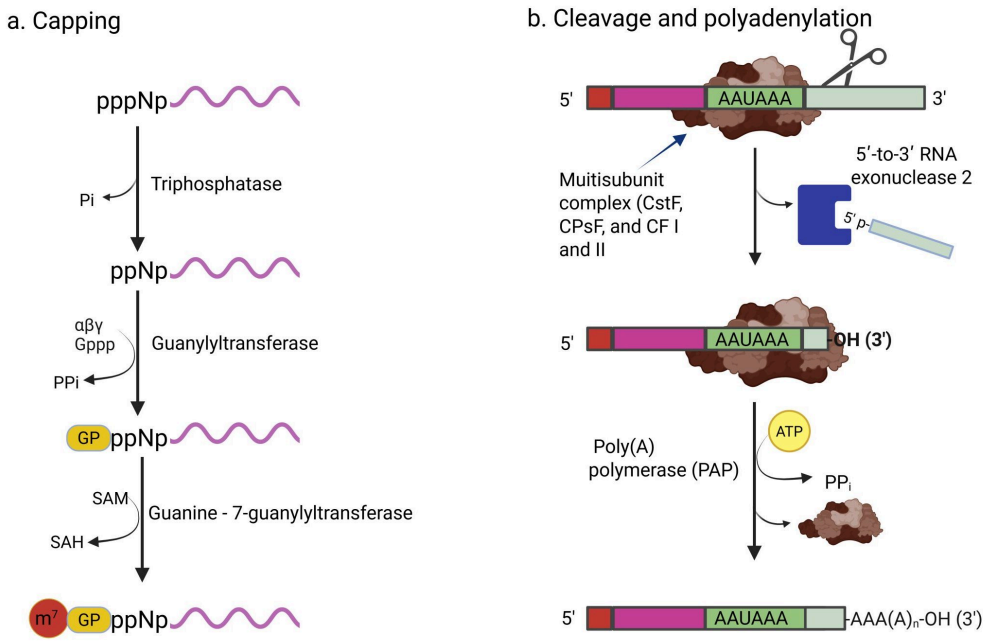


Figure 18 Steps in mRNA processing: (a) 5' capping and (b) cleavage followed by polyadenylation. **5' capping (a)**; The RNA is shown in pink; The mRNA-capping enzyme is bifunctional and has both triphosphatase and guanylyl-transferase activities that remove the γ-phosphate of the nascent transcript and transfer GMP from the GTP donor, respectively. The methyl donor *S*-adenosyl-l-methionine (SAM) is converted to S-adenosyl-l-homocysteine (SAH), which results in the 7-methylguanosine cap (shown in red). During **3' cleavage and polyadenylation (b)**, 3' ends of mRNAs are formed by coupled cleavage and polyadenylation. Cleavage of mammalian pre-mRNAs occurs ∼25 bases downstream of a consensus sequence (AAUAAA) and is carried out by the multisubunit complex (shown in dark and light brown), which comprises cleavage stimulation factor (CstF), cleavage and polyadenylation specificity factor (CPSF) that bears the endonuclease, and cleavage factors I and II (CFIm and CFIIm). Poly(A) polymerase (PAP) adds the poly(A) tail. 3' ends of non-polyadenylated histone mRNAs (not shown) are also made co-transcriptionally by a cleavage complex that has many subunits in common with CstF and CPSF. The 5'-to-3' RNA exonuclease 2 (XRN2) degrades RNA downstream of the cleavage site and facilitates transcription termination. *Image created in BioRender by Faizo Kasule, adapted from Bentley, 2014.*

For more information read: A Plant Poly(A) Polymerase Requires a Novel RNA-binding Protein for Activity

Codons and the Genetic Code

The search for the genetic code revealed that genetic information is stored in nucleotide triplets called **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 amino acids but specify the termination of peptide chain synthesis during translation. The AUG **start codon** initiates polypeptide synthesis and codes for the amino acid methionine.

Ribosomes, tRNA, and Anti-codons

The ribosomes are a large molecular complex that assembles 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. Before initiating protein synthesis, the ribosome occurs in two separate subunits of size 60s and 40s (The “s” stands for Svedberg units and is a measure of a particle’s sedimentation rate during centrifugation).

The tRNA determines the meaning of a codon for a specific amino acid (transfer RNA). Each transfer RNA 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 checks for accuracy in protein synthesis by adding a specific amino acid to a specific tRNA molecule. Thus, a particular amino acid is targeted to each codon triplet of mRNA.

Translation

The process of translation in eukaryotes is conserved and occurs in the cytoplasm on ribosomes with the help of several translation factors and takes place in 3 steps: initiation, elongation, and termination.

Translation initiation

A pre-initiation phase in eukaryotes starts with the charging of transfer RNA (tRNA), which involves amino acid activation using ATP. It is attached to its specific tRNA by an aminoacyl-tRNA synthetase. This forms an aminoacyl-tRNA complex. In eukaryotes, translation starts with methionine delivered by the initiator tRNA (Met-tRNAi), different from internal methionine. The 43S pre-initiation complex (43S PIC) assembles when the 40S small ribosomal subunit binds to multiple eukaryotic initiation factors (eIFs) (Figure 19).

This process starts with the binding of eIF3 to the solvent surface side of 40S, followed by the binding of eIF1 and eIF1A, which causes a confirmation change that opens the mRNA entry channel. In parallel, the eIF2 with GTP bound then binds to the Met-tRNAi, forming a **ternary complex**, which together with eIF5 joins the 40S to complete the 43S complex. During this, mRNA is prepared for loading by the actions of the eIF4F complex, which includes eIF4E that binds the 5' cap, eIF4A, a helicase, and eIF4G. This scaffolding protein bridges mRNA to the ribosome and interacts with Poly(A) binding protein to circularize the mRNA. The resultant mRNA-eIF4F complex is recruited to the 43S complex, forming the 48S pre-initiation complex.

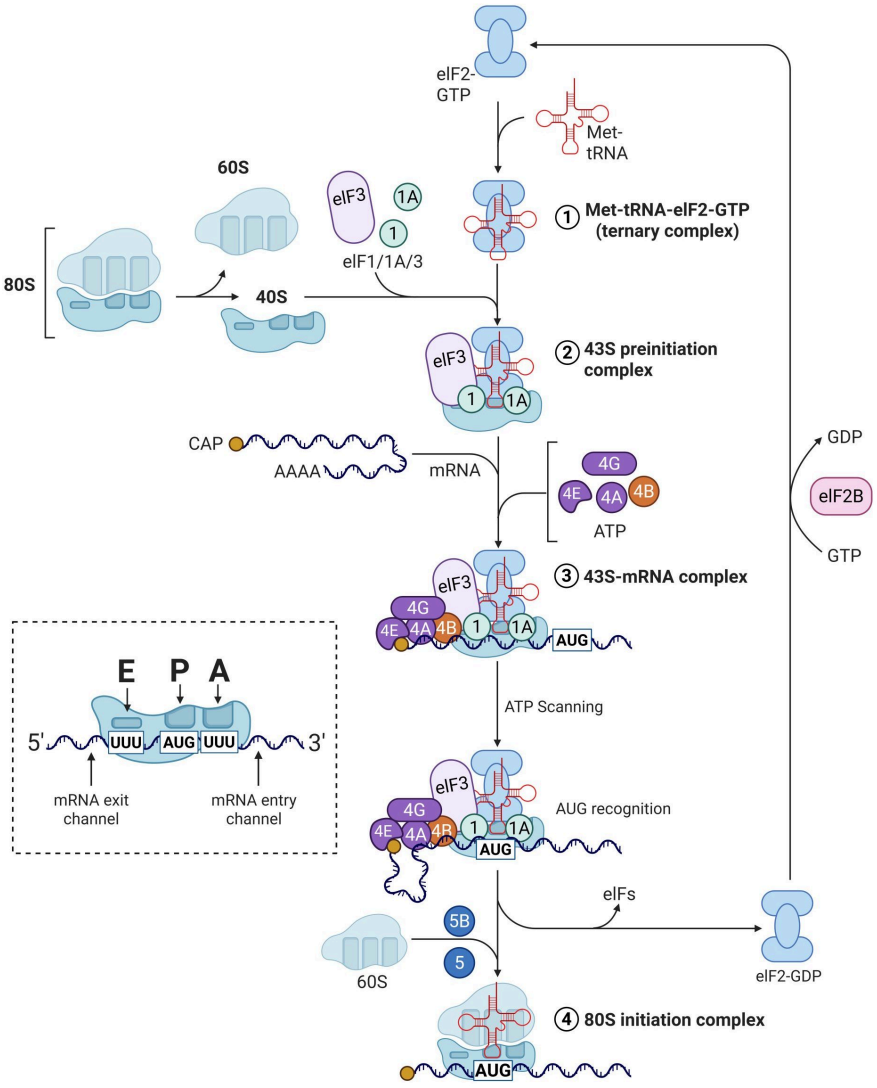


Figure 19 Process of translation initiation in eukaryotes. **Step 1**: Formation of the Met-tRNAi–eIF2–GTP ternary complex, where the initiator tRNA carrying methionine binds to eIF2 loaded with GTP, creating a complex ready to deliver the initiator tRNA to the ribosome. **Step 2:** The 43S pre-initiation complex is then assembled as this ternary complex and associates with the small 40S ribosomal subunit, assisted by eIF3 and other initiation factors. **Step 3:** The 43S–mRNA complex forms when the cap-binding complex eIF4F recognizes the 5' cap of the mRNA and recruits the 43S preinitiation complex, allowing the ribosome to engage with the mRNA. ATP-dependent scanning and AUG recognition occur as the ribosome complex moves along the mRNA using ATP hydrolysis until it identifies the AUG start codon, which base-pairs with the anticodon of Met-tRNAi, triggering structural changes in the complex. Finally, **Step 4:** Formation of the 80S initiation complex takes place when the large 60S ribosomal subunit joins the 40S complex at the start codon, placing Met-tRNAi in the P site of the ribosome. At this point, the full 80S ribosome is assembled, and translation is ready to proceed with methionine as the first amino acid in the polypeptide chain. *Image adapted from BioRender template created by Valeria Yartseva, adapted from* Fraser & Doudna (2007).

During Initiation, the 48S complex scans the mRNA in the 5' to 3' direction until the correct start codon is located, usually AUG with a favorable Kozak context, i.e., with A or G three nucleotides before the AUG (position −3) and a G immediately after the AUG (position +4). The complex of eIF1 and eIF1A main-

tains the ribosome in an open conformation of the 40S subunit and ensures the correct start codon is recognized. Once the correct AUG is located, GTP-bound to eIF2 is hydrolyzed by eIF5B to promote eIF1 dissociation, allowing the 60S ribosomal subunit to join the 48S complex, and collectively they form the functional 80S initiation complex, with Met-tRNAi ^{Met} correctly positioned in the P site, ready to start the elongation phase or peptide chain formation.

Peptide Chain Formation

Elongation in eukaryotic translation begins once the initiator Met-tRNA is correctly positioned in the P site of the assembled 80S ribosome, a structure necessary to align each successive aminoacyl-charged tRNA to transfer 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. Elongation has three main steps: (1) decoding, (2) peptide bond formation, and (3) translocation.

Aminoacyl-tRNAs (aa-tRNAs) enter the A site of the ribosome with the help of GTP-bound elongation factor eEF1A. GTP is hydrolyzed when the correct codon-anticodon pairing is formed, and this releases eEF1A and accommodates aa-tRNA into the A site. A guanine nucleotide exchange factor, eEF1B, recycles eEF1A by exchanging GDP for GTP. The amino acid in the A site forms a peptide bond with the growing polypeptide chain in the P site, and the peptidyl transferase center of the 60S ribosomal RNA catalyzes this. Finally, translocation occurs when the GTPase eEF2 binds the ribosome and hydrolyzes GTP, shifting peptidyl-tRNA into the P site and deacylated tRNA to the E site, where it exits.

Translation elongation proceeds codon by codon in the 5' to 3' direction along mRNA, synthesizing the polypeptide chain from the amino (N-) terminus to the carboxy (C-) terminus. 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. During elongation, the ribosome reads the mRNA templates from the 5' end to the 3' end, adding amino acids sequentially to the C-terminus of the growing chain. Multiple ribosomes can translate the same mRNA simultaneously, forming polyribosomes for efficient translation.

Translation Termination

This occurs when a stop codon (UGA, UAG, or UAA) enters the A site of the ribosome. The process is facilitated by eukaryotic release factors (eRFs). Instead of a tRNA, the eRF1 recognizes all stop codons, assisted by eRF3, a GTPase that helps release the polypeptide chain. This leads to dissociation of ribosomal subunits, and they are recycled. Like other processes, cells maintain quality control mechanisms through various mechanisms, like nonsense-mediated mRNA decay (NMD), which targets premature stop codons by sensing the distance between stop codons and exon junction complexes, while non-stop decay targets and degrades mRNAs that lack stop codons by recruiting the degradation machinery via the ski7 complex.

Stop codon read-through is widespread in monocots and dicots, allowing stop codons to be read as 20 standard amino acids. This occurs in transcripts with simpler structures and is not evolutionarily conserved. This reveals the remarkable plasticity and regulatory potential of stop codon usage in plant gene expression.

For more information read: Readthrough events in plants reveal plasticity of stop codons

Amino Acid Properties

Amino acids are the building blocks of proteins (Figure 20). 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. 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.

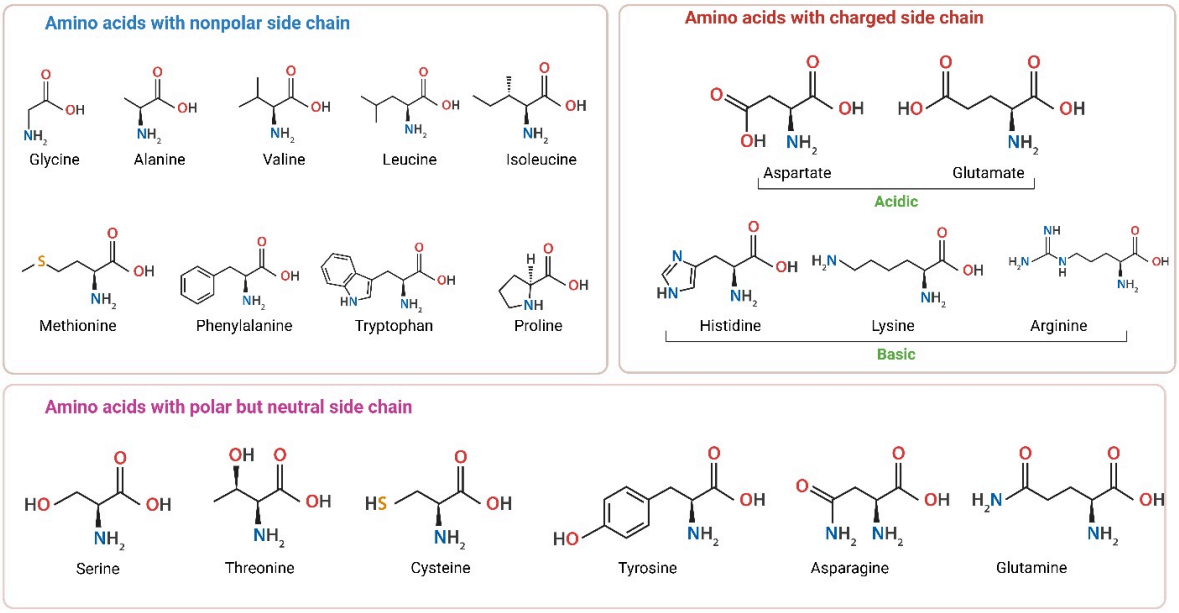


Figure 20 Chart of the 20 standard amino acids

The chemical and physical properties of the side chains provide functionally important properties to amino acids. Some of the important properties of side chains include acidic, basic, or neutral, hydrophobic vs. hydrophilic, and 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.

Protein Structure and Function

Proteins assume complex 3-dimensional structures essential for their functions as enzymes, regulatory factors, or structural proteins. Complex physical-chemical interactions among amino acid side groups determine protein structure.

Several levels of structure are considered. 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 to the protein surface.

Many proteins are multimeric, composed of several polypeptide chains called "subunits", which associate non-covalently or in some cases via disulfide bonds. Such higher-order spatial associations among polypeptides result from interactions among amino acids and result in a quaternary structure. Some multimeric proteins consist of multiple copies of the same polypeptide. Alcohol dehydrogenase, important for flood tolerance, is a homodimer containing two subunits of the same polypeptide encoded at the same gene locus (Schwartz, 1966). Other proteins contain different peptides. ADP-glucose pyrophosphorylase, the rate-limiting enzyme in maize kernel starch synthesis, is a heterodimer containing two different polypeptides encoded by two distinct genes, *shrunken2* and *brittle2* (Hannah, 2005).

For more information read: Starch synthesis in the maize endosperm

For more information read: The genetic control of alcohol dehydrogenase in maize: gene duplication and repression

Chapter Summary

DNA is composed of nucleotides. Phosphate bonds connect nucleotides 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 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. The genetic information of DNA is transferred through transcription to an intermediate molecule called RNA. The starting and stop-

ping of transcription are controlled by specific DNA sequence elements that 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 into amino acids used to build proteins.

References

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21* (3), 381-395. https://doi.org/10.1038/cr.2011.22

Bentley D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature reviews. Genetics*, *15* (3), 163–175. https://doi.org/10.1038/nrg3662

Calado, R. T., & Young, N. S. (2009). Telomere diseases. *The New England journal of medicine*, *361* (24), 2353–2365. https://doi.org/10.1056/NEJMra0903373

Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Regulation of Transcription in Eukaryotes. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9904/

Fraser, C. S., & Doudna, J. A. (2007). Structural and mechanistic insights into hepatitis C viral translation initiation. *Nature Reviews Microbiology*, *5* (1), 29-38. https://doi.org/10.1038/nrmicro1558

Gibilisco, L., Zhou, Q., Mahajan, S., & Bachtrog, D. (2016). Alternative Splicing within and between Drosophila Species, Sexes, Tissues, and Developmental Stages. *PLoS genetics*, *12* (12), e1006464. https://doi.org/10.1371/journal.pgen.100646

Hannah, L. 2005. Starch synthesis in the maize endosperm. Maydica. 50: 497-506. Available at http://www.maydica.org/articles/50_497.pdf

Hernandez, Victoria, " The Meselson-Stahl Experiment (1957–1958), by Matthew Meselson and Franklin Stahl ". Embryo Project Encyclopedia (2017-04-18). ISSN: 1940-5030 https://hdl.handle.net/10776/11481

Kuehner, J. N., Pearson, E. L., & Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology*, *12* (5), 283–294. https://doi.org/10.1038/nrm3098

Lubischer, J. L. (2007). The Cell Cycle, Principles of Control. David O. Morgan. *Integrative and Comparative Biology*, *47* (5), 794-795. https://doi.org/10.1093/icb/icm066

Pierce, B.A. 2002. Genetics: A Conceptual Approach. W.H. Freeman and Company, 405-407, 414-419. New York, NY.

Planells, B., Gómez-Redondo, I., Pericuesta, E., Lonergan, P., & Gutiérrez-Adán, A. (2019). Differential isoform expression and alternative splicing in sex determination in mice. *BMC genomics*, *20* (1), 202. https://doi.org/10.1186/s12864-019-5572-x

Ravi, M. and S.W.L. Chan. 2009. Haploid plant produced by centromere-mediated genome elimination. Nature. 464:615-619. Available at: http://www.nature.com.proxy.lib.iastate.edu/nature/journal/v464/n7288/pdf/nature08842.pdf

Schwartz, D. 1966. The genetic control of alcohol dehydrogenase in maize: gene duplication and repression. Proc. Natl. Acad. Sci. USA. 56, 1431-1436. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC219990/

Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*, *3* (7), a003707. https://doi.org/10.1101/cshperspect.a003707

Yu, W., F. Han, Z. Goa, J.M. Vega, and J.A. Birchler. Construction and behavior of engineered minichromosomes in maize. Proc. Natl. Acad. Sci. USA 21:8924-8929. Available at: http://www.pnas.org/content/104/21/8924.full

Zhang, Y., Li, H., Shen, Y., Wang, S., Tian, L., Yin, H., Shi, J., Xing, A., Zhang, J., Ali, U., Sami, A., Chen, X., Gao, C., Zhao, Y., Lyu, Y., Wang, X., Chen, Y., Tian, Z., Wu, S. B., & Wu, L. (2024). Readthrough events in plants reveal plasticity of stop codons. *Cell reports*, *43* (2), 113723. https://doi.org/10.1016/j.celrep.2024.113723

CHAPTER 2.

GENE EXPRESSION AND REGULATION

Walter P. Suza; Faizo Kasule; Philip W. Becraft; and Madan K. Bhattacharyya

Introduction

Every cell in a plant contains the same set of genes and genetic information. Yet different sets of genes are required for the various functions of different cells or tissues, and 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 gene activity is known as **gene expression**.

The term **expression** in genetics can be used in different ways that are sometimes confusing. Typically, the gene is considered "expressed if a gene product is produced." However, it sometimes occurs that a transcript might be produced but not a protein, or that a protein is produced but in an inactive state. In such cases, although a gene product is made, the biological activity encoded by that gene is absent. For this section, the key point is how the biological activity encoded by a gene is regulated (Figure 1).

The expression of genes in specific plant cells, tissues, and organs, and the timing of this expression, require precise regulation. Expression, or genetic function, can be regulated at any step from transcription, RNA processing, translation, through posttranslational protein modification, as discussed in Chapter 1 under the **"Transcription and Gene Structure"** section. Regulation can be qualitative, where gene expression is either"on" or "off," or quantitative, where gene expression levels can be modulated "up" or "down." Fluctuations in the intensities of external stimuli coupled with changes at the genomic level result in different developmental outcomes or physiological states. Regulation of gene expression at the transcriptional level can occur through chromatin and histone modifications. Also, a gene sequence can be differentially spliced to produce variable-length mRNA products, leading to new protein products with novel functions. Some genes do not encode proteins but short forms of RNAs with regulatory functions, such as the induction of flowering. Finally, protein products can be subjected to modifications such as phosphorylation or dephosphorylation to alter their functions or entirely degraded to turn off a gene.

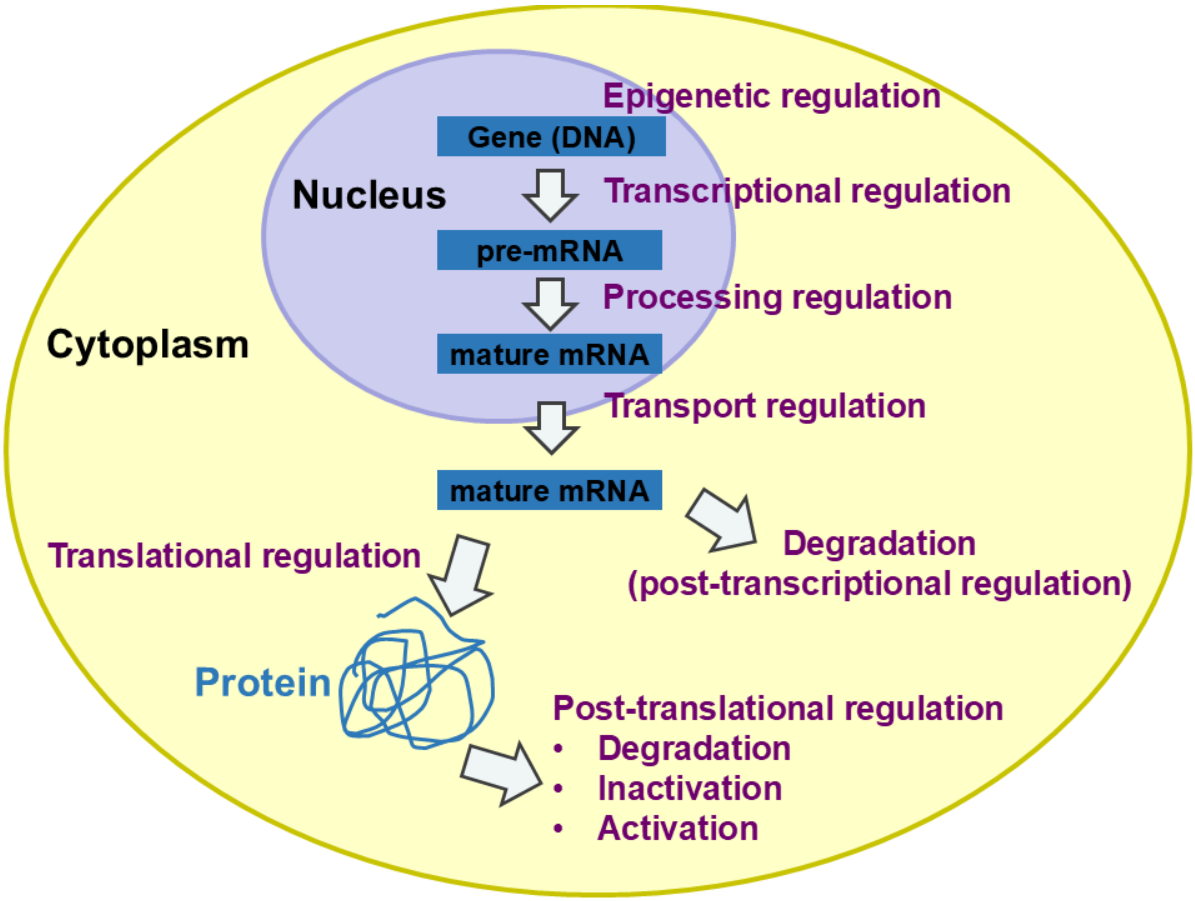


Figure 1 Multilevel regulation of gene expression.

Transcriptional Gene Regulation

Transcription is the first step in gene expression. Therefore, it makes sense in terms of cellular economy to regulate expression at this level as one of the most critical regulatory points. We already described the involvement of RNA polymerase in the transcription process (see the "Transcription and Gene Structure" section in Chapter 1). However, other protein factors are required. Proteins involved in transcriptional regulation are known as **transcription factors**. The interaction of transcription factors with specific DNA sequences regulates the process of gene transcription.

The Concept of Differential (Regulated) Gene Expression

As described in the introduction, not every gene is always expressed. When a gene displays different levels of expression in different circumstances, it is known as **differential expression**. Circumstances that might apply to differential expression 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 with similar expression patterns would be regarded as **co-expressed**.

Promoters

As mentioned, and discussed in Chapter 1, 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**. Promoters contain a core required to bind 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 one hundred to 1000 nucleotides (100–1,000 base pairs).

The complete promoter sequences of different genes expressed similarly may differ. 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 can be interpreted by the cell and control gene expression. Sometimes, *cis* elements promote gene transcription, and other times restrict gene transcription in particular cells and tissues.

Promoter analysis is facilitated by using **reporter genes**. The reporter gene produces easily identifiable and quantifiable effects, which can be used to determine the function of a regulatory region of another gene (promoter, promoter elements, enhancers) in cells, tissues, or organs. Such analyses are critical to crop biotechnology, where targeted gene expression 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 (He et al., 2020). An example of a reporter gene is *RUBY* (Figure 2), which is a fusion of three genes: (1) *CYP76AD1* encodes a cytochrome P450 enzyme that catalyzes the conversion of tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA) and oxygenation of L-DOPA to cyclo-DOPA; (2) the next enzyme in the reporter gene fusion, *DODA (* L-DOPA 4,5-dioxygenase), makes an enzyme that cleaves L-DOPA to form **betalamic acid**. The enzyme converts betalamic acid and **cyclo-DOPA** spontaneously to form **betanidin**, (3) **glucosyl transferase (GT),** which glycosylates betanidin to produce **betalain,** a natural red pigment. Because betalain has a very bright red color, which is vividly visible to the naked eye, the *RUBY* system allows rapid and non-destructive monitoring of plant transformation and gene expression without the need for specialized equipment, expensive chemicals, or invasive treatments (He et al., 2020). The three gene fusions are separated by *P2A* peptides (Figure 2).

The ectopic expression of *RUBY* leads to visible pigment accumulation in plant tissues in many plants, like *Arabidopsis*, tomato, eggplant, carrot, and maize, among others. Strong promoters like Cauliflower Mosaic Virus 35S (*CaMV35S)* and Ubiquitin have been used to drive the constitutive *RUBY* expression throughout Arabidopsis and maize. In contrast, seed-specific promoters such as the *At2S3* from Arabidopsis restrict *RUBY* expression to seeds .

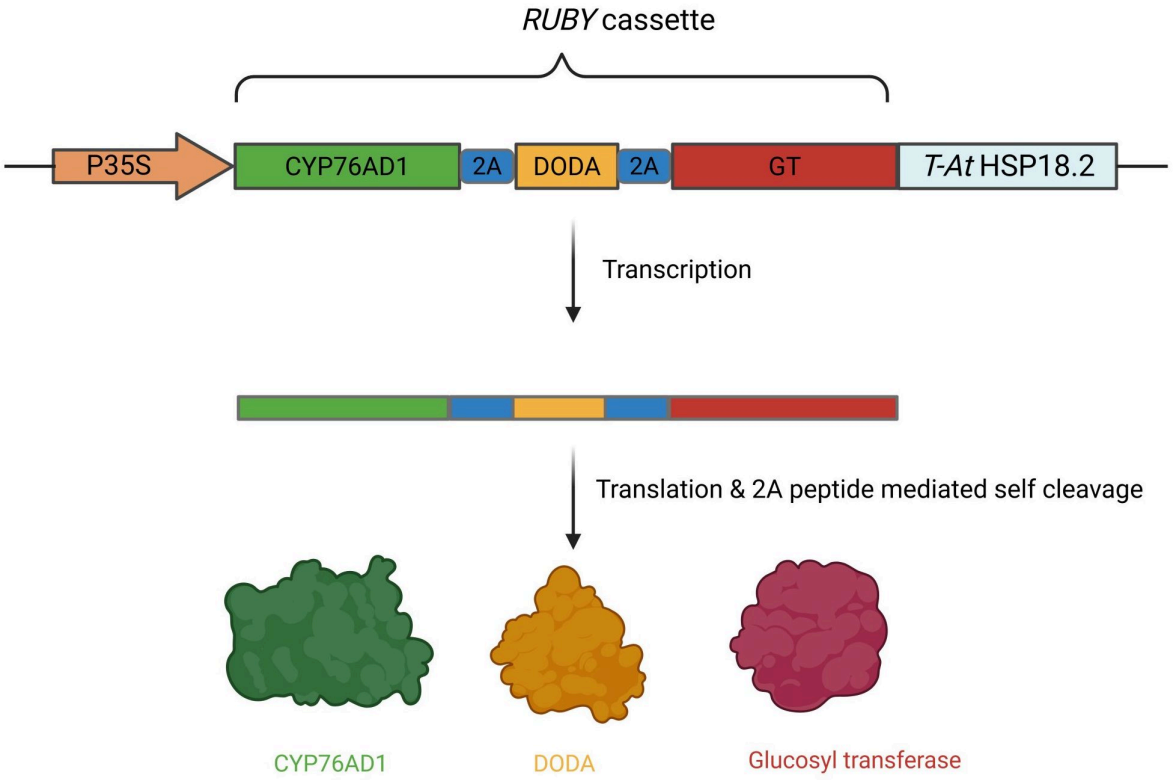


Figure 2 Schematic representation of betalain biosynthesis engineering in plants. All three betalain biosynthetic genes (CYP76AD1, DODA, and GT) were fused into a single ORF and regulated through CaMV35S enhanced promoter and HSP18.2 terminator. In between the genes, sequences that encode 2A peptides were placed. The 2A peptides undergo self-cleavage, thus releasing the individual enzymes for betalain biosynthesis. The ORF of 2 A-linked betalain biosynthesis genes is named RUBY. *Image created by Faizo Kasule.*

Figure 3 shows an example of *RUBY* expression in various maize tissues using the P35S promoter and *Arabidopsis* HSP18.2 terminator. Using this constitutive 35S promoter to drive *RUBY* expression, betalain accumulation was observed from early stages and across the different tissues and developmental stages in maize (Figure 3). This indicates that *RUBY* serves as an excellent marker for maize transformation (Lee et al., 2023).

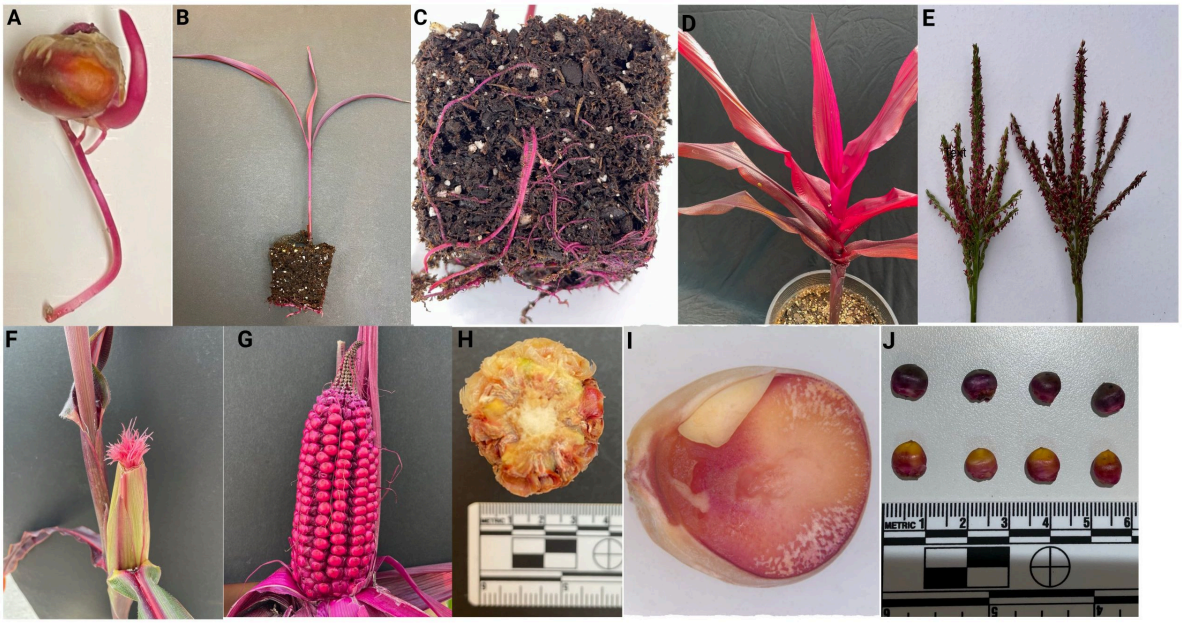


Figure 3 P35S Promoters direct the expression of RUBY in different transgenic maize plant tissues. A) germinating T0 seed, B) T0 seedling, C) Transgenic root architecture, D) Transgenic plant with leaves showing betalain accumulation, E) Tassel and pollen of T0 plant, F) Encased ear with silks, G)Transgenic ear, H) Cross section of harvested shelled ear, I) Cross section of kernel and J) Mature kernels with dark betalain pigment: *Image created by Faizo Kasule. Pictures were provided by Sehiza Grosic and Dr. Minjeong Kang, Iowa State University.*

Read: The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants

Enhancers

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

Transcription Factors

RNA polymerase binds the promoter at the TATA box and drives gene transcription in cooperation with other proteins called transcription factors. Transcription factors interact with RNA polymerase to facilitate its binding to the promoter and regulate its activity. Some transcription factors, known as "basal transcription factors," are fundamental to the 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 the 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, depending on what other *cis* elements and/or transcription factors might be present.

A. About 6% of the 30,000 genes in the Arabidopsis genome encode transcription factors. Most of these genes occur in families. For more information read: Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes

B. Transcription factor biotechnology is gaining interest in efforts to solve crop production challenges in unfavorable environments. For more information read: Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres

C. Read: Regulating the Regulators: The Future Prospects for Transcription-Factor-Based Agricultural Biotechnology Products

Epigenetic Regulation

Chromatin Structure and Histone Modification

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

Transcriptional regulation often involves modification of chromatin structure mediated by posttranslational 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 (FLD, p462) stabilizes nucleosomes and represses transcription. On the other hand, histone acetylation by histone acetyltransferase 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 regulate chromatin condensation or decondensation.

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 to 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 4) as demonstrated by the low level of methylation in regions of the genome undergoing active transcription.

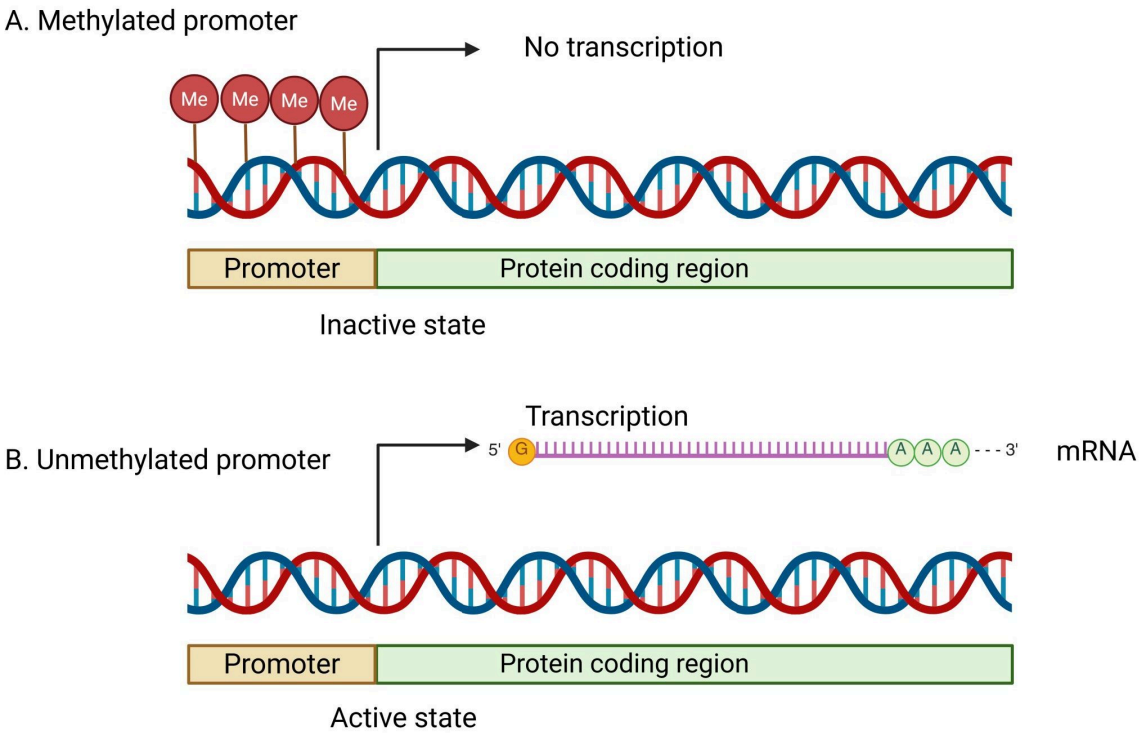


Figure 4 Absence of methylation correlates with gene expression. A) The promoter region may be methylated during development or environmental conditions such that the gene is inactivated. B). The same gene is activated upon the removal of the methyl groups. *Image created by Faizo Kasule.*

Researchers use the DNase sensitivity assay to determine the genomic regions undergoing active transcription (Chapter 1). DNAse-sensitive regions contain lower methylation than regions that are resistant to the enzyme. The specific methylated DNA sites can be determined using restriction enzymes that recognize the same sequence but differ in their ability to digest DNA if their target site is methylated. For example, the enzyme HpaII cleaves CCGG only if not methylated. However, MspI cleaves CCGG if it is unmethylated or methylated. Figure 5 shows an example of a DNA sequence containing stretches of CCGG that are variably methylated. Using the example of SNP analysis in *Plant Breeding Methods*, an experiment can be designed where DNA is isolated and subjected to the two restriction enzymes. The products of the two enzymes can be visualized on a gel as shown in Figure 5.

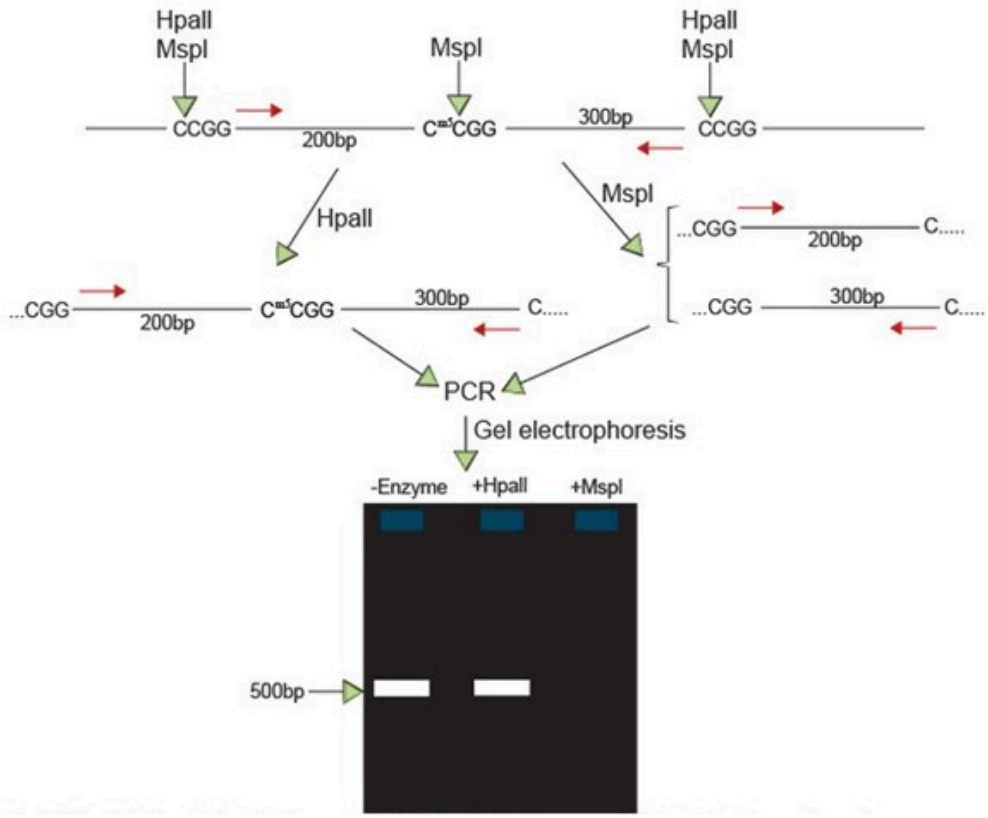


Figure 5 The DNA strand (depicted as a single line) containing two unmethylated cytosines and one 5-methylcytosine is subject to separate restriction analyses with HpaII and MspI enzymes. The HpaII enzyme will cleave at the two CCGG sequences that are unmethylated. The MspI enzyme can cleave at the two CCGG sequences, and also the methylated sequence (C ^{m5} CGG). The treatment of this DNA with HpaII will result in a single product. However, MspI will cleave at three sites, resulting in two DNA fragments. There are two primer attachment sites on the undigested strand (red arrows) for PCR analysis of the products. The two primer sites are retained in an intact product of HpaII, which allows for a PCR amplification of a product that is 500bp in length. The products of MspI also contain primer sites, but they are not intact and cannot be amplified by PCR with the original primers. After gel electrophoresis, the PCR product from undigested DNA is of similar length to that of DNA digested with HpaII. The lane representing PCR products of DNA digested with MspI is blank, indicating that the reaction did not work.

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 specific 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 transcription levels, higher histone acetylation levels, and lower DNA methylation levels.

Read: The Epigenome and Plant Development

RNA-Level Regulation

Regulation of gene expression occurs at many levels, including post-transcriptionally (Figure 6). From 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. Figure 6 summarizes how post-transcriptional regulation is essential for plant cells to respond to stress.

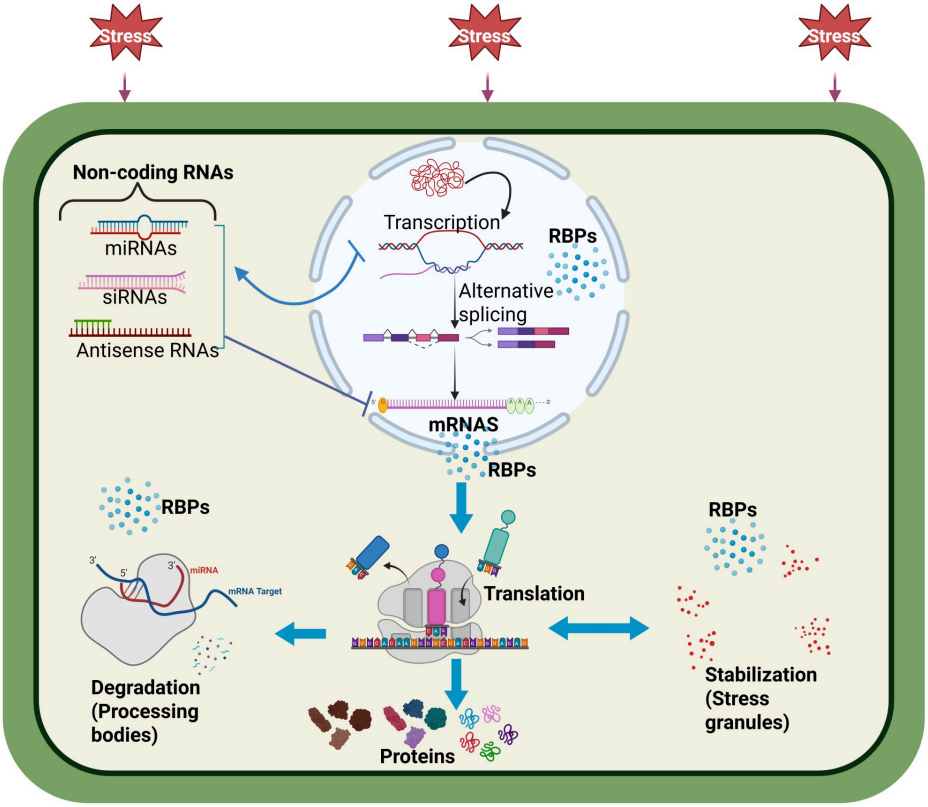


Figure 6 Post-transcriptional RNA regulation in a plant cell responding to stress. RBPs are RNA-binding proteins that function to regulate RNA. *Image created by Faizo Kasule in BioRender adapted from Nakaminami et al. (2012).*

Alternative Splicing

Alternative splicing is the mechanism of alternatively processing pre-mRNAs to generate mRNAs with different exon combinations. In animals, many genes are spliced in alternative ways, leading to multiple proteins from a single gene. According to some estimates, over half of all human genes produce alter-

nately spliced transcripts. Interestingly, as described in the textbook, the sexual differences between male and female flies are determined by alternate RNA splicing and the different resulting proteins. One spliced form of the Tra transcript produces male flies; an alternatively spliced Tra transcript results in female flies.

Over 20% of multi-exon gene transcripts undergo alternative splicing in rice and Arabidopsis.

Examples of alternative splicing in plants include the maize pyruvate orthophosphate dikinase (PPDK) gene involved in C4 photosynthesis (Sheen, 1991). Alternative splicing of the PPDK gene results in a protein with either a short stretch of amino acids that keeps it in the cytoplasm or another that targets it to the chloroplast.

Another example is the Jasmonate Zim-Domain (*JAZ*) genes involved in the signaling pathway of jasmonic acid. Jasmonic acid is required for insect and pathogen defense and reproductive development. In Arabidopsis, the *JAZ10* gene produces alternately spliced transcripts with three splice variants (Chung and Howe, 2009). When one of the splice-variants, *JAZ10.4*, is overexpressed, it causes male sterility by producing a protein that interferes with JA signaling in the flower, preventing the development of stamens.

Read: Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes

Read: 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*

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 significant in plants since plants cannot move to avoid harsh environmental conditions, and stresses are known to induce specific changes in RNA stability (Nakaminami et al., 2011). 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 (Figure 7).

Therefore, controlling ribonuclease access to the mRNA substrate is an essential 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 specific cellular locations to restrict its access to RNA, and controlling the stability of the mRNA by decreasing accessibility to nucleases.

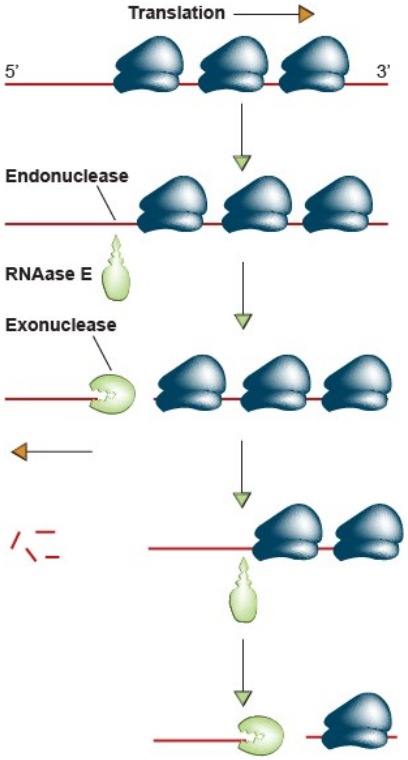


Figure 7 mRNA degradation by ribonucleases. Four regions of an mRNA are essential for its overall stability. These include the 5' untranslated region and the cap, coding region, 3' untranslated region, and the poly(A) tail. The cap at the 5' end protects the mRNA from exonucleases that degrade RNA from 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 insufficient. It only provides stability when bound by 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.

Read: RNA regulation in plant abiotic stress responses

Small RNAs and RNA Interference

RNA interference (RNAi) refers to RNA-mediated regulation of gene expression. It is a natural mechanism for moderating gene expression in the cells of all eukaryotic organisms, from plants to humans. RNAi usually results in decreased expression or silencing of a target gene through several different mechanisms, including transcriptional silencing, translational silencing, or mRNA degradation. In plants, mRNA degradation is perhaps the most common mechanism of RNAi. These processes are mediated by small RNAs that target mRNA or chromosomal genes through sequence homology. Several small regulatory RNAs exist, including microRNAs (miRNAs) and small interfering RNAs (siRNAs). The formation of siRNAs or miRNAs is initiated by the formation of double-stranded RNA (dsRNA), which is then processed by Dicer endonucleases to form the small RNAs of 21-24 nucleotides. This RNA then associates with a protein complex called RISC and targets cellular mRNAs with homologous sequences.

View: RNA interference animation

RNAi is an important mechanism by which endogenous genes are regulated. It commonly functions to refine the pattern of gene expression. For example, a maize gene called *tasselseed4* (*ts4*) codes for a microRNA called *miR172*. This microRNA is required to target the degradation of mRNAs for several related genes. Those target mRNAs are transcribed in male flowers but usually do not accumulate there. However, when the *ts4* gene is mutant, *miR172* is not produced. The target mRNAs and their encoded proteins accumulate in male flowers, causing them to develop female characteristics, and hence, seeds form on tassels (Chuck et al., 2007).

RNAi can be accomplished more efficiently by expressing a portion of the target gene that has been engineered as an inverted repeat in transgenic crop plants. Following transcription of this engineered gene, the RNA molecules form a hairpin structure that is then cleaved into small fragments of double-stranded RNA, which interfere with the accumulation and function of the endogenous mRNA molecules of the target gene.

Discovery of RNAi

Although researchers working with worms received full credit for the discovery of RNAi (Fire et al., 1998), related phenomena had been previously observed in plants (Napoli et al., 1990). In the early 1990s, Richard Jorgensen and co-workers, working with petunia, explored ways to increase flower pigmentation through transgenic approaches. Surprisingly, Jorgensen and his team observed that overexpression of transgenes that were intended to increase pigmentation in petunia did not deepen pigmentation. Instead, they observed flowers that showed variegated pigmentation, with some lacking pigments alto-

gether (Figure 8). The researchers concluded that the transgenes and endogenous sequences homologous to the transgenes must have been inactivated, a process they coined **co-suppression** (Napoli et al., 1998). Other laboratories later found that plants respond to RNA viruses by destroying viral RNAs to prevent viral genes from being expressed inside plant cells, and cause disease. Co-suppression and viral RNA targeting processes in plant cells are referred to as **post-transcriptional gene silencing** (PTGS).

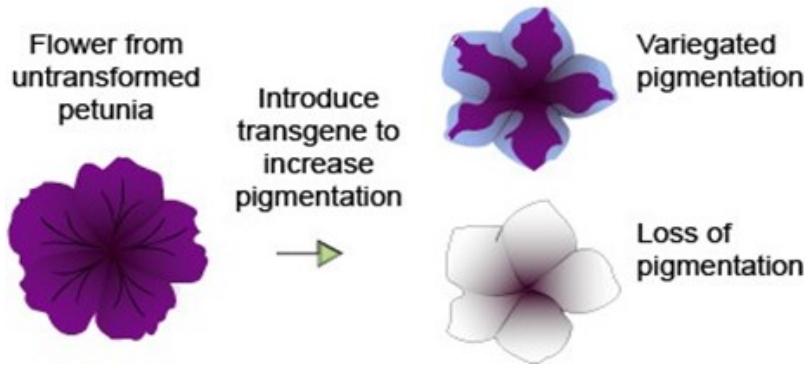


Figure 8 An example of PTGS in plants. The introduction of a transgene for flower pigmentation in petunia results in variegation or total loss of pigmentation due to co-suppression of endogenous genes involved in anthocyanin biosynthesis. *Source: Adapted from Napoli et al. (1998).*

Read: Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*

Read: Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans

Post-transcriptional Gene Silencing

Many RNA and DNA viruses stimulate a process known as posttranscriptional gene silencing (PTGS) immediately after infecting a plant. PTGS is proposed as an anti-viral defense mechanism in plants. The most crucial trigger in PTGS is the production of double-stranded RNA by the virus during viral multiplication. Viral RNAs are recognized as foreign and are targeted for degradation in infected cells through RNAi.

It is now well established that PTGS in plants is an RNA-degradation mechanism that is like RNA interference (RNAi) in animals (Figure 9). In plants, mRNA degradation is perhaps the most common mechanism of RNAi. As mentioned above, RNAi usually results in decreased expression or silencing of a

target gene The formation of siRNAs) or miRNAs) is initiated by the formation of dsRNA molecules which are processed by dicer endonucleases to form the small RNAs of 21-24 nucleotides. Small RNAs then associate with a protein complex called RISC (RNAi silencing complex) and target cellular mRNAs with homologous sequences to result in reduced protein expression (Figure 8). miRNAs act at the post-transcriptional level leading to translational repression, degradation of the endogenous transcripts and silencing of genes through epigenetic means.

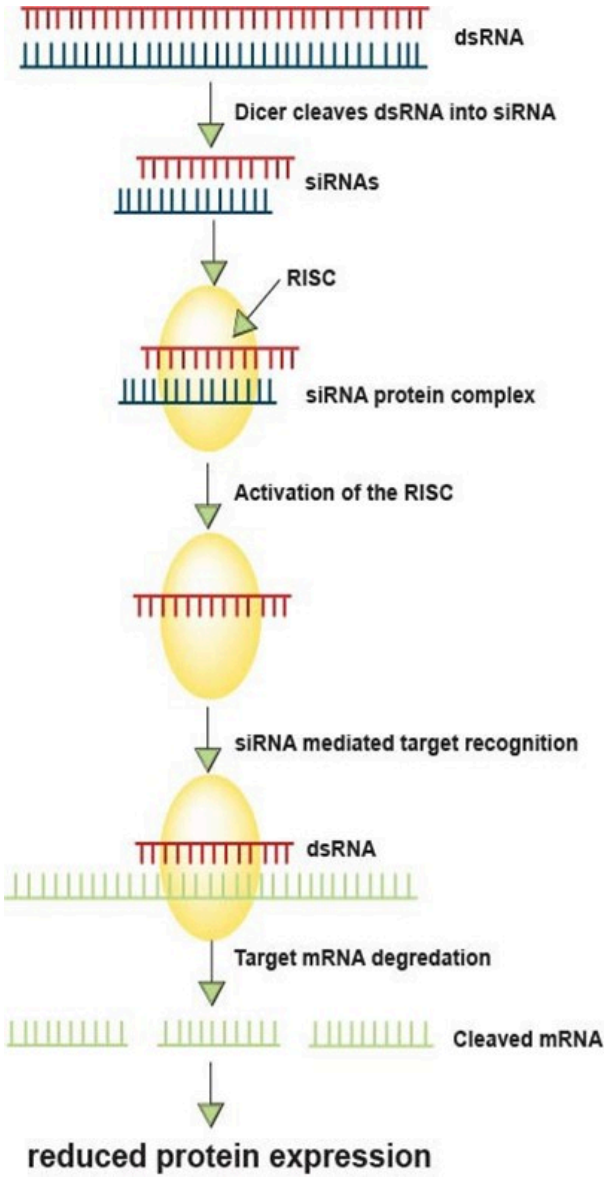


Figure 9 Mechanism of PTGS. *Adapted from Ali et al. (2010).*

Read: The heterochronic maize mutant *Corngrass1* results from overexpression of a tandem microRNA

Read: RNAi-mediated crop protection against insects

Translational Regulation

As described earlier, for most genes, the protein product of the gene performs the biological function. Therefore, regulating the amount of protein production effectively governs gene expression. There are several known mechanisms by which the process of translation is regulated. A detailed consideration of these mechanisms is beyond the scope of this course, but generally, 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 respond to stress quickly 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 mRNA degradation. Ribonucleases (see RNA Stability and Degradation section in Chapter 2) 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 the synthesis of proteins with incorrect functions, which will negatively impact cellular processes. During transcription, errors could add or omit nucleotides, altering the proper codon sequence and creating mutant proteins (Chapter 3).

Protein-Level Regulation

After translation, proteins are subject to various modifications that can regulate their activity. There are many 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; 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 processing certain proteins, for example, in removing 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 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 the conditions of the cellular environment. The redox state can result in oxidation or reduction of proteins, particularly sulfhydryl side groups. Cellular pH can affect the charge of ionizable side groups.

Common Types of Protein Regulation

All 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 in which this occurs. Again, we will briefly consider a few of the more common mechanisms.

One major way protein modification can regulate protein function is by altering its activity. This can be true for many types of proteins, including enzymes, transcription factors, signaling proteins, and structural proteins—for example, a group of transcriptional regulators known as WRKY transcription factors function in several stress responses. WRKY8 is important for pathogen defense. Upon infection, WRKY8 becomes phosphorylated, which activates its DNA-binding activity and results in the activation of genes involved in pathogen defense (Ishihama et al., 2010).

> Read: Phosphorylation of the *Nicotiana benthamiana* WRKY8 Transcription Factor by MAPK Functions in the Defense Response

Cells are compartmentalized (Figure 10) 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.

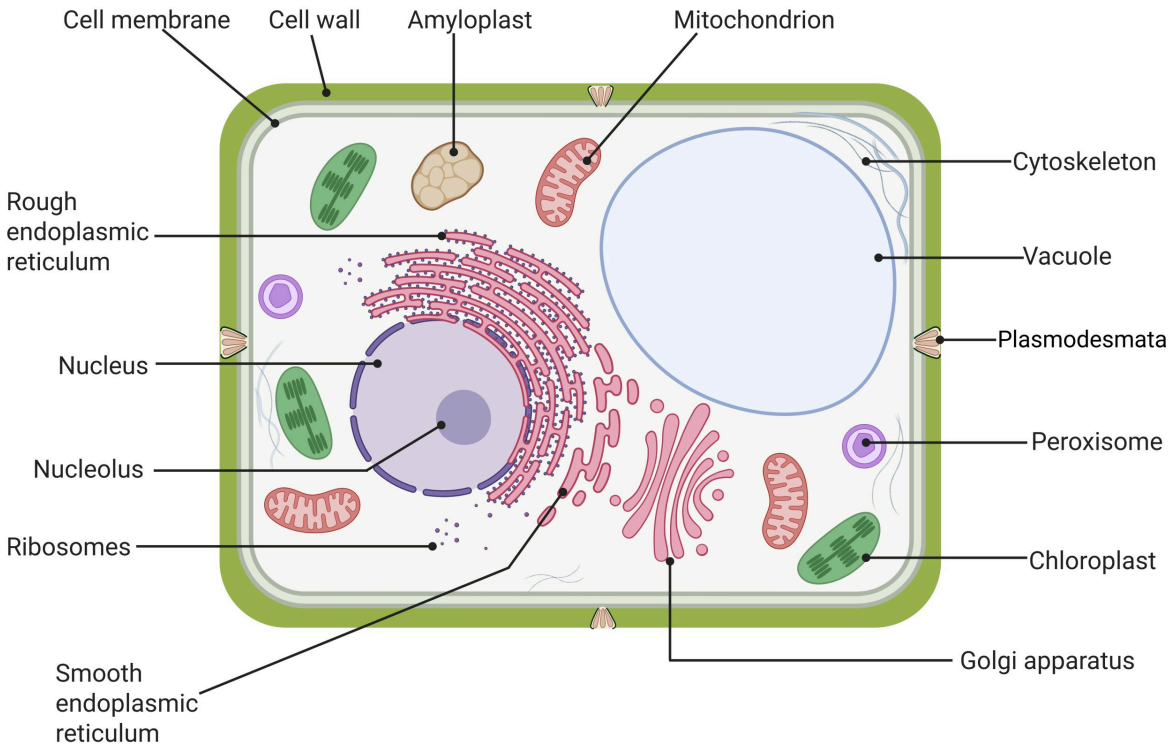


Figure 10 Plant cells are compartmentalized, allowing the separation of metabolic processes. *Image created in BioRender by Faizo Kasule.*

Compartmentalization is regulated for some proteins. For example, upon exposure to light, the phytochrome protein moves into the nucleus, affecting light-regulated gene expression.

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 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-systemin 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 (Figure 8) are rich in proteases that function similarly in protein degradation as 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 means 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 with 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 the growth hormone gibberellic acid.

Proteins have lifetimes that range from a few minutes to weeks or more. Cells continuously make proteins from mRNA molecules and break them down into 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

seedling growth are derived from the degradation of seed storage proteins. Conversely, in annual crop plants, many amino acids in seed storage proteins are derived from proteins degraded in leaves and other plant parts during senescence.

Chapter 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 within the coding region. An enhancer can increase transcription levels when added to genes they are not normally associated with. RNA polymerase binds the promoter at the TATA box and drives gene transcription in cooperation with other proteins. 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 chromatin modification 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 with different exon combinations. 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 mRNA degradation. 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.

References

Ali N, Datta SK, Datta K. RNA interference in designing transgenic crops. GM Crops. 2010 Jul-Sep;1(4):207-13. doi: 10.4161/gmcr.1.4.13344. PMID: 21844675.

Chuck, G., Cigan, A.M., Saeteurn, K., and Hake, S. 2007. The heterochronic maize mutant *Corngrass 1* results from overexpression of a tandem microRNA. Nat. Genet. 39:544-549. Available at http://www.nature.com/ng/journal/v39/n4/pdf/ng2001.pdf

Chung HS, Howe GA. 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. 2009 Jan;21(1):131-45. doi: 10.1105/tpc. 108.064097. Epub 2009 Jan 16. PMID: 19151223; PMCID: PMC2648087.

He, Y., Zhang, T., Sun, H., Zhan, H., & Zhao, Y. (2020). A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research*, *7* (1), 1-6. https://doi.org/10.1038/s41438-020-00390-1

Ishihama, N., Yamada, R., Yoshioka, M., Katou, S., and Yoshioka, H. 2011. Phosphorylation of the *Nicotiana benthamiana* WRKY8 Transcription Factor by MAPK Functions in the Defense Response. The Plant Cell. 23: 1153-1170. Available at http://www.plantcell.org/content/23/3/1153.full.pdf

Lee, K., Kang, M., Ji, Q., Grosic, S., & Wang, K. (2023). New T-DNA binary vectors with NptII selection and RUBY reporter for efficient maize transformation and targeted mutagenesis. *Plant Physiology*, *192* (4), 2598-2603. https://doi.org/10.1093/plphys/kiad231

Nakaminami, K., Matsui, A., Shinozaki, K., and Seki, M. 2012. RNA regulation in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA). Gene Regulatory Mechanisms. Available at http://www.sciencedirect.com/science/article/pii/S1874939911001386

Nelson, D.E., Repetti, P.P., Adams, T.R., Creelman, R.A., Wu, J., Warner, D.C., Anstrom, D.C., Bensen, R.J., Castiglioni, P.P., Donnarummo, M.G., Hinchey, B.S., Kumimoto, R.W., Maszle, D.R., Canales, R.D., Krolikowski, K.A., Dotson, S.B., Gutterson, N., Ratcliffe, O.J. and Heard, J.E. 2007. Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc. Natl. Acad. Sci. USA: 16450-16455. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2034233/pdf/zpq16450.pdf

Sheen, J., 1991. Molecular mechanisms underlying the differential expression of maize pyruvate orthophosphate dikinase genes. The Plant Cell 3:225-245. Available at http://www.plantcell.org/content/3/3/225.full.pdf

CHAPTER 3.

MUTATIONS AND TRAIT VARIATION

Walter P. Suza; Philip W. Becraft; Faizo Kasule; and Madan K. Bhattacharyya

Introduction

Crop breeding and genetic research rely on genetic variation, which is synonymous with DNA variation. Mutations are the source of genetic variation that drives trait improvement. Although most mutations are rare and deleterious and are selected against during evolution, some create beneficial genetic variation. Therefore, it is often of interest for plant breeding or genetic studies to induce genetic mutations by exposing seed or plant tissue to certain agents, called mutagens. The chapter on mutations and variation in *Crop Genetics* extensively covered the genetic and molecular basis of mutations and variation. The practical utilization of genetic variability from induced mutations was covered in *Plant Breeding Methods*. Therefore, for more information, it may be helpful for you to review the texts.

Mutation and Epimutation

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

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 another or a purine by another purine. The less common type is the **transversion**, in which a purine may be substituted by a pyrimidine or vice versa. As described and discussed in the next section, these can have various effects on gene function depending on where they occur in a gene.

Insertions and Deletions

Insertions and deletions, collectively known as indels, are other types of DNA mutations that occur frequently. They 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 in multiples of 3 are referred to as in-frame insertions and deletions, resulting 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.

A common source of inserted DNA is transposons, which are DNA sequences that can move from one genomic site and insert into another. Transposons were discovered in maize by Barbara McClintock. She was awarded the Nobel Prize in Physiology and Medicine in 1983 for this discovery (Figure 1).

Figure 1 Barbara McClintock, recipient of the 1983 Nobel Prize in Physiology and Medicine. *Image Source: Cold Spring Harbor Laboratory.*

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 induced by many plant transposons is instability. 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.

Epimutations

As previously described, epigenetic mechanisms, including DNA methylation, histone modification, and RNA interference (RNAi), are essential to regulate gene expression for plant development and defend genomes against widespread mutagenic activity of transposons. Epimutations are heritable changes in gene expression that occur without any alteration in gene sequence (Daxinger and Whitelaw, 2010). Some epimutations may affect the function or regulation of a single gene. In contrast, others may inter-

fere with the epigenetic mechanisms that control transposons, leading to increased transposon activity that can negatively impact growth and development. For example, the *SUPERMAN* gene of Arabidopsis has known epialleles that show heritable changes in flower morphology, producing extra anthers, even though the DNA base sequence is not altered. These alleles show heritable changes in the DNA methylation pattern (Jacobsen and Meyerowitz, 1997). The Arabidopsis *decrease in DNA methylation1* (*ddm1*) mutant (Vongs et al., 1993) provides a good example of the impact of genome-wide epimutations on plant growth and development. The wild-type *DDM* gene controls DNA methylation, and when compared to the wild type, the *ddm1* mutant shows progressive loss of DNA methylation, enhanced transposon activity, and developmental defects after several generations of inbreeding (Kakutani et al., 1996; Hirochika et al., 2000).

- Read: Transgenerational epigenetic inheritance
- Read: Hypermethylated SUPERMAN epigenetic alleles in Arabidopsis
- Read: Arabidopsis thaliana methylation mutants
- Read: Developmental abnormalities and epimutations associated with DNA hypomethylation mutations
- Read: Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation.

Effects of Mutations on Gene Function

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 (Figure 2).

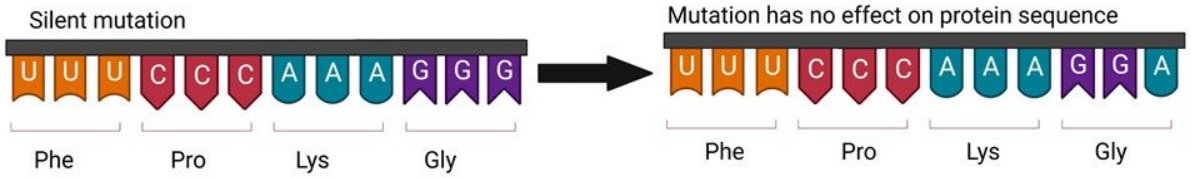


Figure 2 Effects of mutations on gene function: example of a silent mutation. *Image created by Faizo Kasule*.

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 (Figure 3).

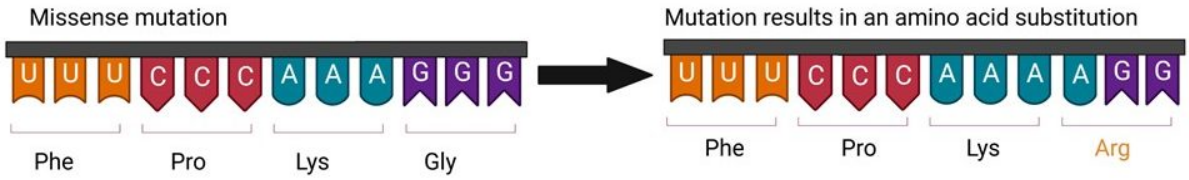


Figure 3 Effects of mutations on gene function: example of a missense mutation. *Image created by Faizo Kasule*.

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, such as in Figure 4.

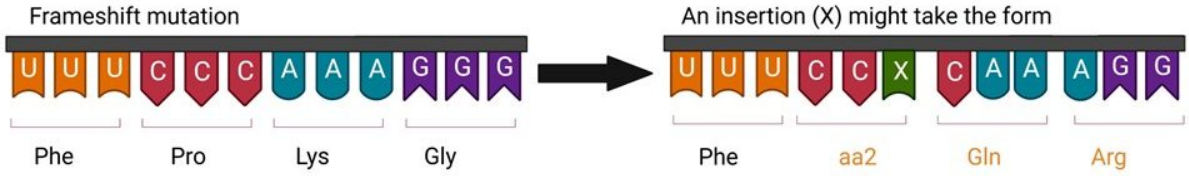


Figure 4 Effects of mutations on gene function: example of a frameshift mutation. *Image created by Faizo Kasule*.

Because the new sequence of codons is different from the original, the entire amino acid sequence is changed from the point of insertion (nucleotide sequence will be CCX CAA AGG and change in amino acid sequence accordingly if X is inserted; note that the change from A to X is a point mutation in the figure not an insertion since third C residue of the codon is not carried over to the next codon). 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 results in the loss of the function of the protein.

Amino Acid Additions or Deletions

Alternatively, if insertions or deletions are in multiples of 3, amino acid additions or deletions can arise. Consider the following 3-base insertion (Figure 5). If such insertions or deletions occur in important regions of the protein, they might impair function.

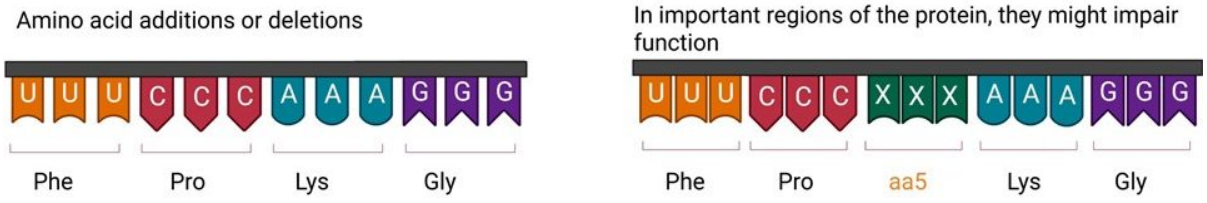


Figure 5 Effects of mutations on gene function: example of an insertion/deletion (indel) mutation. *Image created by Faizo Kasule*.

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 6).

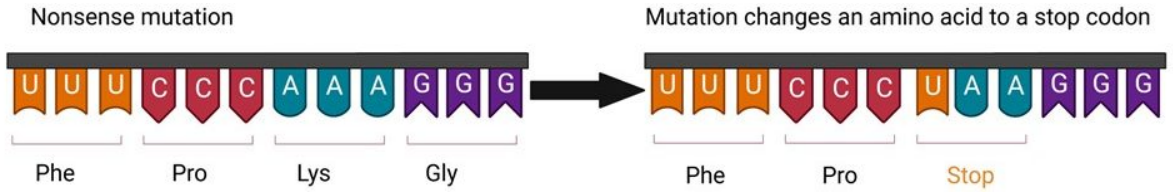


Figure 6 Effects of mutations on gene function: example of a nonsense mutation. *Image created by Faizo Kasule*.

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 be expressed at inappropriate times, places, or levels, which can affect a plant's phenotype.

Teosinte, the progenitor of modern maize, produces multiple branches (Figure 7A) due to reduced apical dominance. In teosinte and modern maize, apical dominance is controlled by 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 7B).

Figure 7 Regulatory region mutation of the tb1 gene converted teosinte (A) to modern maize (B). (C) Ears of teosinte and maize, and (D) Maize and teosinte kernels. *Image created by Faizo Kasule, picture courtesy: Prof. Kan Wang, Mercy Azanu, and Faizo Kasule, Iowa State University.*.

- Read: The evolution of apical dominance in maize

- Read: Identification of a functional transposon insertion in the maize domestication gene *tb1*

Introns

A mutation in an intron may lead to aberrant pre-mRNA splicing. For example, a change from GT to A-T in the splice site sequence of intron 3 of the maize *brittle-2* gene involved in starch biosynthesis abolishes the use of the splice site. Consequently, the *bt2* mutants accumulate multiple non-wild-type-sized transcripts that encode non-functional proteins and thus produce shrunken or brittle kernels with high sugar content due to low starch synthesis (Lal et al., 1999).

Read: A splice site mutant of maize activates cryptic splice sites, elicits intron inclusion and exon exclusion, and permits branch point elucidation

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 called 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 the normal allele can provide the necessary function in a heterozygote.

In corn, a recessive loss-of-function mutation in the *y1* gene, which encodes the phytoene synthase enzyme required for yellow carotenoid pigment synthesis, causes a white kernel phenotype often seen segregating in sweet corn hybrids (Figure 8).

Figure 8 An ear of corn from a self-pollinated plant that was heterozygous for the recessive y1 mutation. Homozygous mutant kernels are segregating and visible as white kernels. *Figure courtesy of Maize Genetics and Genomics Database*.

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 inappropriately high levels, 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 at 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 (Figure 9).

Figure 9 Ethylene induces tomato ripening as seen in the normal tomato on the left. In the green ripe tomato mutant (right), ripening is impaired due to decreased ethylene sensitivity. This is a dominant neomorphic mutation caused by ectopic expression of the GR protein in the fruit. Where it is not normally expressed. The protein appears to interfere with the function of ethylene receptors (Barry and Giovannoni, 2006).

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. The subject of RNAi will be discussed in greater detail in Chapter 6.

Read: Ripening in the tomato Green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling

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 an enzyme required for normal starch biosynthesis (Rahman et al., 1998). In the *su1* mutant, with defective starch biosynthesis, sugar concentration 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).

Examples

Breeding for increased sweetness in sweet corn [PDF]

Characterization of SU1 isoamylase, a determinant of a storage starch structure in maize

Starch synthesis in maize endosperms

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 10). At the molecular level, the cell response to GA involves several processes, including the destruction of repressor proteins, referred to as DELLAs.

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 to destroy the repressor in response to GA. Mutations in the DELLA domain of *Rht-1* genes result in a protein product resistant to degradation, leading to failure to signal a GA response. The inability 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, proving that the DELLA protein's new function is sufficient to induce dwarfism (Peng et al. 1999).

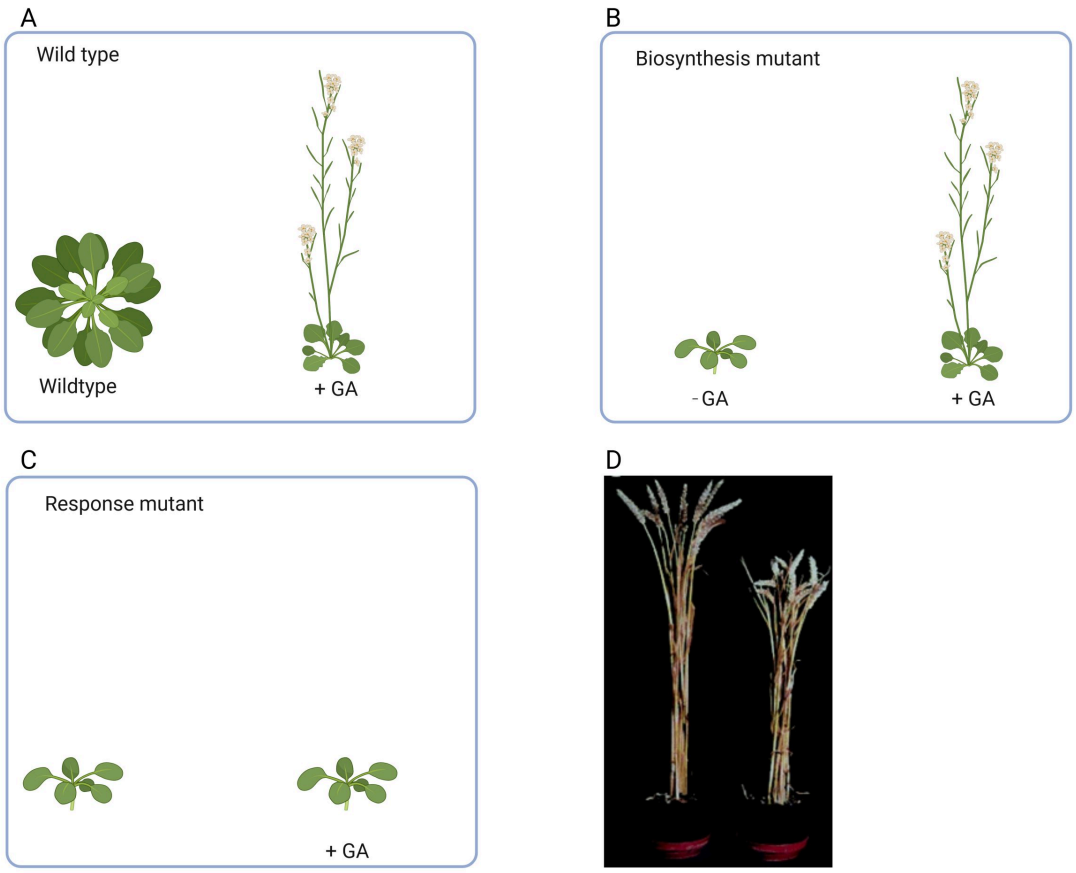


Figure 10 GA promotes plant growth. (A) When applied to wild-type plants, GA induces height. (B) A mutation that blocks GA biosynthesis will result in short plants, and exogenous addition of GA to a biosynthetic mutant restores growth. (C) However, a mutation that affects GA signaling, e.g., rht-1, leads to failure to sense GA. Thus, exogenous addition of GA does not restore the growth of a GA response mutant (C). (D) The wheat *rht-1* mutant (Peng et al.1999) provided a key trait for increased yield. *Image created by Faizo Kasule.*.

Read: 'Green revolution' genes encode mutant gibberellin response modulators

Semidominant traits often imply that the level of gene function is proportional to the expression of a trait. A classic example of this occurs in flower pigmentation. As seen in Figure 11, the white flowers are homozygous for a recessive allele that carries a null mutation for a gene required for anthocyanin biosynthesis. In the homozygous dominant, red flowers are produced. In the heterozygote, with only one functional copy of the gene, only half as much pigment is produced, resulting in a pink flower color.

Figure 11 Flowers of different genotypes for the flower color gene. *Image source: Rameshng, CC BY-SA 3.0, via Wikimedia Commons*.

Read: *Mutation Breeding* in *Plant Breeding Methods*.

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 mutations can arise spontaneously, or they can be induced by experimental methods.

Spontaneous Mutations

Spontaneous mutations arise "naturally." That is, the scientist makes no effort to increase the mutation rate. Spontaneous mutations generally occur at very low rates—approximately one mutation per gene per million gametes.

Mistakes in DNA Replication

Very rarely, incorrect bases are incorporated or 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

Several chemical agents are present in the environment with the 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. Of course, radiation can also be applied experimentally, and the use of radiation-induced mutations in plant breeding was discussed in *Plant Breeding Methods*. Several pathogens are also potentially mutagenic, including *Agrobacterium tumefaciens* (see Chapter 8).

Experimental Mutations

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 limits the type of tissues that 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 an alkylating agent and induces 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. The pollen and egg cells arise from different progenitor cells that may or may not be derived from the same ancestral cell mutagenized in the seed, depending on the reproductive development of the specific species. Therefore, a breeding scheme must be devised to generate families that segregate 1:2:1 +/+: *m*/+*: *m*/m**, where *m** represents a new mutant allele generated by the mutagenesis. These seeds can

be sown, and plants screened for traits of interest. The required breeding for self-pollinating species is very simple, whereas it would be much more elaborate and labor-intensive for dioecious species.

Pollen Mutagenesis

Pollen mutagenesis is the method of choice for a few species, such as maize. Pollen is treated with EMS and applied to the silks of an acceptor plant. The benefit is that the resultant seed is not chimeric; instead, all the embryo cells 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 helpful 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. Various approaches to gene cloning will be discussed in more detail in later sections of the course.

Early transposon mutagenesis experiments in snapdragon (*Antirrhinum majus*) provided key information about genes that control floral development (Carpenter and Coen, 1990). From a mutagenesis experiment using the Tam3 transposon (Figure 12), Coen and co-workers (Coen et al. 1990) discovered a mutant, which they referred to as *flo-613*.

The *flo-613* mutant produced inflorescence meristems in place of the floral meristem. Since they knew the sequence of the Tam3 transposon, these researchers were able to isolate the gene that was "tagged" by the transposon. The gene is called *floricaula* (*flo*) and belongs to a class of homeotic genes (Coen et al. 1990). Homeotic genes control organ morphogenesis, and mutations in homeotic genes transform one structure into another

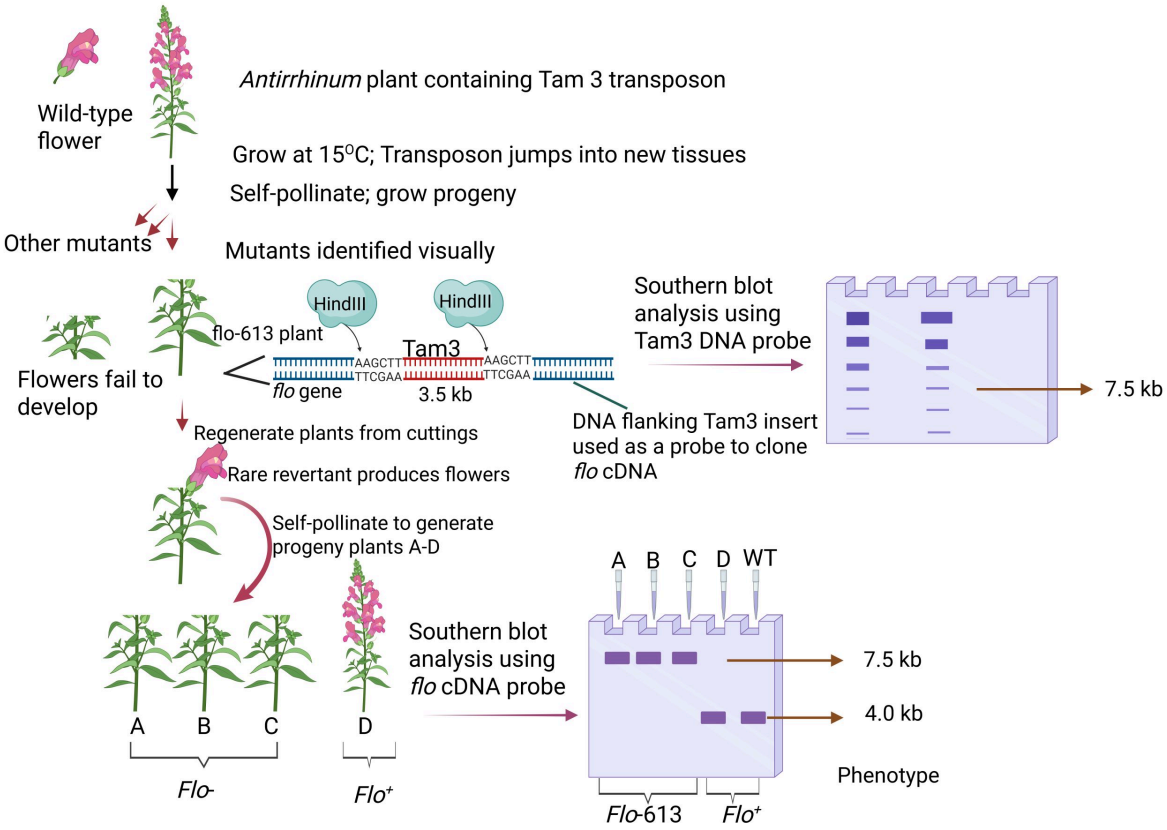


Figure 12 Cloning a plant gene by the activation tagging approach. Adapted from Watson, J.D., Gilman, M., Witkowski, J., and Zoller, M. (1997). Recombinant DNA. New York, NY: Scientific American Books.

Another important property of transposons that helped in the isolation of *floricaula* is their genomic instability, which results in unstable mutations. This occurs because the transposon can spontaneously excise from the gene it first mutated by insertion. Suppose excision of the transposon from the gene (e.g., *floricaula*) leads to reversion to wild wild-type phenotype. In that case, that is evidence that the cloned gene is responsible for the mutant phenotype observed.

Another strategy researchers use 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. The process of plant transformation will be discussed in detail in subsequent lessons.

Figure 13 An example of a variegated soybean mutant caused by the excision of a transposon from a flower pigment gene (Xu et al., 2010).

Arabidopsis has been used for T-DNA mutagenesis for several reasons. One is that it has a small genome size, helping to limit the number of transformants needed to obtain a mutation in the gene of interest. However, other plant species, including crops such as rice and other grasses, are now subjected to T-DNA mutagenesis to study gene function. T-DNA mutagenized seed can be obtained from various organizations. A few examples of these organizations can be found in the web links below.

Like the example of transposon mutagenesis above, a gene can be "tagged" by a T-DNA, and the tag is used to isolate the gene. However, T-DNA insertions are stable and thus disrupt the gene permanently (Figure 13).

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 successfully identified many genes that were not amenable to other mutational approaches.

- Read: *Floricaula:* A homeotic gene required for flower development in *Antirrhinum majus*
- Read: Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in Soybean [Glycine max (L.) Merr.]

Chapter 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.

References

Barry, C.S., and Giovannoni, J.J. 2006. Ripening in the tomato green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling. Proc. Natl. Acad. Sci. USA 103: 7923-7928. Available at http://www.pnas.org/content/103/20/7923.full.pdf

Carpenter, R., and Coen, E.S. 1990. Floral homeotic mutations produced by transposon mutagenesis in *Antirrhinum majus*. Genes & Dev. 4: 1483-1493. Available at http://genesdev.cshlp.org/content/4/9/1483.refs

Coen, E.S., Romero, J.M., Doyle, S., Elliot, R., Murphy, G., and Carpenter, R. 1990. *Floricaula*: A homeotic gene required for flower development in *Antirrhinum majus*. Cell 63: 1311-1322. Available at http://www.sciencedirect.com.proxy.lib.iastate.edu/science/article/pii/009286749090426F

Daxinger, L., and Whitelaw, E., 2010. Transgenerational epigenetic inheritance: more questions than answers. Genome Res. 20(12):1623-1628. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2989988/pdf/1623.pdf

Doebley, J., Stec, A., and Hubbard, L. 1997. The evolution of apical dominance in maize. Nature 386:485-488. Available at http://naldc.nal.usda.gov/download/68/PDF

Hirochika, H., Okamoto, H., and Kakutani, T. 2000. Silencing of retrotransposons in Arabidopsis and reactivation by the *ddm1* mutation. Plant Cell 12: 357-368. Available at http://www.plantcell.org/content/12/3/357.full.pdf

Jacobsen, S.E., and Meyerowitz, E.M. 1997. Hypermethylated SUPERMAN epigenetic alleles in Arabidopsis. Science. 277(5329): 1100-1103. Available at http://www.sciencemag.org/content/277/5329/1100.full.pdf

Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K., and Richards, E.J. 1996. Developmental abnormalities and epimutations are associated with DNA hypomethylation mutations. Proc. Natl. Acad. Sci. USA 93: 12406-12411. Available at http://www.pnas.org/content/93/22/12406.full.pdf

Lal, S., Choi, J., Shaw, J.R., and Hannah, L.C. 1999. A splice site mutant of maize activates cryptic splice

sites, elicits intron inclusion and exon exclusion, and permits branch point elucidation. Plant Physiol. 121: 411-418. Available at http://www.plantphysiol.org/content/121/2/411.full.pdf

Lertrat, K., and Pulam, T. 2007. Breeding for increased sweetness in sweet corn. Int. J. Plant. Breed. 1: 27-30. Available at http://www.globalsciencebooks.info/JournalsSup/images/0706/IJPB_1(1)27-30o.pdf

Nelson, O., and Pan, D. 1995. Starch synthesis in maize endosperms. Annu. Rev.. Plant Physiol. Plant Mol. Biol. 46: 475-496. Available at http://www.annualreviews.org/doi/pdf/10.1146/annurev.pp.46.060195.002355

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., and Harberd, N.P. 1999. "Green revolution" genes encode mutant gibberellin response modulators. Nature 400: 256-261. Available at http://www.nature.com/nature/journal/v400/n6741/pdf/400256a0.pdf

Pierce, Benjamin A. 2005: Genetics: A Conceptual Approach 2. W.H. Freeman and Company: 471-502. New York, NY.

Rahman, A., Wong, K., Jane, J., Myers, A.M., and James, M.G. 1998. Characterization of SU1 isoamylase, a determinant of storage starch structure in maize. Plant Physiol. 117: 425-435. Available at http://www.jstor.org.proxy.lib.iastate.edu/stable/pdfplus/4278295.pdf

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat. Genet. 43: 1160-1165. https://doi.org/10.1038/ng.942

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. 1993. *Arabidopsis thaliana* DNA methylation mutants. Science 260 **:** 1926–1928. https://doi.org/10.1126/science.8316832

Watson, J.D., Gilman, M., Witkowski, J., and Zoller, M. 1997. Recombinant DNA. Scientific American Books, New York, NY.

Xu, M., Brar, H. K., Grosic, S., Palmer, R.E., and Bhattacharyya, M.K. 2010. Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in soybean [*Glycine max* (L.) Merr.]. Genetics 184:53-63. https://doi.org/10.1534/genetics.109.107904

CHAPTER 4.

GENETIC PATHWAYS

Walter P. Suza; Philip W. Becraft; Faizo Kasule; and Madan K. Bhattacharyya

Introduction

Genes do not function in isolation but in groups of matching sets that perform their biological functions. When different genes function in different sequential steps of a biological process, it 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 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 result in 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 the behavior of those pathways. Also, many biosynthetic and regulatory systems are better described as networks because pathways often branch, 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 describes a process of converting a precursor substrate into a product. In addition to the precursor substrate and product, the biosynthetic pathway 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 development 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 specific amino acids were disrupted by mutations. Figure 1 shows the arginine pathway. Mutants that disrupt this pathway cannot grow on a medium lacking arginine but can grow if arginine is supplied. The same fundamental pathway also produces arginine in plants.

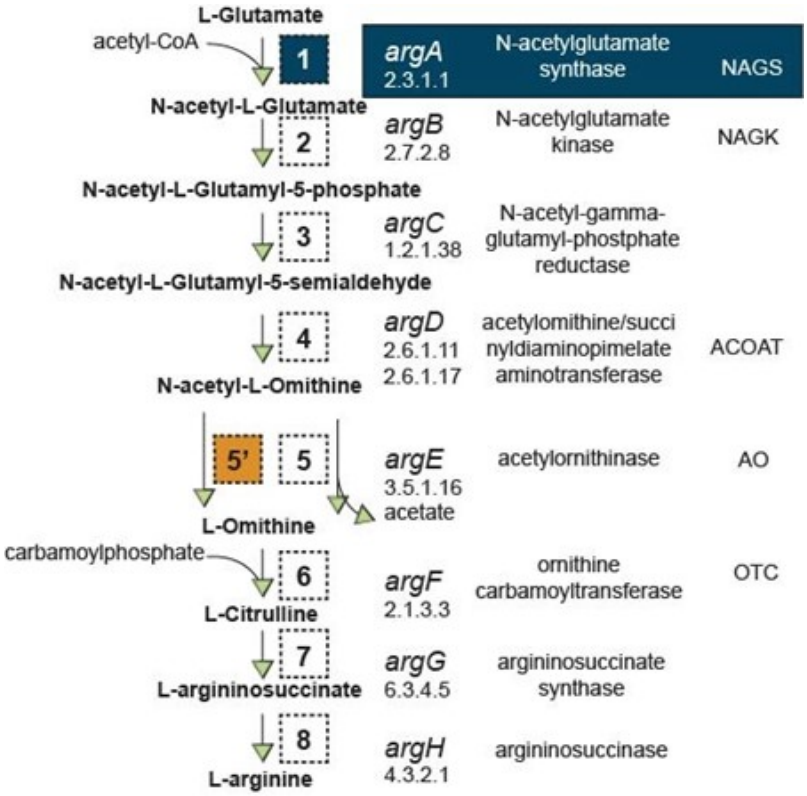


Figure 1 The arginine biosynthetic pathway. The boxed numbers represent chemical reactions. The italicized 4-letter symbol (e.g., argB) represents the gene that encodes the enzyme named to the right that catalyzes each step. *Adapted from Xu et al. (2006)*.

Mutations that disrupt the function of specific 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. This would block step 6 in the pathway (Figure 1), the conversion of L-Ornithine to L-Citrulline. As a result, the intermediate compound L-Ornithine will likely accumulate, whereas all the compounds that occur later in the pathway (i.e., L-Citrulline, L-arginosuccinate, and L-arginine) will be deficient.

For more information, please read *Bioinformatic analysis of an unusual gene-enzyme relationship in the arginine biosynthetic pathway among marine gamma proteobacteria: Implications concerning the formation of N-acetylated intermediates in prokaryotes* (Xu et al., 2006)

The Anthocyanin Pathway

The anthocyanin pathway is slightly more complex, although it is still simple regarding genetic pathways. 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. For example, the range of pigments seen in the petunia assortment below reflects variations in the anthocyanin pathway (Figure 2). 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.

Figure 2 Variation in petunia flower pigment due to variation in anthocyanin pigments. Color variants are due to variation in the biosynthetic pathway, and pattern variants are due to variation in the regulation of the pathway—courtesy of Park Seed Co.

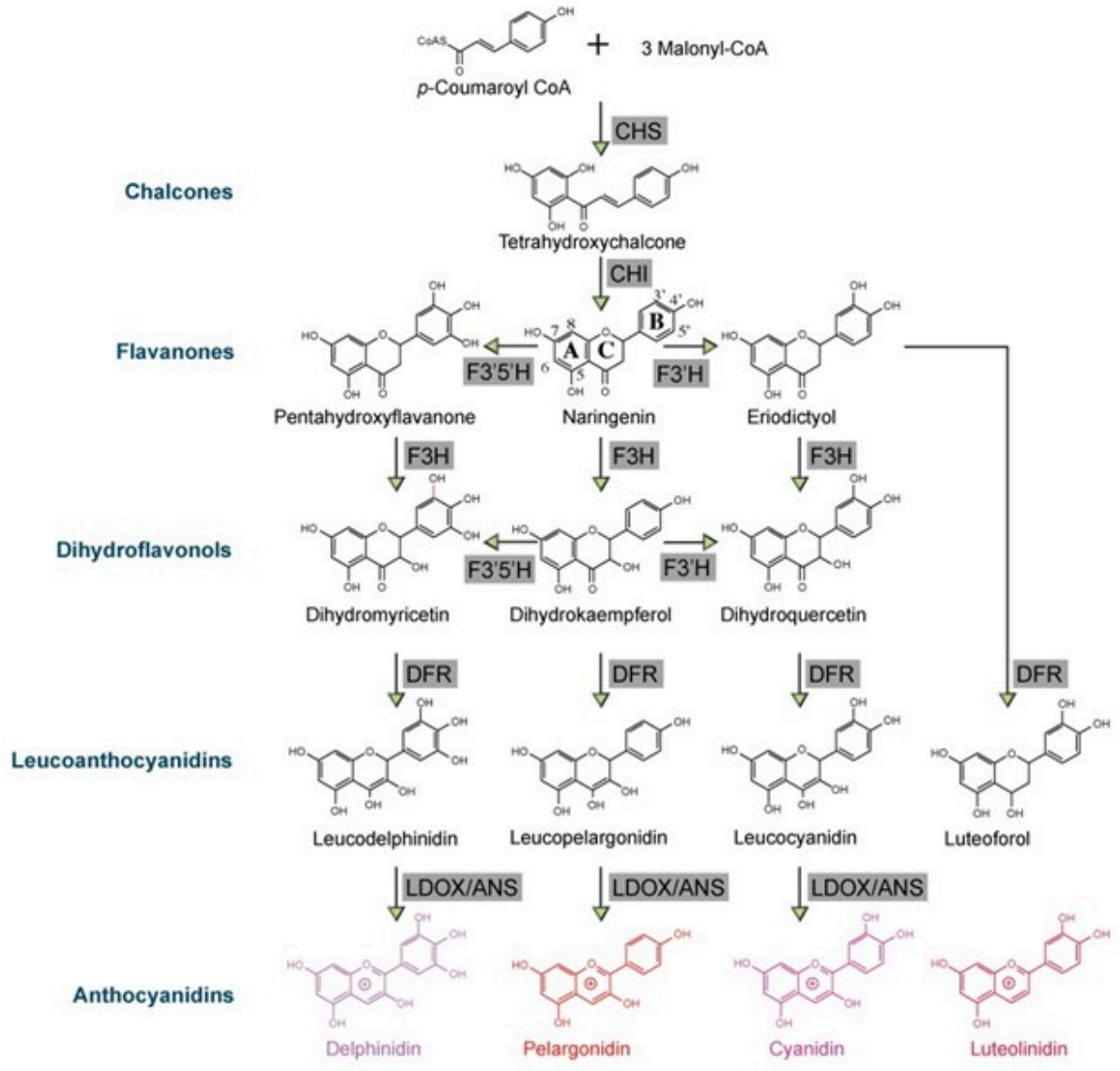


Figure 3 Biosynthesis of the four major classes of anthocyanins (Grotewold, 2006). Names of compounds are indicated and enzymes, in black boxes, are as follows: CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavanone 3'-hydroxylase; F3',5'H, flavanone 3',5'-hydroxylase; DFR, dihydroflavonol 4-reductase; LDOX/ANS, leucoanthocyanidin dioxygenase/anthocyanidin synthase. Modified from Grotewold, 2006, Annu. Rev.. Plant-Biol. 57:761-80.

The anthocyanin biosynthetic pathway is shown in Figure 3. The enzyme chalcone synthase catalyzes the first committed step in the pathway. Subsequent enzymatic steps catalyze the synthesis of intermediate compounds, resulting in the production of anthocyanidins. Chemical modifications, such as the 3' or 3',5' hydroxylation of the B ring, result in the production of different classes of anthocyanidins with varying characteristics of color. Anthocyanidins are subsequently glucosylated and transported into the cell vacuole, where they accumulate and are stored as anthocyanins. Intermediate compounds are col-

orless until the production of anthocyanidins. pH greatly affects the color properties, and the acidic pH of the vacuole is required for the colors normally observed for anthocyanin pigments.

Not all plants produce all these compounds, and some produce yet more varieties. For example, neither the rose nor maize genomes encode a flavanone 3',5'-hydroxylase enzyme, so these species do not normally produce delphinidin.

Note that the word "pathway" does not adequately describe anthocyanin biosynthesis. There is not a single pathway but rather a series of intersecting pathways. For example, the intermediate compound naringenin can be subject to 3 different chemical reactions, leading to different intermediate products and ultimately to any of the four anthocyanidins shown (Figure 3). And one of the products, dihydrokaempferol, can again be subject to 3 different chemical reactions, leading to 3 different anthocyanidins. Many biosynthetic systems are far more complex, and multiple pathways often intersect. For example, carbohydrate metabolism and amino acid metabolism interconnect at multiple points. For this reason, biosynthetic systems are often referred to as biosynthetic or metabolic **"networks."**

> For more information, please see *The Genetics and Biochemistry of Floral Pigments* (Grotewold, 2006).

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 4. 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 (not shown). The *bz2* mutant is deficient in an enzyme related to glutathione-S-transferase, which is required to transport 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 to produce cyanidin, which generates purple anthocyanin. In the mutant, only pelargonidin is synthesized, producing a red pigment.

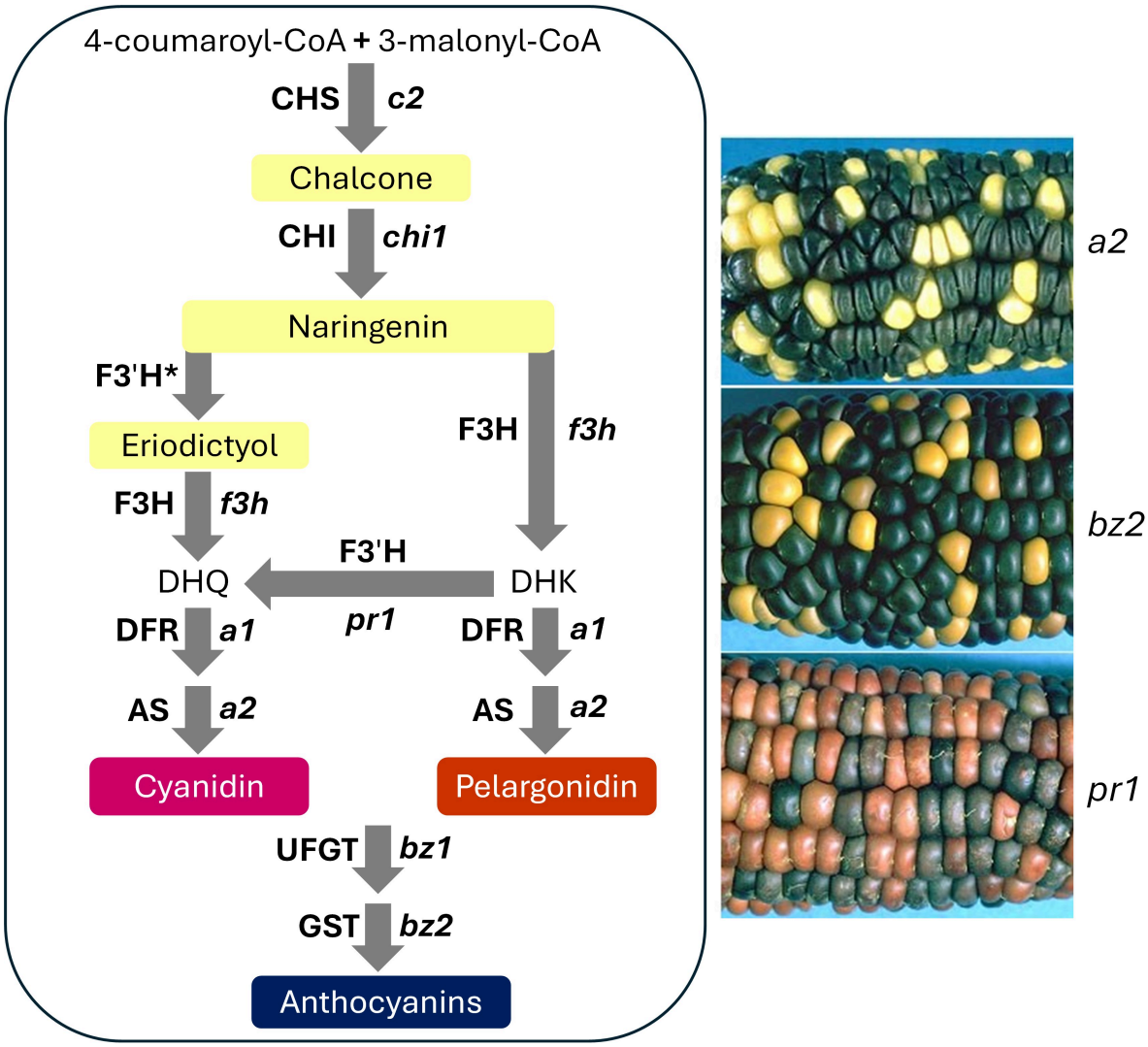


Figure 4 The maize anthocyanin pathway. Ears shown in the right are segregating for *a2* (*anthocyaninless2*), *bz2* (*bronze2*), and *pr1* (*red aleurone1*) kernels, respectively. Normal kernels are dark purple, nearly black, and the mutant kernels are lighter colors. Corn ear images courtesy of M.G. Neuffer and MaizeGDB.

Learning Activity

To answer the following questions, refer to the pathways illustrated in Figs 3 and 4. 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 4.

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

A. no change

B. increase

- 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.

C. decrease

D. can't predict

- **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.

What would you predict to happen to the levels of pelargonidin in a *pr1* mutant compared to normal?

A. no change

B. increase

- Once more, the expectation would be **B**, that pelargonidin would increase. Normally, some fractions 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.

C. decrease

D. can't predict

- **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.

The anthocyanin biosynthetic pathway has been subjected to extensive biotechnological manipulation. Some examples are reviewed in the accompanying article (Tanaka et al, 2008). One striking example is the production of lavender roses by the addition of a pansy gene encoding flavanone 3',5'-hydroxylase. This enzyme is normally not encoded by the rose genome, so the delphinidin pigment produced in these

transgenic roses produces beautiful flower colors not naturally found in roses (Figure 5). Similar strategies can be applied to other biosynthetic pathways to produce novel products for enhanced nutrition, pest resistance, or industrial applications.

Figure 5 Transgenic roses produce novel anthocyanin pigments. (From Tanaka et al., 2008)

For more information, please see *Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways* (Tanaka and Ohmiya, 2008).

Regulatory Pathways

Introduction

As mentioned, a regulatory pathway is a situation where one gene or gene product controls the expression or activity of another gene or gene product, which may also perform a regulatory function. As we discussed, gene function can be regulated at many points, from transcriptional to post-translational regulation. We will examine a couple of specific examples of regulatory pathways, but first, let us 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.

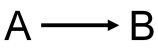
Positive vs. Negative Regulators

A key feature of regulatory pathways is that regulators might function positively or negatively. A positive regulator 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. In contrast, 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 is repressed.

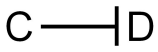
Example 1:

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



Example 2:

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



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 helpful way to think about this is that each component can have an "on" or an "off" state. Of course, many proteins can have intermediate activity levels, but for simplicity's sake, we will 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 example #1, if **A** is in the "on" or active state, it will function to switch **B** to the "on" or active state.

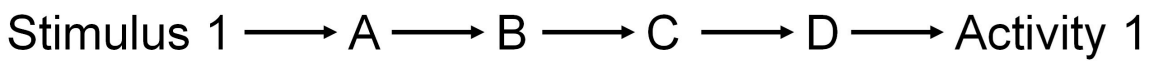
If a negative regulator is in the "on" state, it will function to switch the next factor to the "off" or inactive state. For the example #2, if **C** is in the "on" or active state, it will switch **D** to the "off" or inactive state.

Steps of Regulatory Pathways

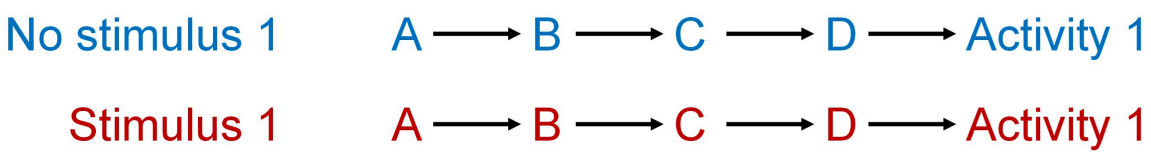
Regulatory pathways sometimes contain many, or sometimes few, steps. They can also contain a mix-

ture 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, **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 occurs in response to the stimulus.



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



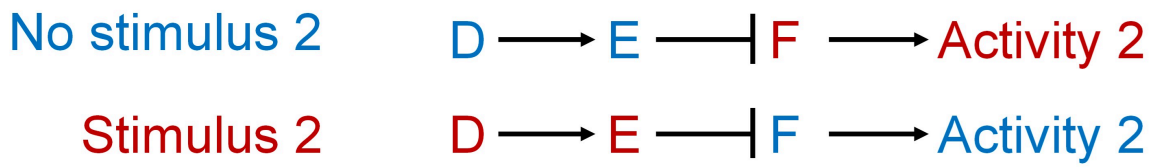
Or we can make a table to help keep track of the states of each component in the presence or absence of a stimulus.

Stimulus 1 →	A →	B →	C →	D →	Activity 1
no	off	off	off	off	no
yes	on	on	on	on	yes

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

In the preceding example, components **D** and **E** would be inactive in the absence of a 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 to 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**.



OR

Stimulus 2	D	E	F	Activity 2
no	off	off	on	yes
yes	on	on	off	no

Learning Activity

Try to complete each of the following tables to determine whether the hypothetical **Activity** will be *activated* or *repressed* in response to each **Stimulus**.

activated

Stimulus 3 →	**D ⊣**	**E ⊣**	**F →**	**Activity 3**

repressed

Stimulus 4 →	**D ⊣**	**E ⊣**	**F ⊣**	**Activity 4**

Answer

Click to reveal

activated

Stimulus 3 →	**D ⊣**	**E ⊣**	**F →**	**Activity 3**
no	off	on	off	no
yes	on	off	on	yes

repressed

Stimulus 4 →	**D ⊣**	**E ⊣**	**F ⊣**	**Activity 4**
no	off	on	off	yes
yes	on	off	on	no

Now that we have considered regulatory pathways from a hypothetical perspective, let us consider a couple of 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 in 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 aspects 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 6 shows examples of several rice mutants with varying levels of BR deficiencies.

Figure 6 Wild-type rice and several mutants (*osdwarf4-1*, *d11-4*, *brd1-1*), including a double mutant (*osdwarf4-1/d11-4*). BR deficiency caused dwarfism and altered growth habits. Bar, 30 cm. Image Credit: (Sakamoto et al., 2006).

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 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 7. Do not be concerned with memorizing the names of all these factors but focus on understanding the regulatory logic. The overall regulatory relationships of the pathway are depicted in panel **a**. The main target of regulation is a pair of closely related transcription factors called BZR1 and BZR2. As shown in panel **a**, 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 required for its activity.

As shown in Figure 7b, 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 another protein called BSU1, 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 from BIN2, thereby inactivating BIN2. Thus, BSU1 is a negative regulator of BIN2. When BIN2 is inactivated, that allows BZR1/2 to accumulate, enter the nucleus, and regulate the expression of genes that effect the BR response, including promoting plant growth and yield.

> For more information, please see *Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice* (Sakamoto, 2006).

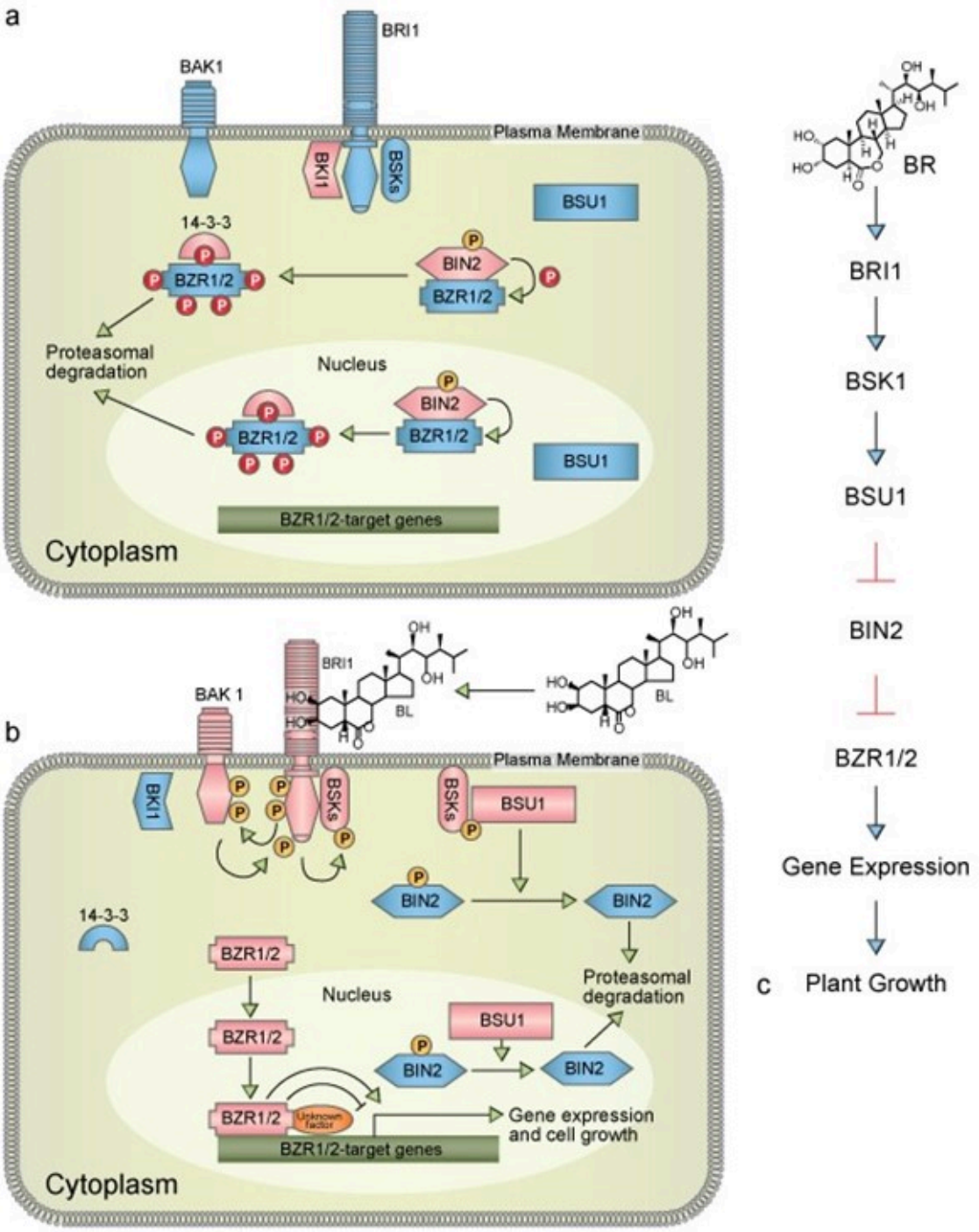


Figure 7 The brassinosteroid signaling pathway. a. The state of the pathway in the absence of BR hormone stimulus. b. Pathway activity in response to BR hormone. c. Diagrammatic representation of the regulatory relationships among BR pathway components. a and b are from Kim and Wang (2010).

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 mutation phenotypes will differ depending on whether the mutation affects the pathway's positive or negative regulator. A loss-of-function mutation in the *bri1* gene will block the perception of the hormone. Therefore, the pathway will not be activated, resulting in a dwarf phenotype. 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 the effects of deregulating hormone responses are more complex. Let us 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 the pathway's positive or negative regulator. 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

Several genetic principles make more sense considering that biological processes are controlled by multiple genes acting together. 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 are 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. Let us consider these concepts in detail within the contexts of biosynthetic and regulatory pathways.

Complementary Gene Action in Biosynthetic Pathways

As mentioned, one simple means by which complementary gene action can occur is when mutations

disrupt genes required for different steps in a pathway. Recall that gene complementation can only occur with recessive mutations, which, except in rare instances, would be loss-of-function mutations. Refer to the maize anthocyanin pathway shown in Figure 4A. Recessive loss-of-function mutations in the *c2* gene produce a "colorless" phenotype virtually identical to mutations in the *a2* gene. Now consider the following cross:

Parents: *A2+/A2+; c2/c2* (colorless) X *a2/a2; C2+/C2+* (colorless)

↓

F_1 : *A2+/a2; c2/C2+* (purple pigmented)

The first parent has a "colorless" mutant phenotype because it is homozygous for two nonfunctional copies of the *c2* gene. As such, it does not encode a functional chalcone synthase enzyme; therefore, it cannot produce any pigmented products of the anthocyanin pathway. The second parent carries two non-functional copies of the *a2* gene, does not encode a functional anthocyanidin synthase enzyme, and cannot synthesize anthocyanin pigments. The F_1 progeny received one functional copy of *A2+* and one functional copy of *C2+* from the parents; therefore, they encode functional anthocyanidin synthase and chalcone synthase enzymes. Hence, the progeny can produce purple anthocyanin pigments, and these two mutants complement.

It can also happen that two or more genes are required for a single enzymatic step. This occurs in the case of heterodimeric enzymes, enzymes composed of two different protein subunits, with each subunit encoded by a different gene. An example is ADP-glucose pyrophosphorylase, a key enzyme in starch biosynthesis (Hannah, 2005). Maize mutants disrupted in this enzyme are defective in starch production and accumulate high levels of sugars, starch precursors. This is the basis for some types of commercial sweet corn. ADP-glucose pyrophosphorylase is a heterodimeric enzyme requiring proteins encoded by two different genes, *Sh2* and *Bt2*. Mutations in either gene disrupt the function of the ADP-glucose pyrophosphorylase enzyme, therefore, are phenotypically very similar. As discussed in the previous paragraph, if a *sh2* mutant is crossed to a *bt2* mutant, the F_1 progeny inherits a functional copy of each gene and can produce a functional enzyme.

Epistasis in a Biosynthetic Pathway

The concept of epistasis can be readily explained in the context of a biosynthetic pathway. Consider again the anthocyanin pathway (Figure 3 and Figure 4) and the *a2* and *bz2* maize mutations depicted in Figure 4. Suppose an individual was homozygous for both mutations; what would be the phenotype of this double mutant? Because the *a2* mutation blocks the anthocyanidin synthase step, the pathway would stop with the production of leucoanthocyanidin intermediate compounds, which are colorless. Whether any subsequent steps, including those catalyzed by the *Bz2* gene product, are functional makes

no difference since no substrates are produced for them to act on. Thus, the double mutant phenotype would be colorless like *a2*. The *bz2* phenotype would be masked, and *a2* would be epistatic to *bz2*.

Learning Activity

Predict the phenotype of an *a2*; *pr1* double mutant

1. Predict the phenotype of an *a2*; *pr1* double mutant
2. Can't tell
3. Red like *pr1*; *pr1* would be epistatic to *a2*
4. They would all be light red; both genes would be partially epistatic
5. Colorless like *a2*; *a2* would be epistatic to *pr1*
    - The correct answer is **D) colorless like a2; a2 would be epistatic to *pr1***. Because of the *pr1* mutation, the left-hand branch of the pathway in Figure 4.4A would be blocked, and all the flux would be through the right-hand branch. However, this branch would be blocked by the *a2* mutation, preventing the conversion of the colorless leucopelargonidin to the pigmented pelargonidin. Thus, the double mutant would be colorless, and *a2* would be epistatic to *pr1*.

In a strictly linear pathway, such as the one shown for arginine biosynthesis (Figure 1), the rule of thumb is that the "upstream" gene (the one earlier in the pathway) is epistatic to the more "downstream" gene. However, as we see, this rule breaks down in branched pathways.

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 with similar roles in the pathway (e.g., both positive regulators of the pathway) would be expected to cause similar mutant phenotypes. If such mutants in different genes were crossed together, the F_1 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 us revert 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. What do we expect if the two mutations are combined into a double mutant plant? The *bri1* mutations disrupt the receptor function required to initiate signaling through the pathway. However, BIN2 functions to inhibit the activity of the 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 the *bri1* mutation blocks previous steps. 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*.

Learning Activity

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

Predict the phenotype of an *a2*; *pr1* double mutant
A gain-of-function mutation in the BZR2 gene

1. Unregulated growth
    - The correct answer is **A) unregulated growth**. 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. No effect
3. Can't predict
4. dwarf
    - The correct answer is **D) dwarf**. 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.

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

1. Can't predict
2. No effect
3. Dwarf
    - The correct answer is **C) dwarf**. 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.
4. Unregulated growth

Chapter Summary

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, resulting in 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 the activity of a pathway or network being targeted. The same genetic principles outlined in this section are used to design biotechnological strategies to alter pathway activities to produce the desired outcome on the trait of interest.

References

Grotewold, E. 2006. The genetics and biochemistry of floral pigments. Annu. Rev. Plant Bio. 57:

761-780. https://doi.org/10.1146/annurev.arplant.57.032905.105248

Hannah, L. 2005. Starch synthesis in the maize endosperm. Maydica. 50: 497-506. Available at http://www.maydica.org/articles/50_497.pdf

Kim, T.W., and Wang, Z.Y. 2010. Brassinosteroid signal transduction from receptor kinases to transcription factors. Annu. Rev. Plant Biol. 61: 681-704. Available at http://www.annualreviews.org/doi/pdf/10.1146/annurev.arplant.043008.092057

Park Seed Co. 1868. Available at http://parkseed.com

Sakamoto, T., Morinaka, Y., Ohnishi, T., Sunohara, H., Fujioka, S., Ueguchi-Tanaka, M., Mizutani, M., Sakata, K., Takatsuto, S., Yoshida, S., Tanaka, H., Kitano, H., and Matsuoka, M. 2006. Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. Nat Biotechnol. 24: 105-109.

Tanaka, Y., and Ohmiya, A. 2008. Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways. Current Opinion in Biotechnology.

19: 190-197. Available at http://wenku.baidu.com/view/59eedc68a45177232f60a2c2.html

Xu, Y., Glansdorff, N., Labedan, B. 2006. Bioinformatic analysis of an unusual gene-enzyme relationship in the arginine biosynthetic pathway among marine gamma proteobacteria: implications concerning the formation of N-acetylated intermediates in prokaryotes. BMC Genomics. 7:1471-2164.

CHAPTER 5.

DNA MARKERS

Walter P. Suza; Faizo Kasule; Thomas Lübberstedt; and Madan K. Bhattacharyya

Introduction

Molecular or DNA markers reveal sites of variation in DNA. Variability in DNA facilitates the development of markers for **mapping** and detecting traits. Any DNA sequence can be genetically mapped, such as genes that lead to plant phenotypes. The prerequisite is that a polymorphism must be available for the sequence to be mapped; that is, there must be two or more different alleles. This can basically be a **single nucleotide polymorphism (SNP)**, a single nucleotide variant at a specific position within the target sequence, or an insertion/deletion (**INDEL**) polymorphism. Target sequences can be amplified by various methods, including the **Polymerase chain reaction (PCR),** and subsequently visualized to generate "molecular phenotypes" comparable to visual phenotypes, which can be observed using appropriate equipment. Various molecular methods have been developed to visualize SNPs or INDEL polymorphisms at a low cost and high throughput, as discussed earlier. The main use of those SNPs and INDEL polymorphisms is as **molecular markers**. By genetic mapping as described above, linkage between genes affecting agronomic traits or morphological characters, and DNA-based SNP or INDEL markers can be established. It can be more effective in the context of plant breeding to select indirectly for markers (DNA or non-DNA) than directly for target traits. Reasons can include lower costs for marker analyses, the ability to run multiple such assays (for DNA markers) in parallel, the ability to select early and discard undesirable genotypes, or to perform selection before flowering, as well as codominant inheritance of markers, among others. This chapter discusses various types of markers used in plant breeding.

General Properties of DNA Markers

DNA markers are readily detectable DNA sequences whose inheritance can be monitored. The advantage of DNA-based markers is that they are independent of environmental factors. An ideal DNA marker should possess the following properties:

1. Be highly polymorphic
2. Display co-dominant inheritance (to discriminate homozygotes from heterozygotes)
3. Occur at high frequency in the genome

4. Display selective neutral behavior
5. Provide easy access
6. Be simple to evaluate with the available set of tools
7. Display high reproducibility, and
8. Facilitate easy exchange of data between laboratories

Historically, DNA markers can be grouped into three main categories: (1) hybridization-based markers, e.g., restriction fragment length polymorphism (RFLP) markers; (2) PCR-based markers, e.g., amplified fragment length polymorphism (AFLP), and simple sequence repeat (SSR); and (3) sequence or chip-based markers, e.g., some procedures for detecting single-nucleotide polymorphism (SNP) markers. Examples of molecular markers belonging to the above three categories are further discussed. Recall that you learned about the principle of PCR and the application of PCR in molecular marker analysis in *Plant Breeding Methods* and *Molecular Plant Breeding*.

DNA Versus Non-DNA Markers

Genetic markers are broadly classified into two groups. (1) DNA markers: those based

on the **detection of DNA**. (2) Non-DNA markers: those based on visually distinguishable traits, also referred to as **morphological markers** (e.g., flower color or seed shape); and those based on gene products, referred to as **biochemical markers** (e.g., RNA, protein, and other cellular metabolites).

The advantage of DNA markers is that they are not affected by environmental factors. However, the presence of a particular DNA sequence may not always lead to the expected expression of a trait of interest. This is because the expression of a specific allele depends on environmental conditions and interaction with other genes. Thus, even though an allele with a known effect on a particular trait is present, it might not result in the expected phenotype. Therefore, DNA markers serve as a measure of an individual's genetic potential. The equivalent in human genetics is the concept of risk. Based on DNA information, it is possible to predict a patient's risk of developing a particular condition (e.g., a 30% chance of getting pancreatic cancer at a certain age). However, whether this condition is expressed depends on other circumstances. In contrast, if RNA- or metabolite-based biomarkers for this cancer type are available, the onset of this condition can be predicted with high accuracy. Thus, non-DNA markers are indicative of an individual's realized potential.

The advantage of morphological markers, also known as visible markers, is that they are generally easy to score. However, morphological markers are affected by environmental conditions, making their use less reliable across environments. Also, morphological markers are limited in number compared to the abundance of DNA markers. Biochemical markers are affected by the developmental stage of the plant

and the cell type from which they are isolated. This must be carefully selected. This is a major difference compared to DNA-markers, which are stable and valid, independent of the tissue from which the respective DNA has been isolated. The World Health Organization defines a **biomarker** as any parameter that can be used to measure an interaction between a biological system and an environmental agent, which may be chemical, physical, or biological. Therefore, the diagnosis of a disease condition and the identification of possible treatments require the use of biomarkers. In conclusion, the term "biomarker" is broadly defined and may encompass both DNA- and non-DNA-based markers. However, the term biomarker is sometimes used in a narrower sense to refer to biochemical non-DNA markers.

To understand the relationship between DNA markers and non-DNA markers, review the pathways by which genetic information in **deoxyribonucleic acid (DNA)** is transferred from DNA to **ribonucleic acid (RNA)** molecules (called **transcription**), and then transferred from RNA to a protein (termed **translation**) by a code that specifies the amino acid sequence of the protein (see Chapter 1). Epigenetic mechanisms (Chapter 2), including DNA methylation and demethylation, as well as histone acetylation/deacetylation, may also impact gene expression (see Figure. 1).

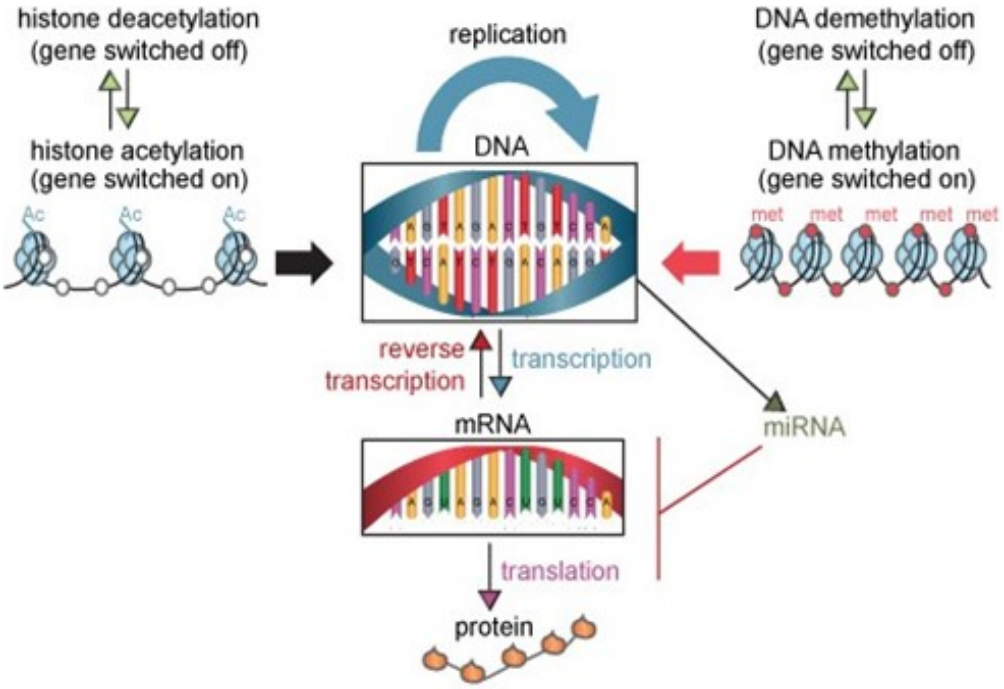


Figure 1 Scheme of genetic (and epigenetic) information pathways from DNA to RNA to protein. DNA, RNA, and proteins can be used as markers. If the sequence of the protein is known, it may be used to track the DNA (the gene) from which it was encoded. Variations in the DNA sequence result in variations in RNA and protein sequences. If a change in the amino acid sequence of an enzyme results in a change in its function, the observable phenotype (morphological or biochemical) can be used as a marker. Non-coding RNAs (e.g., miRNA) are also important. Many miRNA genes are expressed in specific tissues and developmental stages to regulate the expression of specific genes by affecting mRNA stability and translation.

Classical DNA Markers

RFLP

RFLPs involve cutting DNA into fragments, followed by separating the fragments in a gel and then transferring them to a nylon membrane for hybridization with a radioactive probe. The membrane is exposed to X-ray film. Patterns of variability in fragment size, or polymorphisms, are viewed on the autoradiograph, and RFLP patterns are scored among genotypes.

Strengths of RFLP

- Co-dominance
- No sequence information is required
- Simplicity not requiring costly instrumentation
- RFLP probe sequences can be used to develop additional markers, e.g., Indel

Weaknesses of RFLP

- Analysis requires large amounts of high-quality DNA
- Low genotypic throughput (few loci detected per assay)
- Difficult to automate
- Use of radioactive probes restricts the analysis to specific laboratories
- Probes must be physically maintained, not allowing sharing between laboratories
- Low selection for polymorphic parental lines makes it challenging to complete the RFLP map

SSR

SSRs are widely used markers based on the high rate of variation in microsatellite loci. SSRs represent a few to hundreds of highly variable tandem copies of DNA repeats. Such tandem repeats, typically consisting of one to four bases, are widespread in higher organisms. Many different microsatellite loci (>100,000) can be present in any plant species. SSRs are a result of slippage during DNA replication or unequal crossover during meiosis. SSRs were discussed in the context of *Plant Breeding Methods*.

Variation in SSRs is observed by developing locus-specific primers that anneal to sequences flanking the repeat region; PCR is subsequently used to amplify the target region. Alleles (fragments) are visualized as bands with different migration patterns on a gel after electrophoresis. However, more recently, capillary electrophoresis has been employed, which also enables multiplexing of up to approximately 16 SSRs

per capillary. It is essential that you are familiar with some of the most common methods for developing PCR primers for SSR analysis.

Strength of SSR markers:

- Hypervariable, multiple alleles (high PIC)
- *In silico* development is straightforward

Weaknesses of SSR **markers:**

- Capability for multiplexing limited (max. 10-15)
- Affects costs/datapoint
- Few intragenic SSRs

AFLP

The AFLP procedure combines RFLP and PCR. In AFLP, genomic DNA is digested with restriction enzymes, followed by a ligation step where adapters are added to both ends of the restriction fragments. PCR is carried out on the adapter-ligated mixture, using primers that target the adapter, but that vary in the base(s) at the 3' end of the primer. Click here to learn about the steps to detect and analyze AFLP markers.

Strength of AFLP markers:

- High marker index
- Amenable to automation
- Robust
- No prior sequence information required
- Special applications: Gene family profiling; Methylation assay
- Established service company: KeyGene

Weakness of AFLP markers:

- Random loci might differ between populations
- Dominant marker system

Current DNA Markers

SNP

SNPs are single-nucleotide differences in the DNA of two individuals. However, the term SNP may be somewhat misleading because differences in alleles at polymorphic AFLP or RFLP loci can also result from single-base changes. A SNP can be linked to a gene of interest or part of a functional gene itself. The advantages of functional markers were discussed in *Plant Breeding Methods*. The DNA used as a SNP marker cannot be analyzed by gel electrophoresis in the same manner as other classical markers because the base pair number and length of the DNA for two alleles will be the same. Some of the principles exploited in SNP detection are shown in Figure 2 below.

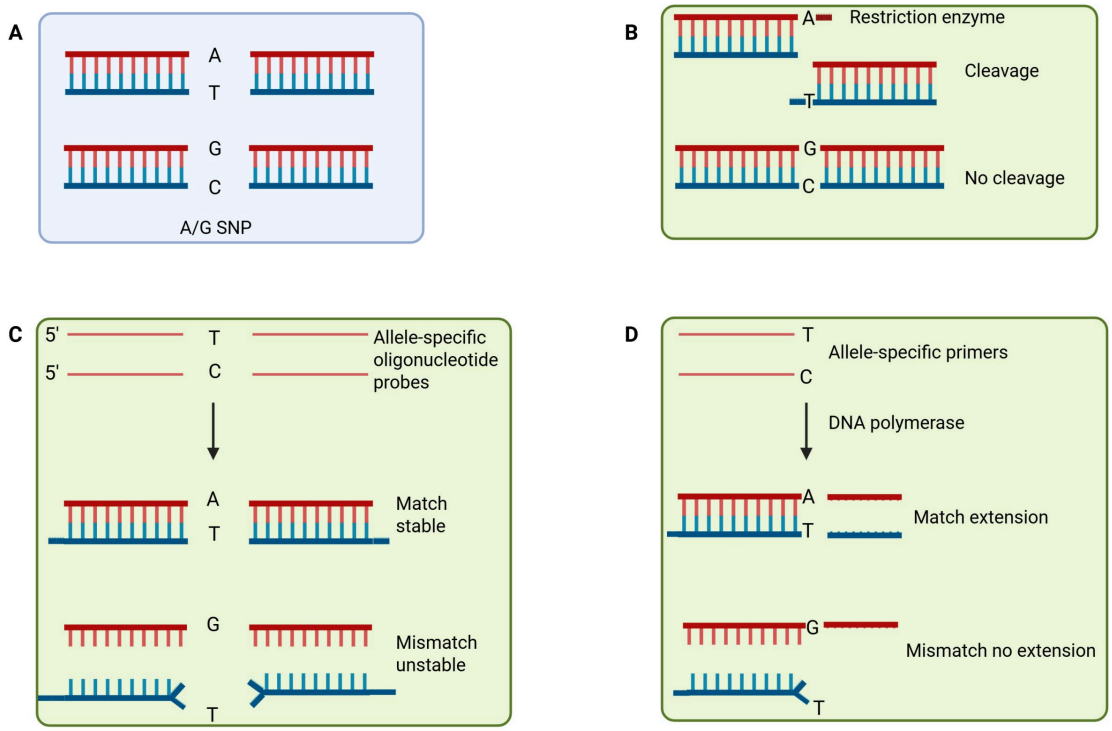


Figure 2 Examples of methods used in SNP genotyping, adapted from Syvänen (2001).

In Figure 2A, an A-to-G transition is described, including the various methods that can be used to detect the A- and G-alleles. In Figure 2B, restriction enzymes may be used for allele-specific cleavage of the target DNA when a SNP changes the restriction site for the enzyme. An example of this method was discussed in *Plant Breeding Methods*. In Figure 2C, two short probes are used to detect the polymorphism by hybridization. Only the probe that perfectly matches the target will be stable; a mismatch will be unstable. The probes are usually labeled with a fluorescent dye or radioisotope for detection by a laser

scanner or autoradiography (e.g., Southern blot analysis). In Figure 2D, a method called **primer extension** is described. Primer extension uses two allele-specific primers that anneal to the target sequences adjacent to the SNP and have a nucleotide that is complementary to the SNP at the 3' end. Only primers that match the target sequence ideally will be extended by the DNA polymerase. Defining the scale of the genotyping project is crucial for selecting an appropriate SNP identification approach. The table below provides cost estimates for SNP genotyping of 1000 individuals.

Overall, various genotyping approaches are available, ranging from low-throughput to high-throughput. Some platforms permit users to select custom SNPs, but the highest-throughput assays are available only in fixed content. Not all custom SNPs are compatible with every format, and multiple SNPs may be required to carry out most projects targeting specific SNPs. However, there are still trade-offs in terms of throughput, specifically between the number of samples and the number of SNPs to be analyzed. Ultimately, cost will dictate how a SNP project is designed. Regardless of the study, design, quality control, and tracking are crucial to the project's success. Laboratory Information Management Systems (LIMS) are important in every study design. The following are examples of SNP genotyping systems that are commonly used by plant breeders.

> Read: Syvänen, A. 2001. Accessing genetic variation: genotyping single-nucleotide polymorphisms. Nat Rev Genet. 2:930-942

Taqman Assay

TaqMan SNP assays are based on PCR using four oligonucleotide primers: (1) A set of forward and reverse primers that are designed and tested for each SNP, and (2) Two hydrolysis (Taqman) assay probes conjugated with fluorescent dyes and quenchers. Taqman probes are designed to anneal within a region of the PCR fragment resulting from the annealing of the forward and reverse primers. The quencher ensures that a dye does not fluoresce before Taqman probes have annealed to their target during PCR. The PCR reaction is catalyzed by a polymerase enzyme with 5' to 3' exonuclease activity. The 5' to 3' exonuclease activity is required to cleave the quencher from the dye, allowing fluorescence to be produced during PCR amplification.
Watch this video to learn more about how Tagman works.

Watch this video to learn more about how Tagman works.

MassArray System

The MassArray system utilizes highly multiplexed PCR reactions to simultaneously screen multiple mutation sites by primer extension, combined with Matrix-Assisted Laser Desorption/Ionization-Time of

Flight mass spectrometry (MALDI-TOF-MS). The system provides a rapid and quantitative readout, allowing for the detection of mutations, gene copy number, methylation status, and the level of expression of allelic variants. Up to approximately 20 SNPs, multiplied by about 400 samples, can be analyzed at a time. The three key steps in SNP analysis using the Sequenom system are (1) Target amplification, (2) Primer extension, and (3) Signal detection and ratio analysis.

> Read: Polyploid SNP Genotyping Using the MassARRAY System

> Watch: MassARRAY® System by Agena Bioscience
> Watch: MassARRAY® System by Agena Bioscience

GoldenGate Assay

The GoldenGate assay involves the addition of biotin to genomic DNA, which immobilizes the DNA on avidin-coated particles that bind biotin. The assay utilizes three oligonucleotide primers; two of these (P1 and P2) are specific for the two SNP alleles, and the third (P3) is a locus-specific primer tagged with a sequence for capture on solid support. In the reaction, allele- and locus-specific primers anneal to the genomic DNA, followed by extension using a DNA polymerase. After extension, the products are ligated to the tag sequence by a ligase. PCR primers containing fluorescence labels recognize the P1, P2, and P3 sequences. Extension products, containing fluorescence labels, are captured on the BeadArray, which includes complementary tag sequences for fluorescence detection.

> Read: Single-nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay.

> Read: Toward genome-wide SNP genotyping

INDEL

Insertions and deletions (Indels) cause changes in the DNA sequence through the addition or removal of nucleotides (Figure 3). Indels can range in size from one or a few bases to multiple megabases. Small

deletions—from a few base pairs to kilobases in length—most often arise from unequal crossover during meiosis.

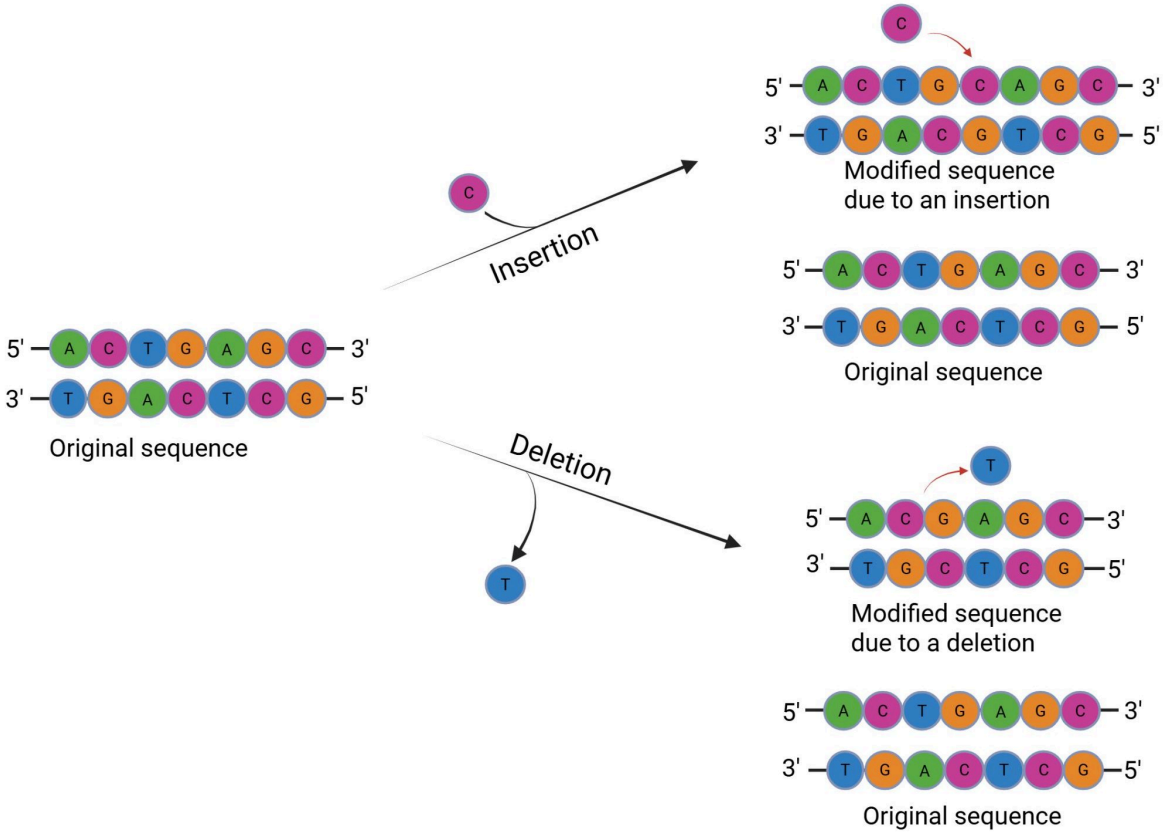


Figure 3 An insertion mutation caused by the insertion of a Cytosine base (C) into the genome, and a deletion mutation caused by a deletion of a Thymine base (T) in the genome. Indels can arise during DNA replication, when the nascent and template strands slip out of alignment, hence creating extra or missing nucleotides that are fixed as insertions or deletions in the new DNA strand.

Read: Indel arrays: an affordable alternative for genotyping.

General Application of DNA Markers

The rapid development of molecular genetics tools during the 1980s and 1990s enabled the identification of genetic variation at the molecular level, based on DNA sequence changes and their association with phenotypic differences. These DNA variations, or **polymorphisms**, can be exploited by plant breeders as **markers** for traits of interest. The growing availability of sequence data from large-scale sequencing projects, along with advances in next-generation sequencing technologies and bioinformatics, has dramatically reduced the cost of marker discovery and application. The principles and use of PCR and

molecular markers were discussed earlier in *Crop Genetics*, while their application in plant breeding is covered in greater depth in *Molecular Plant Breeding*. In this chapter, we briefly highlight examples of how DNA markers are used in **genetic fingerprinting**, **gene tagging**, and **backcrossing**.

Genetic Fingerprinting

Genetic fingerprinting is a method that utilizes the unique characteristics of DNA to classify individuals into distinct or similar groups. Because the genomes of different individuals contain polymorphisms, a unique DNA profile can be established for a particular organism. This profile is specific to that individual, and as unique as a fingerprint.

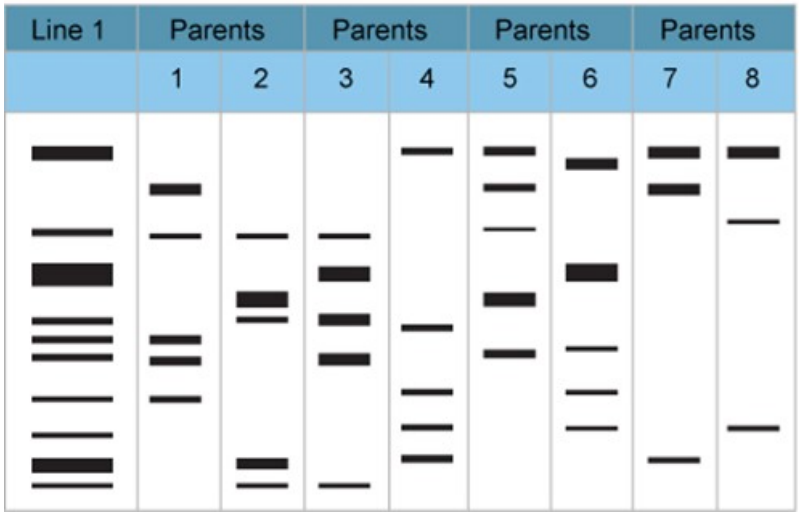


Figure 4 Genetic fingerprinting can be used to confirm parentage.

The concept of fingerprinting is increasingly being applied to determine the ancestry of plants and animals. Genetic fingerprinting can be used in the breeding of endangered species or commercially important crops, as it helps guarantee the authenticity of the plants. With the ability to obtain highly specific DNA profiles, genetic fingerprinting can help prove that new varieties qualify for intellectual or variety protection. For commercially important crops that are difficult to characterize phenotypically, genetic fingerprinting is a crucial tool for identifying genetic diversity within breeding populations. One of the earliest methods of genetic fingerprinting used hybridization-based RFLP markers. Examples of genetic fingerprinting data are provided in Figure 4 and Figure 5.

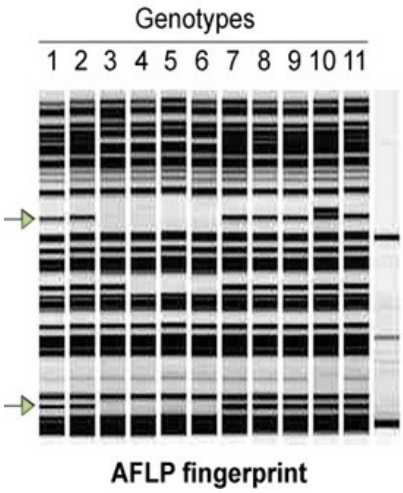


Figure 5 Example of a genetic fingerprint based on AFLP analysis.

Learning Activity

How can genetic similarity be estimated among the different genotypes in Figure 5?

Answer

Click to reveal

- By calculating genetic similarity (i.e., genetic similarity = number of common bands divided by the number of all bands).

Genetic fingerprinting can be applied at all phases of cultivar development. The phases are described below.

Phase 1: Identifying Genetic variation:

1. In parent selection
2. In recurrent selection
3. In assigning individuals to heterotic pools
4. In choosing genetic resources

Phases 2 and 3: Developing variety parents or testing hybrids:

1. To measure heterozygosity to predict hybrid performance
2. In conducting backcrossing

Phase 4: Seed multiplication and variety protection:

1. To ensure the purity of hybrids and blends
2. For variety approval
3. To identify “essentially derived varieties” (EDV)

Gene Tagging

Recall that in *Crop Genetics*, you learned that linkage analysis tests for the co-segregation of a marker and a trait of interest by calculating the frequency of recombination. Using sets of markers flanking a trait of interest, the likelihood of a recombination event occurring between two markers is related to the distance between them. Thus, the further the marker is from the gene controlling the trait, the greater the chance there is that recombination will occur between the gene and the marker. This implies that markers closest to the trait will segregate with the trait of interest. A marker linked to a gene controlling a gene of interest can serve as a “tag” for that gene/trait. An example of gene tagging is provided in Figure 6.

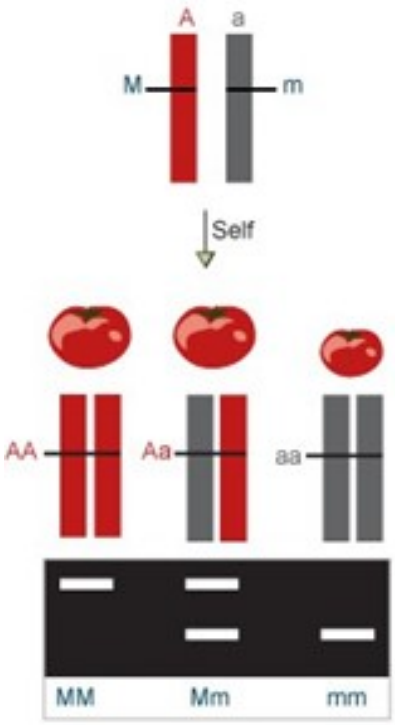


Figure 6 An example of gene tagging with a molecular marker is completely linked to a trait of interest. A hypothetical gene, "A", controls fruit size in tomatoes and is dominant over "a". A co-dominant marker is available to identify individuals carrying either of the fruit size alleles. The marker "M" is linked to the dominant allele, and "m" to the recessive allele. The detection of markers M and m by PCR produces fragments that can be separated by gel electrophoresis.

Learning Activity

What might be the benefit(s) of tagging a gene with a molecular marker for a trait that can be phenotypically scored?

Answer

Click to reveal

- To allow early selection, allowing individuals without the trait to be curled early
- Phenotypic markers are affected by the environment
- The marker evaluation can help to fix the desired allele (A)
- For transgenic traits, plants homozygous for the recessive allele (wild type) can be identified for regulatory purposes

Backcrossing

Marker-assisted backcrossing involves three steps:

- Step 1: Selection of donor allele at the markers linked to target genes to reduce loss of target allele due to recombination. In this step, markers are useful if a recessive allele controls the trait or when multiple resistance genes are to be obtained from the donor. Additionally, markers are useful for identifying environmentally sensitive genes and for expensive phenotypes, such as grain quality.
- Step 2: Selection of recurrent parent allele at other linked markers. This step helps reduce linkage drag when introgressing wild or exotic germplasm.
- Step 3: Selection of recurrent parent allele at unlinked markers throughout the genome (**background selection**). It is all a matter of probability to identify backcross progeny that are like the recurrent parent (Figure 7). Markers specific to the recurrent parent, referred to as **background markers**, can therefore be used to identify progeny that are genetically more similar to the recurrent parent.

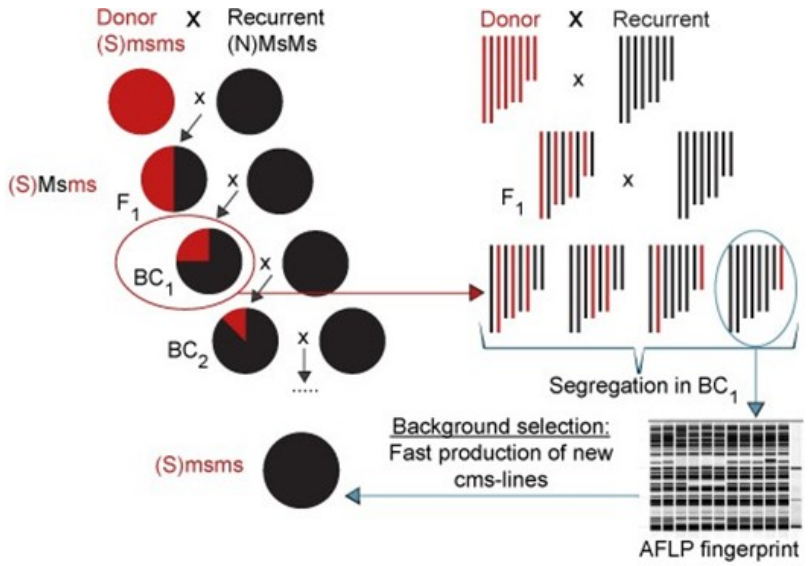


Figure 7 Development of male sterility by marker-assisted backcrossing in maize. A male sterile donor is crossed with a fertile recurrent parent. Red and black charts depict the proportions of donor and recipient genomes, respectively. Red and black bars represent the chromosome segments of the donor and recurrent parent, respectively. Progeny containing the most significant proportion of the recurrent parent genome can be detected as early as the BC1 generation (red circle) using molecular markers and genetic fingerprinting (blue circle). Overall, the use of markers helps increase the pace of production of new male sterile lines.

Markers are useful for **foreground selection** of lines that have the donor allele in a heterozygous condition. An example of the use of markers for foreground selection is illustrated in Figure 8. Without a

marker, it would be difficult to distinguish between progeny that are heterozygous for the male sterility trait (Msms) and those that are homozygous (MsMs), as both scenarios result in fertile plants. The use of a co-dominant marker linked to Ms/ms heterozygotes helps identify heterozygotes and eliminates the need to expend time and resources for selfing and scoring individuals based on pollen production.

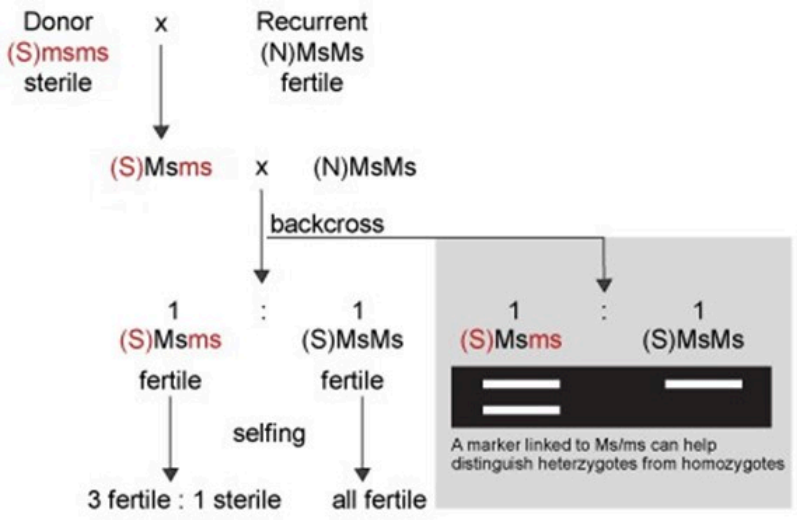


Figure 8 The use of molecular markers for foreground selection. Backcross of (S)Msms to (N)MsMs produces fertile plants, but of different genotypes (Msms or MsMs). Selfing the MsMs BC1 progeny will produce all MsMs fertile plants. Selfing of BC1 Msms progeny will produce fertile and sterile plants in the ratio of 3:1. The use of a linked marker will help eliminate additional work to self and phenotypic screening of the plants.

Biochemical markers

RNA-Based Markers

Chapter 1 discussed the steps in gene expression, including **transcription** (the production of mRNA from DNA) and **translation** (the production of proteins from mRNA). Not all genes produce mRNA that can be translated into proteins. Specific genes are transcribed into non-coding RNAs (e.g., microRNAs–miRNAs or short-interfering RNAs–siRNAs) that serve a regulatory function during plant growth and development. A gene can be either "on" or "off" depending on the cell type, stage of development, and environmental signals, meaning that at any moment each cell makes coding and non-coding RNA from only a proportion of its genome.

Protein-Based Markers

Common protein markers are **isozymes**. Isozymes are enzymes with similar function derived from more than one locus. Isozymes are encoded by gene families resulting from duplication events. Isozymes are different from **allozymes** in that allozymes represent one enzyme derived from a single locus. Isozymes are analyzed using a procedure called electrophoresis, a technique that separates macromolecules on a

gel through the application of an electric field and specific chemical staining (Figure 9). To be proper as markers, isozymes must be electrophoretically resolvable (i.e., bands can be clearly separated for visualization on a gel) and detectable by various in-gel assay methods.

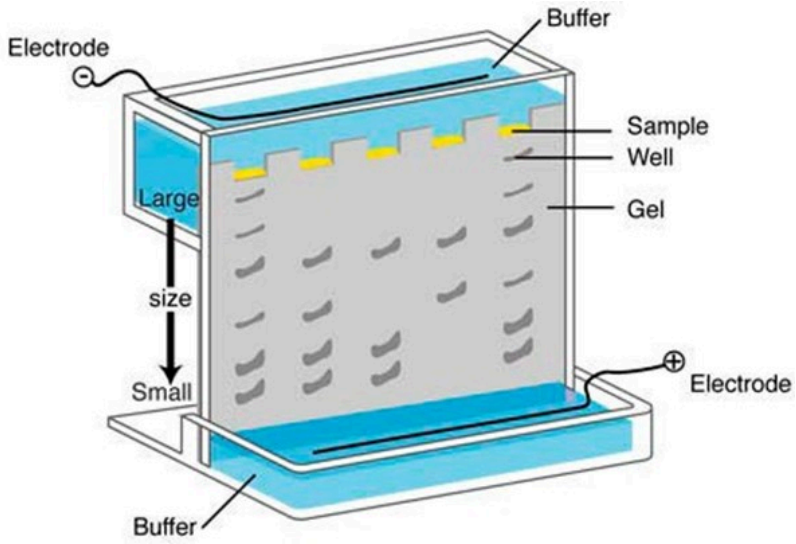


Figure 9 Electrophoresis is a laboratory technique used to evaluate isozymes. *(Illustration from NIH-NHGRI 2011)*.

The advantage of isozymes is that they are robust and highly reproducible. Additionally, isozymes exhibit codominant expression, meaning that both homozygotes can be distinguished from the heterozygote, and neither allele is recessive. However, isozymes are gene products, so they reveal only a small subset of the actual variation in DNA sequences between individuals and do not reveal variation in the non-coding regions of the genome. Other limitations of isozymes as markers include: (i) data complexity as a result of dimers or multimers of the enzymes; (ii) multi- allelic and multi-locus systems can interpret the banding patterns difficult; (iii) the system is limited to those enzymes that can be detected in situ, resulting in a narrow coverage of the genome; (iv) relatively few biochemical assays are available to detect isozymes; and (vi) the assay is based on a phenotype, and thus sensitive to the environment.

Currently, isozymes are used mainly for germplasm identification and population genetics studies. Other examples of the application of proteomic approaches are listed below.

1. Two-dimensional polyacrylamide gel electrophoresis was used to detect polymorphic protein markers in several plant species.
2. A proteomic approach was used to identify protein markers in lung cancer
3. Isozymes were useful in developing the linkage map for tomato
4. Evaluation of isozyme uniformity in a wild extinct insular plant [*Lysimachia minoricensis* J.J. Rodr. (Primulaceae)

Metabolite-Based Biomarkers

In human health, transitions from a healthy state to a diseased state can often be characterized by changes in key cellular metabolites. A similar concept can be applied to plants, where specific metabolites can serve as biomarkers during growth and development. For example, Tarpley et al. (2005) identified a set of biomarker metabolites associated with different developmental stages in rice.

The main advantage of metabolite-based markers is that their concentrations are more directly linked to phenotypic traits than those of DNA markers. Consequently, establishing a set of metabolite biomarkers for a plant species can be valuable for predicting agronomic performance under various environmental conditions (Sulpice et al., 2009, 2010; Steinfath et al., 2010).

Two major techniques commonly used for metabolite profiling are: (1) mass spectrometry (MS) and (2) nuclear magnetic resonance (NMR) spectroscopy. Specific methods include gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), and NMR. For instance, Skogerson et al. (2009) used these approaches to develop metabolite profiles for different wines as biomarkers of wine sensory properties. Such wine biomarkers can serve as cost-effective alternatives to traditional sensory panels and may also have applications in regulatory contexts, such as detecting product adulteration.

When evaluating biomarkers, there is typically a trade-off between the breadth of metabolic coverage and the quality of metabolite data. In other words, targeted analysis of a single metabolite or a specific metabolite class often yields higher-quality data than untargeted analyses that cover multiple chemical classes.

Chapter Summary

DNA markers are readily detectable DNA sequences whose inheritance can be tracked across generations. The growing availability of sequence data from large-scale genome sequencing projects, combined with advances in next-generation sequencing (NGS), marker systems, and bioinformatics tools, has significantly reduced the cost of marker discovery and application.

A marker linked to a gene controlling a trait of interest can serve as a “tag” for that gene, making it valuable for both foreground and background selection in backcrossing programs. Additionally, genetic fingerprinting using DNA markers is widely applied in breeding programs for endangered species and commercially important crops to verify genetic identity and ensure authenticity.

Beyond DNA markers, other types of molecular markers include RNA-based markers, protein-based markers, and metabolite-based markers.

References

Syvänen, A. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat. Rev. Genet. 2:930-942. https://doi.org/10.1038/35103535

Skogerson, K., Runnebaum, R., Wohlgemuth, G., Ropp, J.D., Heymann, H., and Fiehn, O. 2009. Comparison of gas chromatography-coupled time-of-flight mass spectrometry and H nuclear magnetic resonance spectroscopy metabolite identification in white wines from a sensory study investigating wine body. J. Agric. Food Chem. 57: 6899-6907.

Steinfath, M., Strehmel, N., Peters, R., Schauer, N., Groth, D., Hummel, J., Steup, M., Selbig, J., Kopka, J., Geigenberger, P., Dongen, J., T., V. 2010. Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. Plant Biotechnol. J. 8:900-911.

Sulpice, R., Trenkamp, S., Steinfath, M., Usadel, B., Gibon, Y., Witucka-Wall, H., Pyl, E., Tschoep, H., Steinhauser, M.C., Guenther, M., Hoehne, M., Rohwer, J.M., Altmann, T., Fernie, A.R., and Stitt, M. 2010. Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis Accessions. The Plant Cell. 22: 2872–2893

Tarpley, L., Duran, A.L., Kebrom, T.H., and Sumner, L.W. 2005. Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. BMC Plant Biol. 5: 8

CHAPTER 6.

RECOMBINANT DNA TECHNOLOGY

Walter P. Suza; Faizo Kasule; and Madan K. Bhattacharyya

Introduction

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

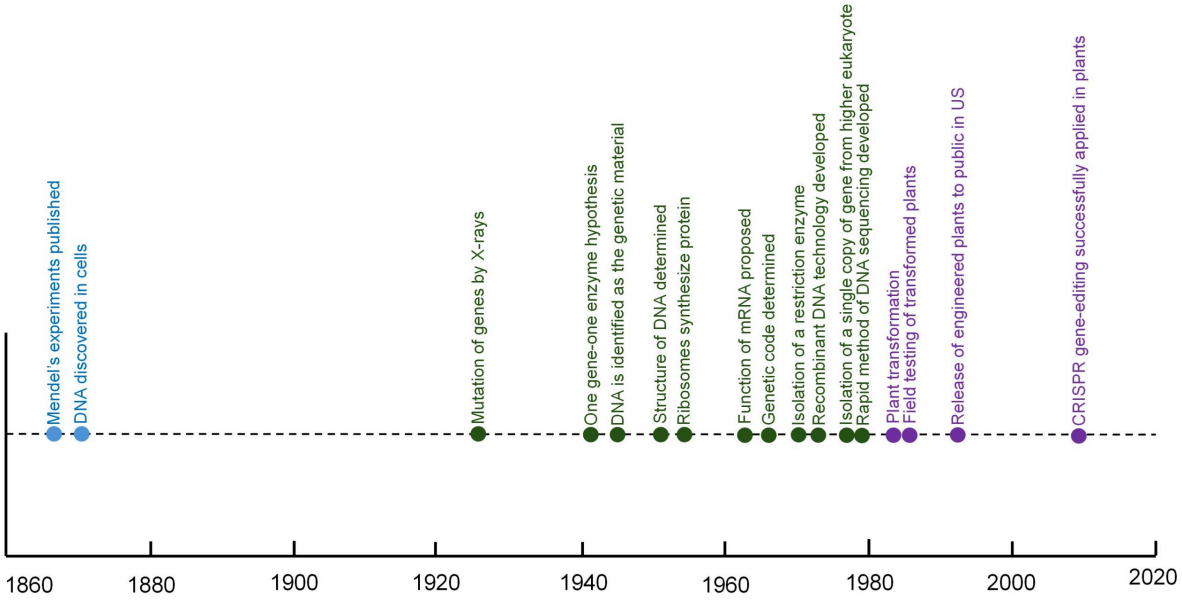


Figure 1 Events related to the development and application of recombinant DNA technology in plant breeding.

rDNA technology has resulted in breakthroughs in crop biotechnology to ensure protection from insects and weeds. The power of rDNA technology comes from our ability to study and modify gene function by manipulating genes and transforming them into cells of plants and animals. Several molecular biology tools are used to arrive at this, including DNA isolation and analysis, molecular cloning, gene expression quantification, gene copy number determination, transformation of the appropriate host for replication,

or transfer into crop plants, and analysis of transgenic plants. An introduction to rDNA procedures, such as analyzing or combining DNA fragments from one or several organisms, including the introduction of the rDNA molecule into a cell for its replication, or integration into the target cell's genome, was discussed in *Plant Breeding Methods*.

Outcomes

The 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 humans to produce various proteins for pharmaceutical purposes. The production of GM plants became possible after Bob Fraley and others succeeded in using *Agrobacterium tumefaciens* to transform plant cells with rDNA 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.

The following points summarize the outcomes of rDNA technology in agriculture.

- From 1996-2007, 1.7 billion acres were cultivated with GM crops in 12 developing and 11 developed countries.
- In 2007, world acreage for GM crops was soybean 57%, maize 13%, and canola 5%. The developing countries accounted for 43% of the world's acreage of GM crops. U.S. acreage for GM crops was soybean 90%, cotton 85%, and maize 50%.
- More than 1 billion humans have consumed foods derived from rDNA.
- More than 90% of farmers who plant GM crops are small, resource-poor farmers in developing countries.
- GM crops have contributed about $17-18 billion US dollars to the economies of developed and developing countries.

The preceding information was adapted from Vasil, (2008b). Some current statistics on the adoption of genetically engineered crops in the U.S. can be found here:

DNA as starting material

Preparation of Plant DNA

Samples must be immediately frozen to prevent degradation due to enhanced cellular metabolism. It is important to remember that for recombinant DNA procedures to work, a pure DNA sample must be obtained. The challenge is that plant 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, usually contain different compositions of metabolites, for example, proteins, lipids, and carbohydrates. These compounds must be separated from the DNA during isolation. To achieve this, scientists use different molecules' chemical and physical properties 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 nonpolar solutions. Therefore, the final step in many DNA purification protocols involves alcohol precipitation. Other compounds, for example, proteins, can also be easily separated from DNA by altering the salt concentration in the extraction buffer.

DNA isolation methods from plant tissues are becoming more robust and simpler as new technologies emerge. Despite the versatility of the latest DNA isolation technologies, one must understand the science behind various procedures in DNA isolation. Consider the protocol in the following article by Richards et al. (1994), describing methods of preparation of genomic DNA from plant tissue:

Article: Preparation of Genomic DNA from Plant Tissue

As you review the article, note how various compounds and enzymes are incorporated at different steps of the procedures.

Learning Activity

In the above article by Richards et al. (1994) in *Current Protocols in Molecular Biology* , test your understanding by considering the function of each of the following:

1. ionic and non-ionic detergents
2. protease
3. cesium chloride
4. isopropanol

Determine the function of each item by reviewing the linked article (Preparation of Genomic DNA from Plant Tissue)

To date, DNA isolation is done routinely through new inventions (referred to as kits) by different private companies. Look at the examples of commercially available plant DNA isolation kits.

- DNeasy Plant Mini Kit
- ChargeSwitch gDNA Plant Kit
- Extract-N-Amp Plant Kits

Determination of DNA Concentration

The simplest way to determine the concentration of DNA is by using a UV (ultraviolet light) spectrophotometer. Each UV spectrophotometer operates slightly differently. Ensure you know which cuvettes to use and how to clean them. Quartz cuvettes are often used in a spectrophotometer if DNA is analyzed at 260 nm. The nitrogenous bases in DNA (and RNA) absorb UV light at about 260 nm wavelength (abbreviated as A260). As a rule, obtaining an A260 value of 1.0 with pure double-stranded DNA means that the concentration of the DNA is 50 µg/ml. To determine the DNA concentration in a sample, use the following simple formula.

Unknown concentration (µg/ml) = (50 µg/ml) x (measured A260 x **dilution factor**)

The **dilution factor** is obtained based on the following:

The DNA you have isolated is for important subsequent steps such as cloning and sequencing. Therefore, you do not want to waste all your valuable samples on concentration determination. Instead, scientists often combine an aliquot of their sample with a larger volume of an appropriate buffer for concentration analysis on a UV spectrophotometer.

For example, a 5 µl aliquot of the original DNA sample mixed with 995 µl of analysis buffer will result in a final volume of 1 ml of DNA-buffer solution. The dilution factor in this case is 200 (1000 µl/5 µl). Such that the following formula can determine the concentration of DNA:

Unknown concentration (µg/ml) = (50 µg/ml) x (measured A260 x 200)

The purity of the DNA can be determined by spectrophotometry because cellular compounds such as proteins and carbohydrates also absorb UV at different wavelengths. Proteins absorb UV at 280 nm, and carbohydrates at 230 nm. Thus, the ratios of the absorbance at 260 nm/ absorbance at 280 nm or 230 nm are estimates of the purity of the DNA sample. For good quality DNA, the ratios should be between 1.65 and 1.85 for A260/A280 and close to 2 for the A260/A230.

Digestion of Plant DNA with Restriction Endonucleases

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

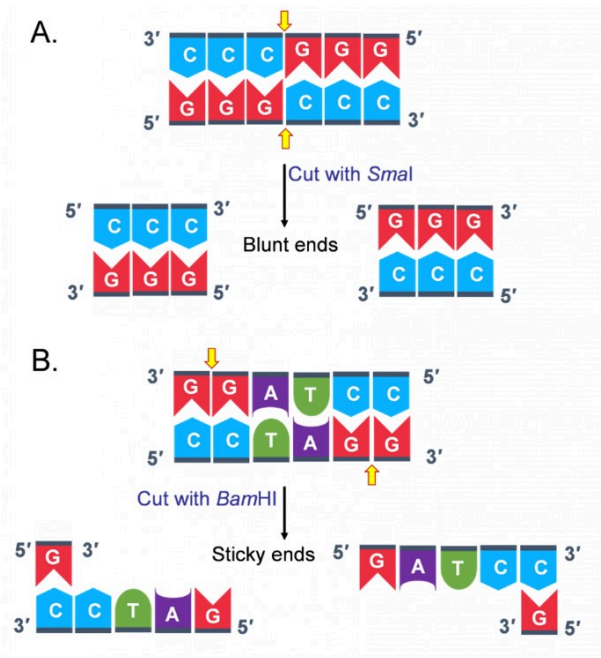


Figure 2 An example of how restriction enzymes cut DNA. (A) Treating the DNA with SmaI results in fragments with blunt ends. (B) Whereas treatment with BamHI produces fragments with "cohesive" or "sticky" ends. *Image by Walter Suza*.

More than 500 restriction enzymes have been identified and can be purchased commercially. One may wish to digest the DNA they have isolated to test for the presence of a SNP marker (as discussed in *Plant Breeding Methods*) or for Southern blot analysis (below). Thus, one may ask how often a restriction enzyme cuts within a genomic sequence. It is not possible to give an exact answer to this question. However, let us assume there are four bases on any strand of DNA. This means the probability of detecting an A (adenine) at a particular location is 1/4. 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 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 know the genome size 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; thus, frequency may differ for a given DNA. Also, methylation of specific bases in genomic DNA can prevent cleavage at some sites. The subject of DNA methylation was covered in Chapter 2.

Watch: **How To Set Up A Restriction Enzyme Digest**

Separation of Digested Genomic DNA on Agarose Gel

The only physical features of nucleic acid fragments routinely used for their characterization are their size and nucleotide sequence. The 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, an interwoven network of agarose chains, interferes with the movement of DNA through the gel (Figure 3). Pore size, which affects the rate of movement of DNA fragments of a given size, depends on agarose concentration. The gel is submerged in an electrolyte solution; the sample is loaded into wells on one end, and current is applied to facilitate the movement of DNA fragments (Figure 3). Since DNA is negatively charged, it will migrate in the electrical field. Fragments separate according to size. The distance of the migration each time is proportional to 1/log MW. Separation of DNA fragments of 100 to 50,000 nucleotides is routine, though not all on the same gel. Larger and chromosome-sized DNA can also be separated by a modified method called **pulsed field electrophoresis**:

Watch: Agarose gel electrophoresis

Watch: Loading and Cutting a Gel

View: **Pulse-Field Gel Electrophoresis**

Polyacrylamide gels can be used for high resolution and separating very small fragments (<200 bp). The principles are the same, but the sieving properties allow the separation of much smaller fragments. Polyacrylamide also enables the resolution of DNA fragments that differ by a single base, and it is used widely for DNA sequencing. Following gel electrophoresis, DNA can be visualized by staining with ethidium bromide or other DNA stains—for example, **SYBR Safe DNA Gel Stain**.

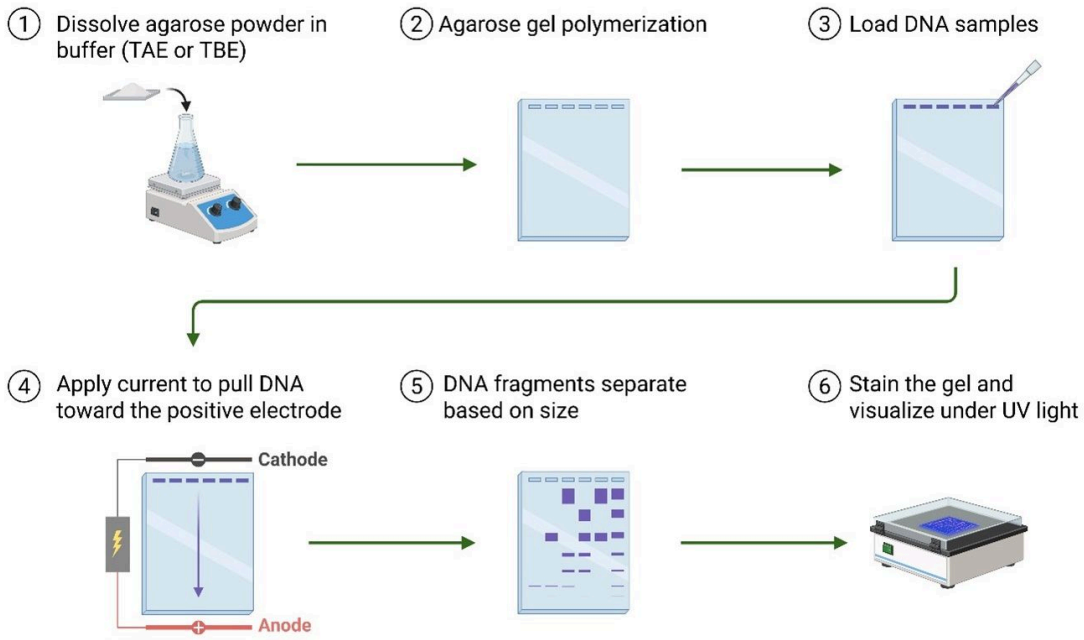


Figure 3 Steps of Agarose gel electrophoresis. Image modified by Faizo Kasule using the template in BioRender. *Adapted from Lee et al. (2012)*.

Learning Activity

Based on the example given above, close to 600,000 fragments can be obtained from cutting maize genomic DNA with a restriction enzyme. If you were to analyze the digested maize genomic DNA by agarose gel electrophoresis, what kind of banding pattern would you observe?

A. Bands

B. Smear

- Agarose gel electrophoresis is not sensitive enough to resolve thousands of fragments of similar size. Thus, although gel electrophoresis works well for analyzing samples with relatively few different fragments, it is not sensitive enough to detect the presence of an individual fragment in a complex mixture like the one produced from the digestion of plant genomic DNA.

C. Ladder

Detection of DNA by Southern Blotting and Hybridization

A key property of duplex DNA is the ability of strands to separate (denature) and then to reform a duplex (renature) spontaneously. This DNA feature is essential for its biological function and is integral to many of the techniques used in rDNA studies. Stable duplex formation in a solution requires a minimum number of complementary base pairs, which depends on the reaction conditions, such as the solution's temperature and salt concentration. When the new duplex is formed between DNA strands from different

sources, for example, DNA cloned on a plasmid with genomic DNA from different organisms, the reaction is called **hybridization**.

A more sensitive method for analyzing complex mixtures or unknown samples involves hybridizing nucleic acids. The procedure for DNA detection on a blot is called **Southern blot hybridization**, after the name of the individual who developed it (Figure 4). As mentioned earlier, a band on a gel represents many fragments of the same length. For a 1 kb fragment, the lower end of ethidium bromide sensitivity is 10–100 ng (0.15 pmol) or 4 x 10^{10} molecules. In contrast, sensitivity is 1 pg or 4 x 105 molecules for hybridization. Also, a band on a gel may represent many different sequences that are approximately the same size. Hybridization also allows the detection of a single species in a complex mixture, since DNA will only hybridize to its complement.

The stringency applied to the process is important in designing a hybridization experiment. **Hybridization stringency** can be thought of as the degree to which hybridization conditions allow imperfectly paired or mismatched duplexes to form. The stringency can be determined by all factors that affect duplex stability (Table 1)**.** At very high stringency, only perfectly paired duplexes will form. As stringency is decreased, small amounts of mismatch will be allowed. A lower stringency is useful for hybridizing a related but not identical sequence in another organism.

Table 1 *Effect of various factors on the hybridization of a nucleic acid duplex*

An increase in	Effect on hybridization rate	Effect on duplex stability	Effect on hybridization stringency
DNA concentration	increases	none	none
Temperature	increases	decreases	increases
Ionic strength	increases	increases	decreases
Fragment length	increases	increases (small)	decreases (small)
pH (high values)	decreases	decreases	increases

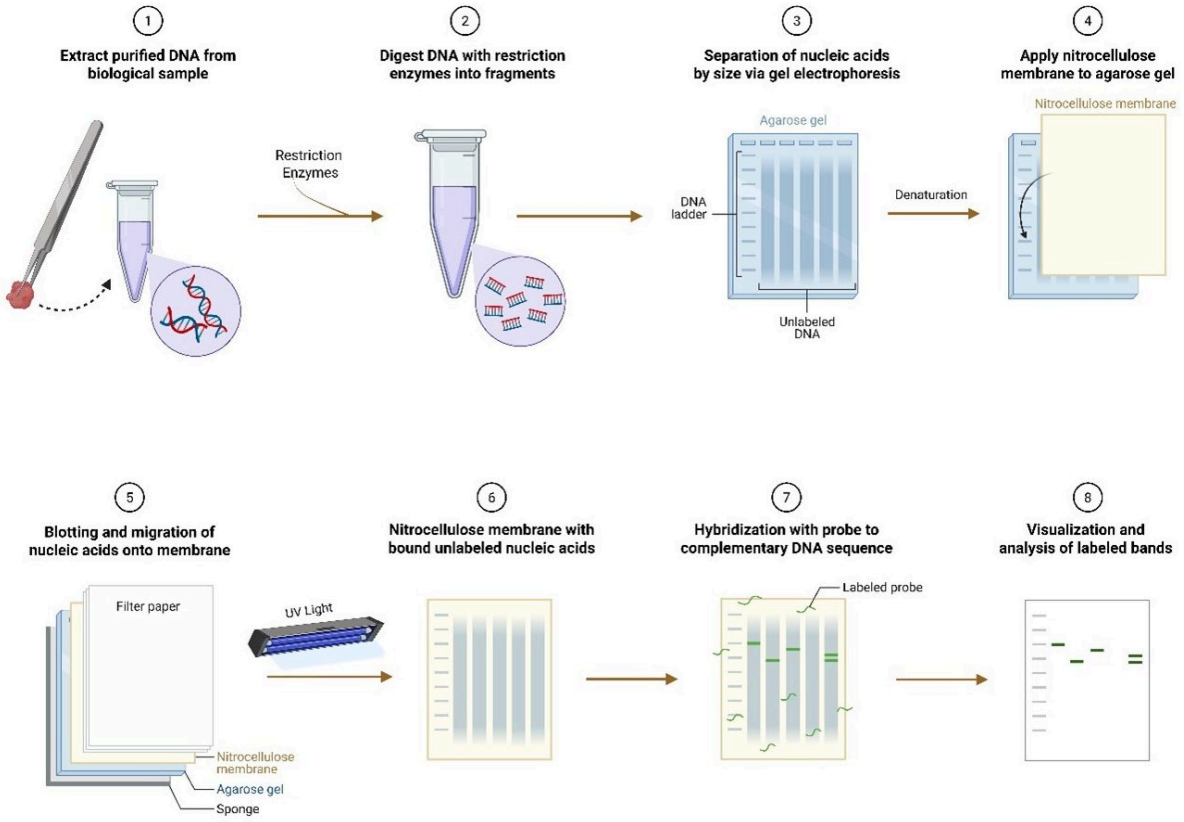


Figure 4 Basic steps of Southern blot analysis, a laboratory method used to study DNA to determine the identity, size, and abundance of specific DNA sequences. *Image modified by Faizo Kasule utilizing the template of Joseph Meacham in BioRender.*

Read this classical paper to learn more about the application of Southern blot analysis.

DNA Sequencing

Sequencing is the determination of the order of the nucleotides on a DNA molecule. A significant milestone in plant biology was reached when the genome of *Arabidopsis thaliana* was published. (The Arabidopsis Genome Initiative, 2000). Thereafter, the scientific community sought to sequence the genomes of several crop plants used for feed and food. Examples of some plants whose genomes have been targeted for sequencing are listed in *Molecular Plant Breeding*.

Principles of Sequencing

Sanger's Dideoxy DNA-Sequencing Procedure

Fred Sanger developed the Sanger DNA-sequencing procedure (dideoxy DNA sequencing) in the 1970s.

The method uses enzymatic reactions to incorporate specific terminators of DNA chain elongation called 2',3'-dideoxynucleoside triphosphates (ddNTPs). The ddNTP molecules can be incorporated into the growing DNA chain through their 5' triphosphate groups. However, because they lack a hydroxyl (OH) group on the 3'-C of the sugar moiety (Figure 5), they cannot form a phosphodiester bond with deoxynucleotide triphosphates (dNTPs) during the sequencing reaction, resulting in termination of DNA chain elongation.

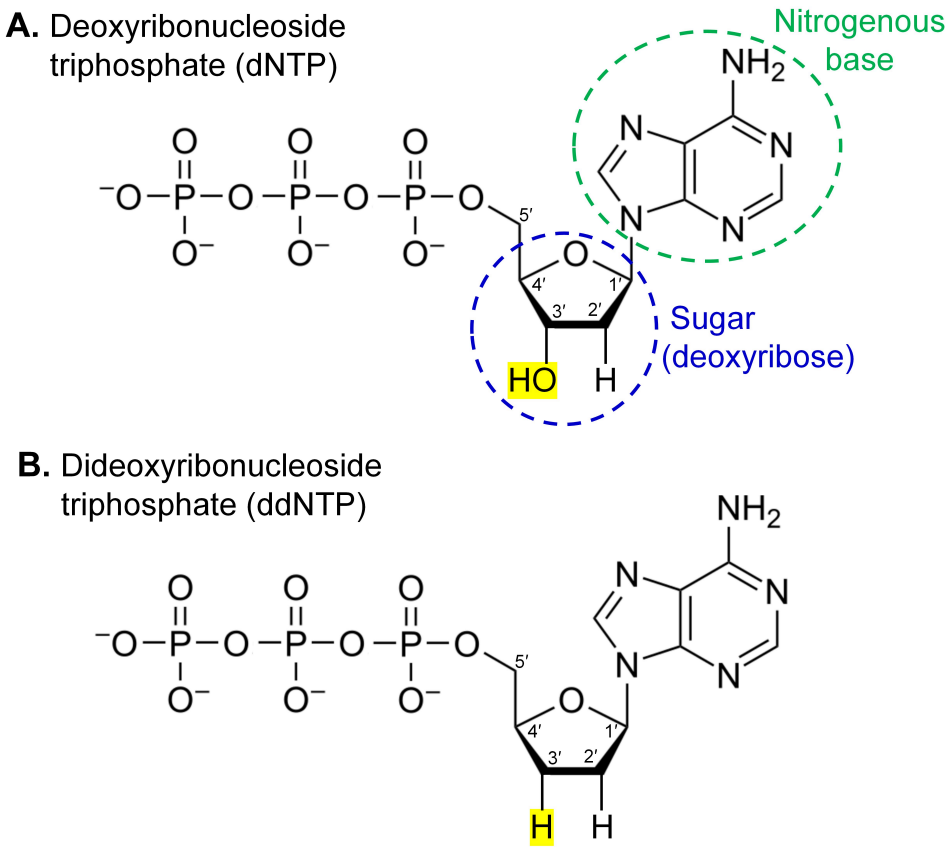
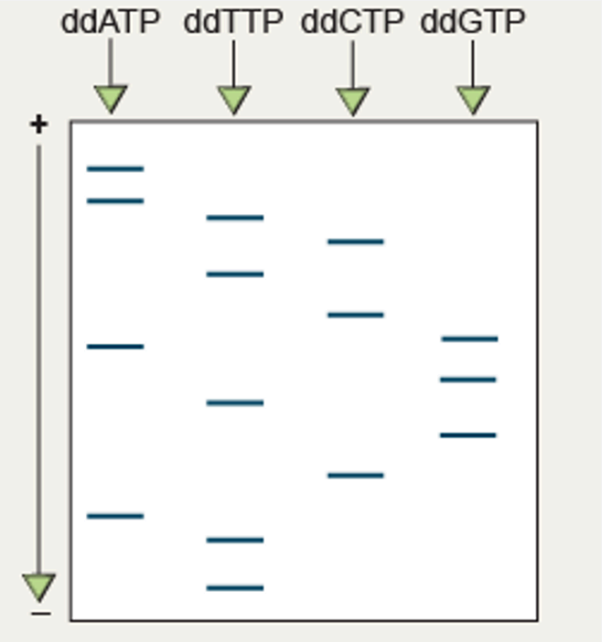


Figure 5 A dideoxyribonucleoside lacks an OH group at the 3'-C of the sugar moiety.

Watch: Video describing the Sanger DNA sequencing

Learning Activity

The figure below shows gel electrophoresis sequencing results based on the dideoxy method. Examine the banding pattern and describe the order of the nucleotides of the original template strand.



Answer

Click to reveal

- The shortest band (the one that traveled the farthest) is scored first. In this case, the shortest band corresponds to the lane loaded with products from a reaction containing ddTTP. Therefore, the sequence of the original strand is: 5' TTACGTGAGCTCTAA 3'.

Maxam & Gilbert DNA-Sequencing Procedure

Allan Maxam and Walter Gilbert developed the Maxam and Gilbert DNA-sequencing procedure in 1977. The procedure is based on the chemical degradation of DNA chains. In this procedure (Figure 6), a segment of DNA is labeled at one end with a radioactive label (^{32}P ATP). A solution containing the labeled DNA is distributed into four different tubes. A chemical that destroys one or two of the four bases (G, A+G, C, C+T) in the DNA is added to each tube. The addition of a chemical called piperidine to the DNA treated with base-modifying chemicals results in cleavage of the strand at the position of the modified base. It is important to control the reaction conditions such that only a few sites are cleaved per DNA molecule. The length of the cleaved fragments depends on the distance between the modified base and the labeled end of the DNA segment. For example, if there are C residues 4, 7, and 10 bases away from the labeled end, the fragments from cleavage at the modified C residue will be 3, 6, and 9 bases in length. The cleaved products of each of the four reactions (G, A+G, C, and C+T) will be evaluated side by side

on an acrylamide gel to separate them based on size. The gel is assessed by autoradiography, and the banding pattern on film is scored to determine the DNA sequence.

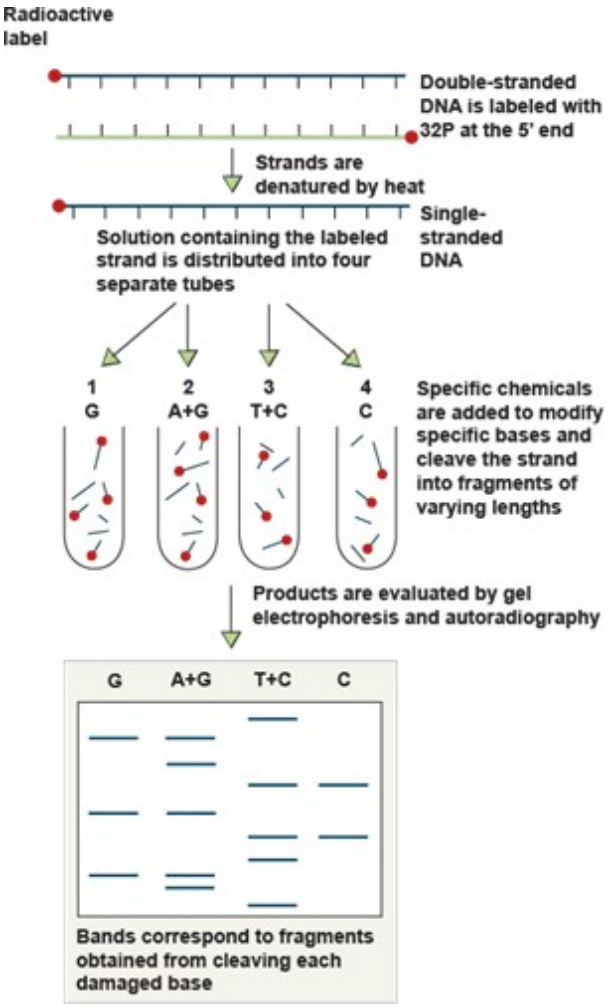


Figure 6 The Maxam and Gilbert DNA-sequencing procedure.

Watch: Maxam Gilbert Sequencing Method Video

Read: A new method for sequencing DNA

Learning Activity

Determine the sequence of the analyzed strand and its complement based on the gel electrophoresis result in Figure 6.

Answer

Click to reveal

- 5' TAGTCGCAGT 3'

Next Generation Sequencing

Next-generation (nextgen) sequencing (NGS) is a high-throughput sequencing method combining parallel processes to produce thousands or millions of sequence data points at once. One main difference of next-generation sequencing from the conventional sequencing technologies described above is that no cloning of DNA fragments in *E. coli* is required for next-generation sequencing. The NGS is divided into (i) second-generation and (ii) third-generation sequencing. The primary difference between the two is whether a single molecule or multiple molecules are used as the template for sequencing. In the second generation, sequencing PCR products representing thousands of copies of a DNA molecule are used as a template. In the third-generation sequencing, a single DNA molecule is sequenced using a single DNA polymerase molecule. Examples of the second-generation nextgen sequencing technologies are:

Pyrosequencing or 454 Sequencing

The **454 sequencing** uses a procedure known as sequencing by synthesis, or pyrosequencing. A video describing pyrosequencing can be viewed here.

Illumina sequencing

The Illumina system uses a terminator-based method to detect single bases as they are incorporated into a growing DNA strand. A video describing Illumina sequencing can be viewed here.

SOLiD Sequencing

The SOLiD system is based on a technique of oligonucleotide ligation and detection. A video with more information about the system can be viewed here.

Ion Torrent Sequencing

Ion Torrent Sequencing, or amplicon sequencing, directly translates DNA bases into digital information on a semiconductor chip. A video with more details on the system can be viewed here.

Examples of third generation sequencing are:

Single Molecule Real Time (SMRT) Sequencing

Pacific Biosciences developed the SMRT system and uses a single DNA polymerase molecule-based approach for real-time sequencing of single DNA molecules. A video with more information about the system can be viewed here:

Oxford Nanopore

In Nanopore Technologies, protein-nanopores are applied to determine the sequence of a DNA molecule. In Oxford Nanopore Technologies, a specifically designed nanopore is inserted into a membrane created by a synthetic polymer. This membrane has very high electrical resistance. Applying a potential across the membrane generates an ionic current through the nanopore. A single DNA molecule passing through the nanopore causes characteristic disruptions in the current, detected as signals and specific to the DNA building blocks, i.e., deoxynucleotides (Figure 5). Each nucleotide produces a unique signal, which calls the bases in real-time. Thus, we gather the sequence of a molecule as it travels through a nanopore in real-time. This video describes Nanopore technology.

Helicos High Speed Gene Sequencing

Illumina described earlier that the Helicos sequencing technology also uses a terminator-based method to detect single bases incorporated into a growing DNA strand. The main difference between Helicos with the Illumina system is that in Helicos, only a single molecule is sequenced, whereas in Illumina, a cluster of copies of a DNA molecule is sequenced. There are also other differences. Watch this video and compare the differences with the Illumina sequencing technology:

Sequence Alignment and Assembling Aspects of Next-gen Sequencing

In general, next-gen second-generation technologies result in a larger number of reads that are shorter than those produced using capillary electrophoresis. Therefore, next-gen sequencing requires more robust algorithms to assemble the large quantity of data generated. Third-generation next-gen sequencing, the SMRT system, has alleviated this problem by producing more than 1,000 bases long sequences.

Several academic and private institutions provide core services for next-generation sequencing, for example, the MGH Next Generation Sequencing Core and Beckman Coulter Genomics.

> Review Fig. 2 in Sequencing crop genomes: approaches
> and applications to learn how the short reads are assembled in a complex genome with repetitive sequences.

Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a method by which millions of copies of a DNA fragment are produced in a test tube in minutes or a few hours. The basic steps in the PCR reaction were discussed in *Plant Breeding Methods*. The method was invented by Kary Mullis in 1985 and exploits certain features of DNA **replication**. Refer to **Crop Genetics** and Chapter 1 for information about DNA replication **.**

The basic materials for conducting PCR using DNA were listed in *Plant Breeding Methods*. To perform a PCR reaction, small amounts of the DNA template are added to special PCR tubes (with thin walls for efficient heat exchange) that contain the following components:

1. Reaction buffer required to maintain optimum pH
2. Forward and reverse oligonucleotide primers are required to recognize the two ends of the target fragment
3. Taq polymerase that maintains activity at very high temperatures for DNA synthesis
4. dNTPs (A, C, G, T), building blocks for the new DNA strand synthesis
5. Magnesium chloride as a cofactor for the activity of Taq polymerase

The PCR reaction is performed in a thermocycler (Figure 7). A thermocycler contains a heat block for which a computer program controls the temperature. It is important to test all the variables in a PCR protocol for determining the best PCR conditions and concentrations of reagents. This is called PCR optimization. Once a PCR reaction is optimized, you may scale up your protocol to include larger volumes and number of samples to be analyzed, e.g., high-throughput PCR assays for screening a population. When hundreds of individuals are screened to search for markers of interest, technicians prepare large batches of “master mix” that contain all the correct concentrations of polymerase, buffer, dNTP, and MgCl2. A large batch of primer solution is also prepared. When thousands of samples are screened, robotics approaches are commonly used.

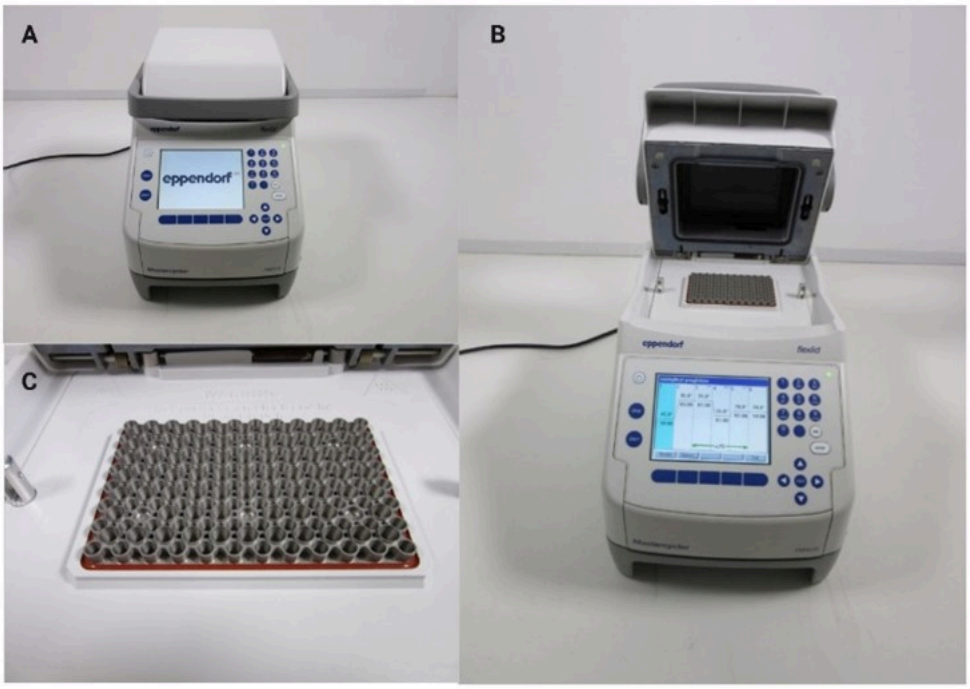


Figure 7 The thermal cycler is used to amplify DNA by PCR in a laboratory. A) Thermal cycler when the lid is closed, B) with the lid open, and C) internal chamber where Sample vials are loaded. *Image created by Faizo Kasule, Picture source: MARSHALL SCIENTIFIC*.

Below are examples of PCR automation systems:

- How to Optimize Polymerase Chain Reaction Setup workflows
- Thermo Scientific Matrix Platemate Plus
- High-precision automated PCR setup
- Automated Pipetting Systems

Reverse-transcription and PCR

Recall in Chapter 1 that you learned about RNA synthesis from DNA through **transcription**. Transcription is an important step in gene expression. The mRNA produced can be isolated and "copied" back to DNA by **reverse transcription** (Figure 8). The first step of reverse transcription mimics a strategy used by retroviruses (e.g., HIV) with RNA genomes. As part of their gene-transmission package, retroviruses also contain an enzyme called RNA-dependent DNA Polymerase, 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 complementary DNA (cDNA) strand. 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.

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

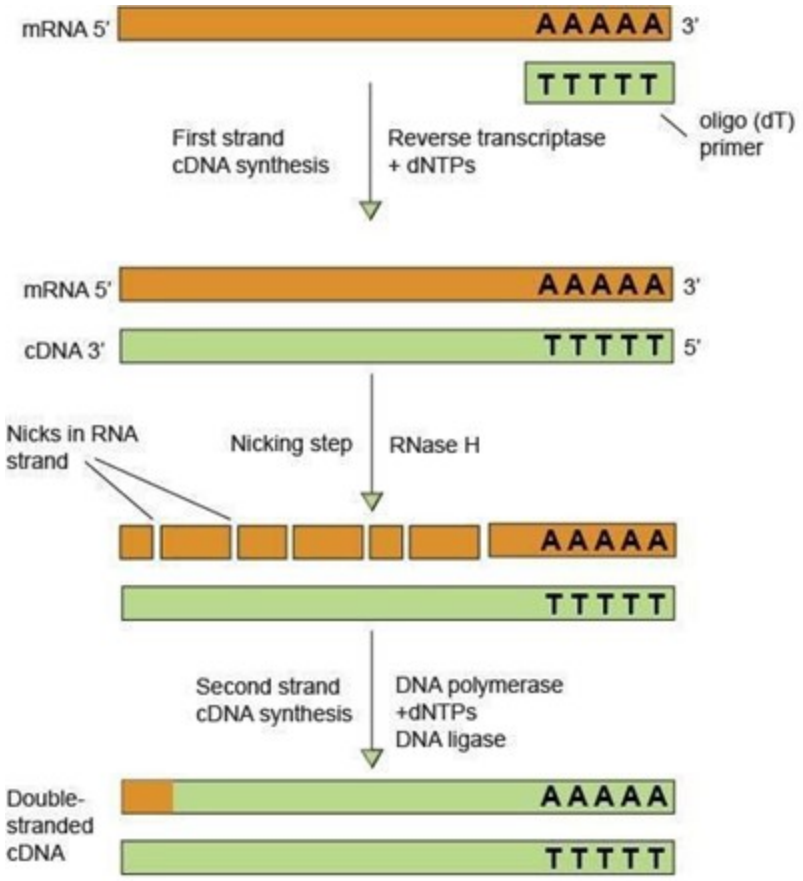


Figure 8 Steps in converting mRNA to double-stranded cDNA by reverse transcription in a test tube. The population of mRNA isolated from plant tissues is combined with oligo(dT) primers that anneal to the mRNA's poly(A) tail to initiate the cDNA's reverse transcription from the mRNA template with dNTPs. The result is hybrid molecules (mRNA-DNA). Treatment with RNase H causes mRNA degradation, 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 premixed 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, the double-stranded cDNA can be amplified by PCR for various purposes, for example, gene cloning.

Reverse Transcription Quantitative PCR (RT-qPCR)

Conventional PCR with end-point detection is sufficient for certain laboratory assays, such as confirming

the presence or absence of a transgene. However, in many cases, a researcher may wish to quantify how strongly the transgene is expressed across different plant tissues.

Unlike standard PCR, which only reveals whether amplification has occurred after the final cycle, quantitative PCR (qPCR), or real-time PCR, monitors the accumulation of PCR products during each amplification cycle. Fluorescent dyes or probes enable real-time detection of double-stranded DNA as it is synthesized. For gene expression analysis, RNA must first be reverse transcribed into cDNA (Figure 8). Using primers specific to the sequence of interest, this cDNA is then amplified by qPCR, allowing quantification of transcript levels. The primary advantage of qPCR over conventional PCR is its ability to measure transcript abundance with high sensitivity and accuracy across a broad dynamic range, rather than simply detecting the presence or absence of a target sequence.

Read: Introduction to Quantitative PCR (qPCR) Gene Expression Analysis

Read: The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments

Watch:

A. One-step vs. two-step RT-PCR
B. Real-time PCR Applications
C. Baselines in Real-Time PCR
D. How to Normalize cDNA Concentrations
E. How to Analyze Real-time PCR Data
F. Real-Time PCR Thresholds and Where to Place Them

Application of PCR in Molecular Mapping

The application of PCR to track alleles that confer partial resistance to the soybean aphid was introduced in *Plant Breeding Methods*. Using PCR, researchers have been mapping DNA markers (e.g., SNPs, SSR, CAPS, etc.) linked tightly to a genomic region that controls a trait of interest. Care must be taken to design oligonucleotide primers that specifically bind only to the target sites for amplification.

Limitations of PCR

Like all other biochemical processes, DNA synthesis by PCR is not perfect; 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, which 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 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 even a minute droplet or aerosol from a pipette tip will contain many amplifiable molecules. It is always essential to run negative controls, which will reveal the presence of contaminating DNA in your PCR assays.

Gene Cloning

Use of PCR in Cloning Genes

Once a particular gene's sequence is known, PCR makes it possible to isolate that gene from any DNA or cDNA sample. Recall in Chapter 1**,** you learned that at the genomic level, a gene is made up of regulatory sequences, coding, and non-noncoding 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. Alternatively, isolating the gene from cDNA would be the strategy if only the coding sequence is required. The cDNA must be 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.

Cloning Vector Definition and Requirements

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

In cloning genes by PCR, restriction enzyme sites (Figure 9) are added at the 5'-end of the primers to facilitate cloning of PCR fragments. A few additional nucleotides (~6 nts) are added at the 5'-end of the restriction sites to enable restriction digestion of PCR products before cloning in a plasmid vector. Alternatively, PCR products may be cloned directly into a **T-vector without restriction digestions** in *E. coli*. A

fter cloning into *E. coli*, the fragment is analyzed by sequencing and then sub-cloned into suitable vectors for expression studies.

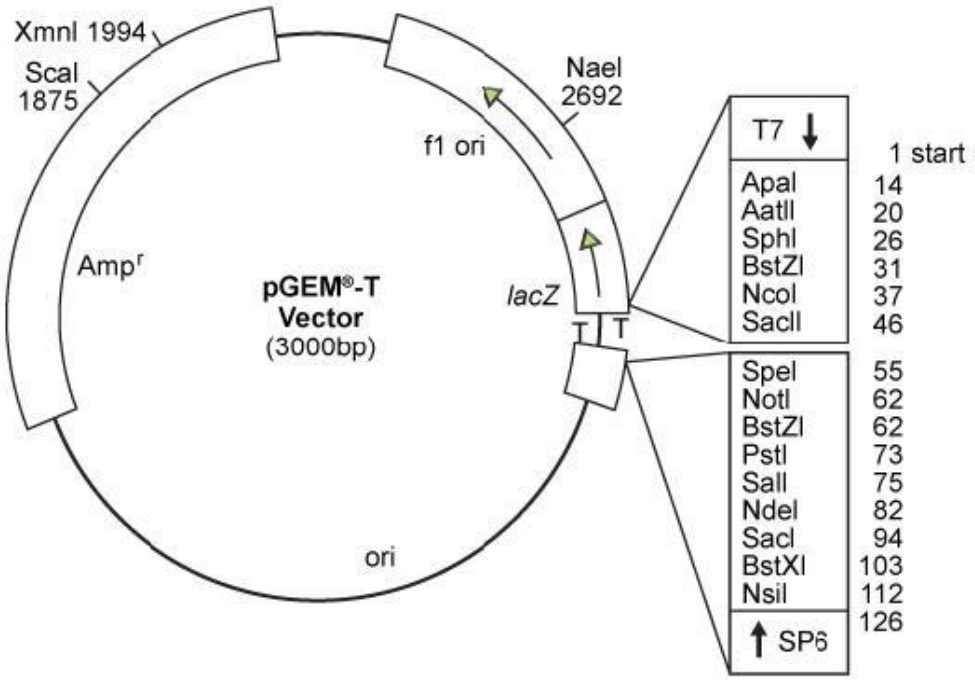


Figure 9 A map of the pGEM®-T vector available from Promega is an example of a plasmid.

Ligase Enzyme and Gene Cloning

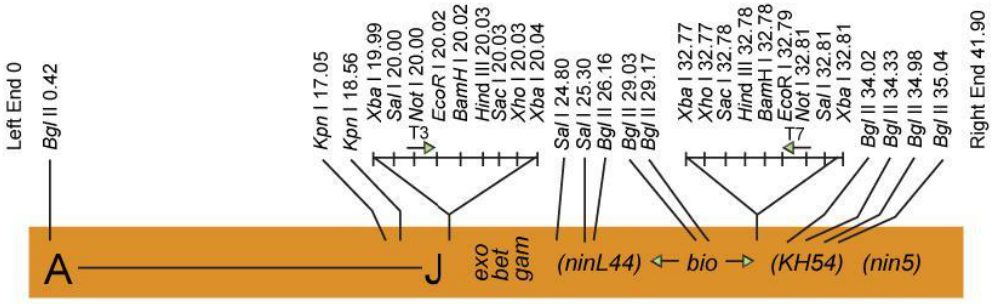
Cutting and joining vector and DNA fragments from different origins results in rDNA. Recall that restriction endonucleases are used to cut DNA. To join DNA molecules together, an enzyme called DNA ligase is used. The enzyme DNA ligase joins restriction fragments by forming new phosphodiester bonds. The 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 °C) and heat (37–42 °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 include 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

One vector type is a **plasmid** (Figure 10), an autonomously replicating extra-chromosomal circular DNA

faithfully passed on to progeny. Plasmids are double-stranded circular DNA ranging from about 1 kb to 200 kb. The most useful plasmids for cloning are 2–10 kb because smaller plasmids are easier to manipulate and usually produce a higher copy number when grown in bacterial host cells.



Lambda DASH II Multiple Cloning Site Regions
[sequences shown: 19.99-20.04kb (top sequence) and 32.77-32.81kb (bottom sequence)]

Xba I Sol I Not I T3 Promoter EcoR I BamH I Hind III Sac I Xho I Xba I
TCTAGAGAGCTGTCGACGCGGCCGCGGAATTAACCCTCACTAAAGGGAACGAATTCGGATCCAAGCTTGAGCTCCTCGAGAAATTCTCTAGA

Xba I Xho I Soc I Hind III BamH I EcoR I T7 Promoter Not I Sol I Xba I
TCTAGAGAATTTCTCGAGGAGCTCAAGCTTGGATCCGAATTCTTCGCCCTATAGTGAGTCGTATTACGCGGCCGCGTCGACAGCTCTCTAGA

Figure 10 *A map of the Lambda DASH II vector available from Agilent Technologies*.

Generally, not more than 10 kb is cloned into plasmids. A vector called bacteriophage lambda (e.g., Figure 10) is used for highly efficient cloning of large DNA fragments. Lambda is a linear duplex of about 48.5 kb with 12 bp cohesive ends, which permit the cloned DNA to circularize when introduced into the cell. Large chromosomal DNA fragments close to 23 kb are stable when introduced into a lambda phage vector.

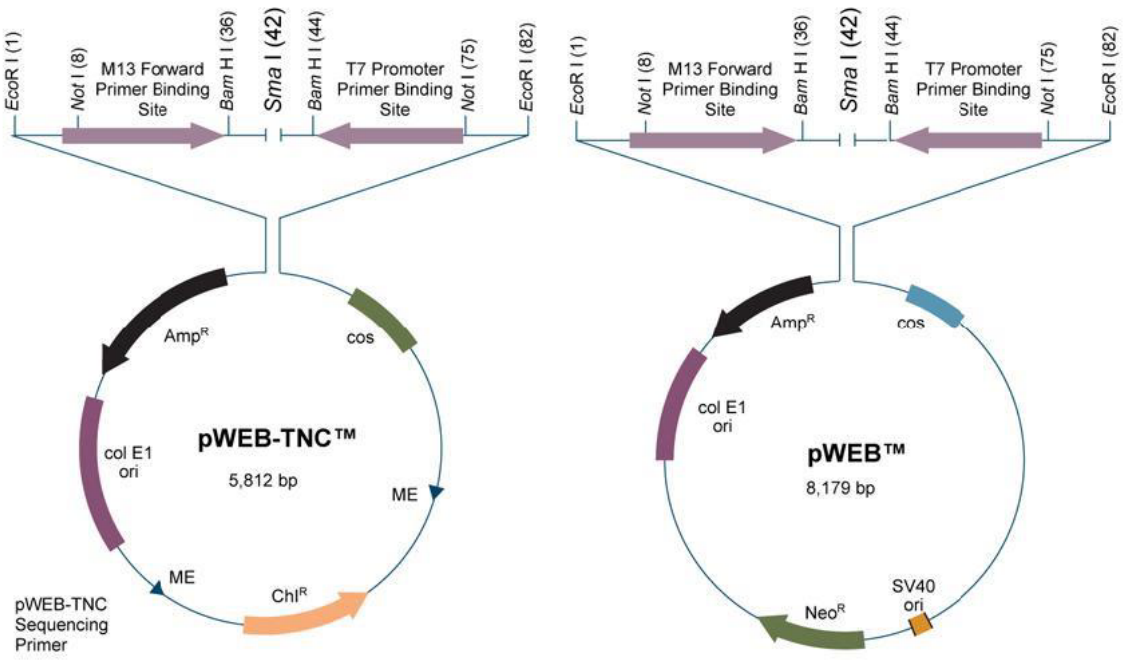


Figure 11 Vector maps for pWEB-TNCTM and pWEBTM available from Epicentre are examples of cosmids.

A third vector type is called a **cosmid** (Figure 11). A cosmid is a plasmid with cohesive ends like lambda, which allows cloning much larger fragments (35–45 kb). A cosmid replicates like a plasmid after being introduced into the host bacterial cell. The large size of cosmids makes them a challenge to work with.

In the examples above (Figures 9 to 11), the T7, SP6, and T3 primer-binding sites can be used to validate the presence of the cloned insert (from several base pairs to about 10 kb) by PCR.

Ti Plasmid Vector and Plant Transformation

Crown galls are plant tumors that develop at the site of infection by certain species of the soil bacterium *Agrobacterium*. Rather than entering plant cells, *Agrobacterium* transfers a DNA segment known as T-DNA from its circular, extra-chromosomal tumor-inducing (Ti) plasmid into the host cell genome. Ti plasmids are retained in *Agrobacterium* because part of their T-DNA encodes genes for unusual amino acids that the bacterium can utilize. The T-DNA also carries genes that disrupt the host plant's hormone balance, triggering abnormal cell growth and tumor formation. Scientists have harnessed *Agrobacterium*'s ability to stably integrate T-DNA into plant genomes as a tool for introducing recombinant DNA into plant cells. They removed the tumor-inducing genes from the T-DNA and engineered the plasmid to replicate in both *E. coli* and *Agrobacterium*. Replication in *E. coli* is advantageous because it allows verification of the cloned gene and produces large amounts of construct DNA for sequencing and later transfers into *Agrobacterium*. Plant transformation using *Agrobacterium* was introduced in *Plant Breeding Methods* and will be explored further in Chapter 7.

Gene Libraries in Plasmid, BAC, and YAC Vectors

There are two types of gene libraries, genomic and cDNA libraries. A **genomic library** is a collection of cellular clones containing sequences of an organism's entire genome. Each clone contains only a single genomic DNA fragment. Genomic libraries are made by digesting genomic DNA with a restriction enzyme, ligating every fragment to a vector, and transforming such rDNA molecules into host cells. A **cDNA library** is a collection of cloned cDNA molecules generated by reverse transcription of mRNA extracted from plant tissues that are actively expressing the genes of interest (Figure 8). Unlike genomic libraries, cDNA libraries lack promoters and introns, since they are derived directly from processed mRNA rather than genomic DNA.

Now, let's ask the following question. How many clones should be present in a library? To answer this question, one must know the genome size of the plant species they are working with. Let us take the example of maize. As mentioned earlier, the maize genome is estimated to contain 2,500 Mb (2,500,000 kb). If you were to reliably digest a 2,500,000 kb fragment from a maize cell into a series of 250 kb restriction fragments, you would generate 2,500,000/250 = 10,000 genomic fragments. To clone such large fragments, you would need to use a special cloning vector, e.g., a bacterial artificial chromosome (BAC) vector, **pBeloBAC11,** if you clone each of the 10,000 genomic fragments into this vector and transform them into *E. coli*, you will obtain a perfect genomic library. The number of clones in this "perfect" library represents what is known as a **genomic equivalent**. Therefore, to compute the number of clones that constitute one genomic equivalent, divide the length of the genome by the average size of the inserts carried by the cloning vector.

In addition to BAC vectors, one can use a yeast artificial chromosome (YAC) vector to clone large plant DNA fragments in individual yeast cells. YAC clones are much bigger and can accommodate large DNA inserts up to 10 Mb.

BAC and YAC vectors are instrumental in the construction of genomic libraries because they can accommodate large DNA inserts, reducing the number of clones required for a genome equivalent library. The fewer the number of clones, the less effort is placed in searching the clones for the sequence of interest. For example, you have seen above that the genomic equivalent of BAC clones with 250 kb inserts contains 10,000 clones. However, a genomic equivalent of plasmid clones with 10 kb inserts would contain 250,000 clones.

Now the main question is what the size of a library should be. As outlined above, we cannot create a perfect library due to sampling issues. The larger the library, the better the chance we can identify our DNA sequence of interest. Usually, the possibility or probability of finding a random DNA fragment in a three-genomic equivalent library is 0.95, while in the six-equivalent library it is 0.99. In general, a six-genomic equivalent library is constructed to start with. If one fails to obtain the fragment of interest, they may continue to construct a larger genomic library in the same or an alternative vector.

Protein expression in bacteria

Developing protein expression systems is an ongoing effort in both industrial and academic research settings. The development of expression vectors has allowed scientists to produce plant proteins in bacterial cells. The most popular expression systems are bacteria, for example, *E. coli*.

The process begins with ligating a DNA fragment containing the coding region of the gene of interest to a plasmid with promoter and other regulatory sequences recognized by the transcriptional machinery of the host cell. Most expression vectors provide options to fuse the gene of interest to a "protein tag" that can be used as a "handle" to isolate the desired protein from a pool of thousands of *E. coli* proteins. For example, the pGEX vector system from Sigma provides the option to link the protein of interest to glutathione-S-transferase (GST) with a short amino acid sequence that can be cleaved by thrombin. The GST-fused recombinant protein is separated from the solution using Glutathione Sepharose, which binds GST. Subsequently, the GST portion on the recombinant protein is removed by digestion with thrombin.

A concern about using bacterial cells as hosts to produce plant proteins is that some plant proteins require additional modifications that can only occur inside a plant cell. One of these modifications is the ordered insertion of sugars into some proteins that occur in specialized compartments of the plant cell. Lack of proper modifications of proteins can affect their folding and function. To overcome this problem, scientists use yeast as hosts to express plant proteins that require additional modifications related to their function in a eukaryotic organism. Alternatively, the protein may be expressed in an insect cell line through baculovirus-mediated infection. However, at times, even yeast or insect cells may fail to provide a suitable environment for expressing proteins of plant origin. In such cases, the best solution is to express the protein in plant cells.

Gene Cloning by Chromosome Walking, Chromosome Jumping

After creating a library of thousands of genomic fragments, the next step is isolating interest gene from the library. Suppose you know by genetic information that a particular trait you are interested in is in a chromosomal region. If you do not know the gene responsible for the trait, how would you identify the gene in a region that may be more than 2 Mb? The answer is, you would start by selecting genomic clones that lie closer to the region of interest and then "walk" through the region by identifying overlapping clones from your genomic library.

> This article by Stein and co-workers is a good example of the use of chromosome walking in wheat. This article used the chromosome walking approach to clone a gene that controls vernalization and mapping of a region that controls chromosome pairing in wheat.

Recall that a genomic library is prepared from fragments obtained from digesting genomic DNA. The gene responsible for your trait will be in your library's thousands of DNA inserts. To identify clones close to the gene of interest, you will first need to find DNA markers linked to the trait you are interested in. This is done by searching for markers that co-segregate with the trait in a plant population. In other words, recombination between your gene (trait) and markers is not seen.

The next step is to screen the genomic library with the co-segregating markers (probes) in isolating clones that hybridize to those markers. The ends of the clones that hybridize to co-segregating markers are used as probes to identify overlapping clones. The procedure is repeated to systematically "walk" toward the gene of interest, one clone at a time. You might think the procedure is lengthy, and you are right. For example, it may take more than a month to walk >100 kb. For that reason, chromosome jumping (Figure 12) is used to cover a region quickly. The chromosome jumping method is described in this article.

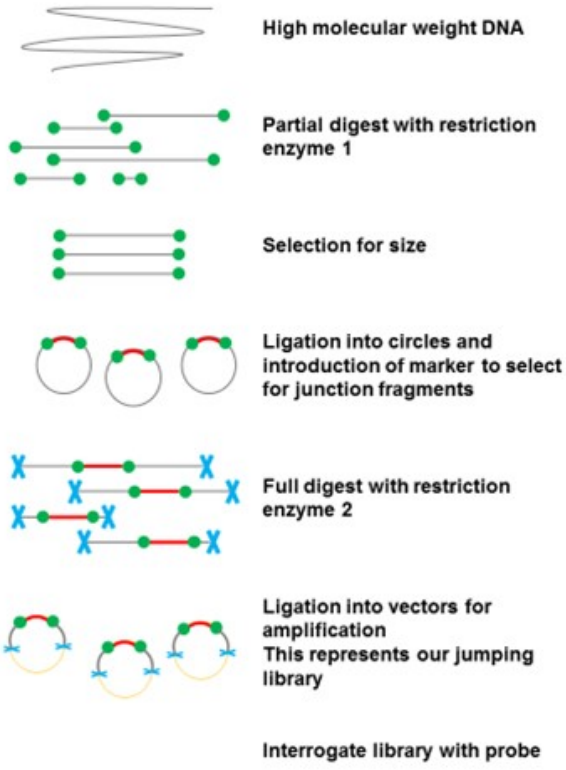


Figure 12 Method for creating a chromosome jumping library. *Image source: AnaSACohen, CC BY-SA 3.0, via Wikimedia Commons.*

It is worth noting that with the advent of next-generation (nextgen) sequencing (see below) technologies and genome sequences for many crop plants, one may avoid arduous traditional map-based cloning or chromosomal jumping experiments in cloning genes based on their map positions. A new approach,

based on next-generation sequencing of the bulk of recessive mutants following bulk-segregant analyses, has been demonstrated for identifying genes in Arabidopsis.

> Read: Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations.

> Read: SHOREmap: simultaneous mapping and mutation identification by deep sequencing.

Application of rDNA Technology in Crop Improvement Research

Recombinant DNA technology has contributed significantly to the development of agricultural biotechnology, from plants that express bacterial toxins to protect themselves from insects to herbicide-tolerant plants that effectively manage weeds.

Gene Silencing as a Tool to Uncover Gene Function in Biochemical Pathways

Recall from Chapter 2 that RNA interference (RNAi) can be accomplished more efficiently by expressing a portion of the target gene, usually around 400 bp, that has been engineered as an inverted repeat in transgenic crop plants. Following transcription of this engineered gene, the RNA molecules form a hairpin structure that is then cleaved into small fragments of double-stranded RNA, which interfere with the accumulation and function of the endogenous mRNA molecules of the target gene. Below is an example of how scientists applied RNAi to investigate gene function in the jasmonic acid (JA) pathway (Figure 13).

JA is a key plant hormone that regulates development and mediates responses to stresses such as leaf injury. Research by Paul Staswick and his team at the University of Nebraska–Lincoln revealed that amino acid conjugates of JA, particularly jasmonoyl-L-isoleucine (JA-Ile), serve as crucial signaling molecules (Staswick et al., 2002; Staswick & Tiryaki, 2004; Suza & Staswick, 2008). In Arabidopsis, the *JASMONATE RESISTANT 1* (*JAR1*) gene encodes an enzyme that conjugates JA with isoleucine (Figure 13A). Building on this knowledge, Suza et al. (2010) applied RNAi (Figure 13B) to identify and characterize a tomato gene predicted to encode an enzyme that catalyzes JA-Ile formation in response to leaf wounding (Figure 14), analogous to the role of JAR1 in Arabidopsis.

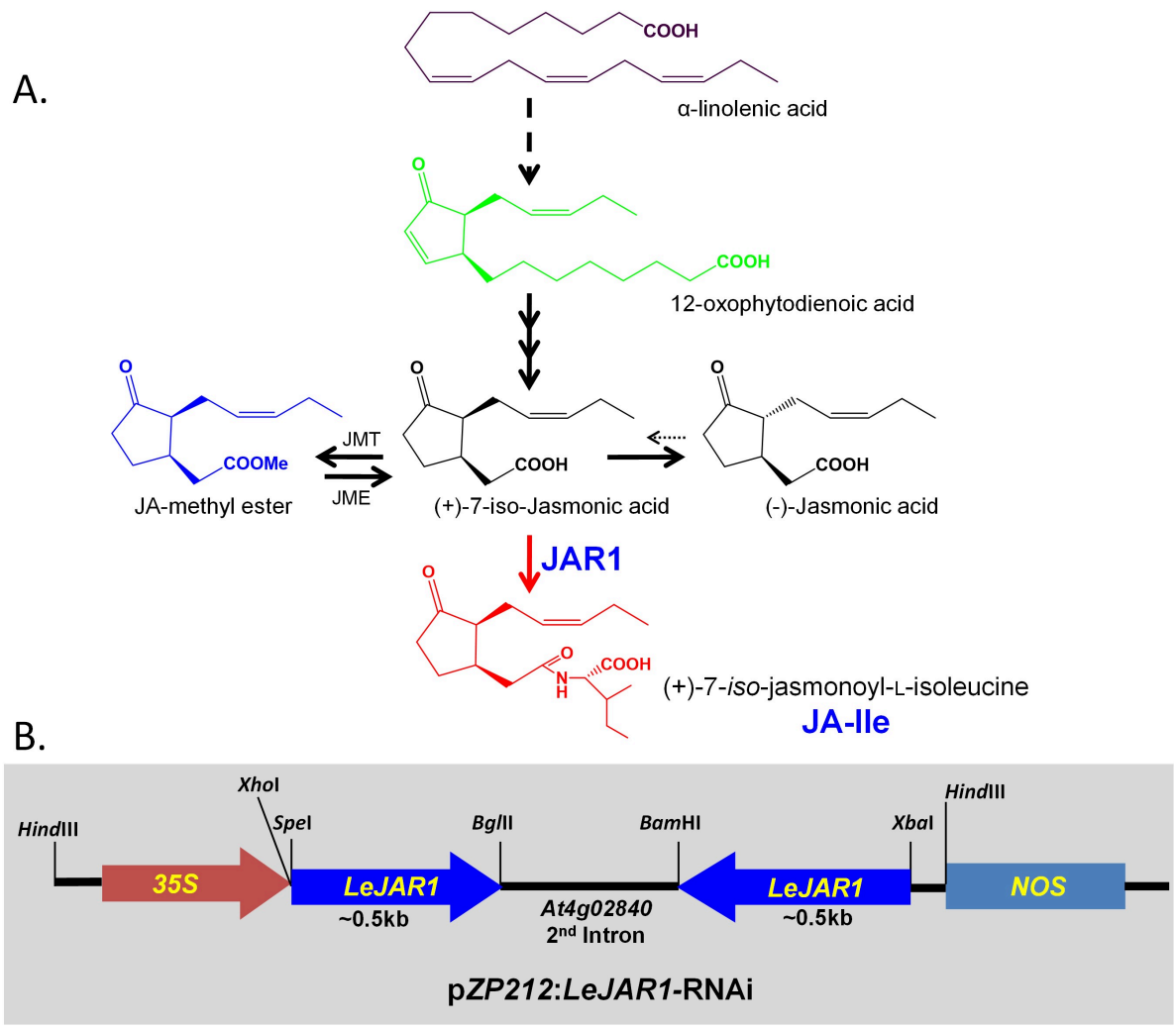


Figure 13 (A) The jasmonate biosynthetic pathway leading to the bioactive jasmonic acid-amino acid conjugate, jasmonoyl-L-isoleucine (JA-Ile). Dashed and multiple arrows represent multiple reaction steps in the pathway. (B) RNAi construct for silencing tomato *JAR1* (*LeJAR1*). Short sequences of *LeJAR1* are cloned into a vector in both forward and reverse orientations, separated by an intron from the *At4g02840* gene. The intron acts as a spacer, enabling the formation of a "hairpin" RNA structure that effectively triggers RNAi. The Cauliflower mosaic virus 35S promoter drives expression of the RNAi cassette, and transcription is terminated by the nopaline synthase (*NOS*) terminator sequence.

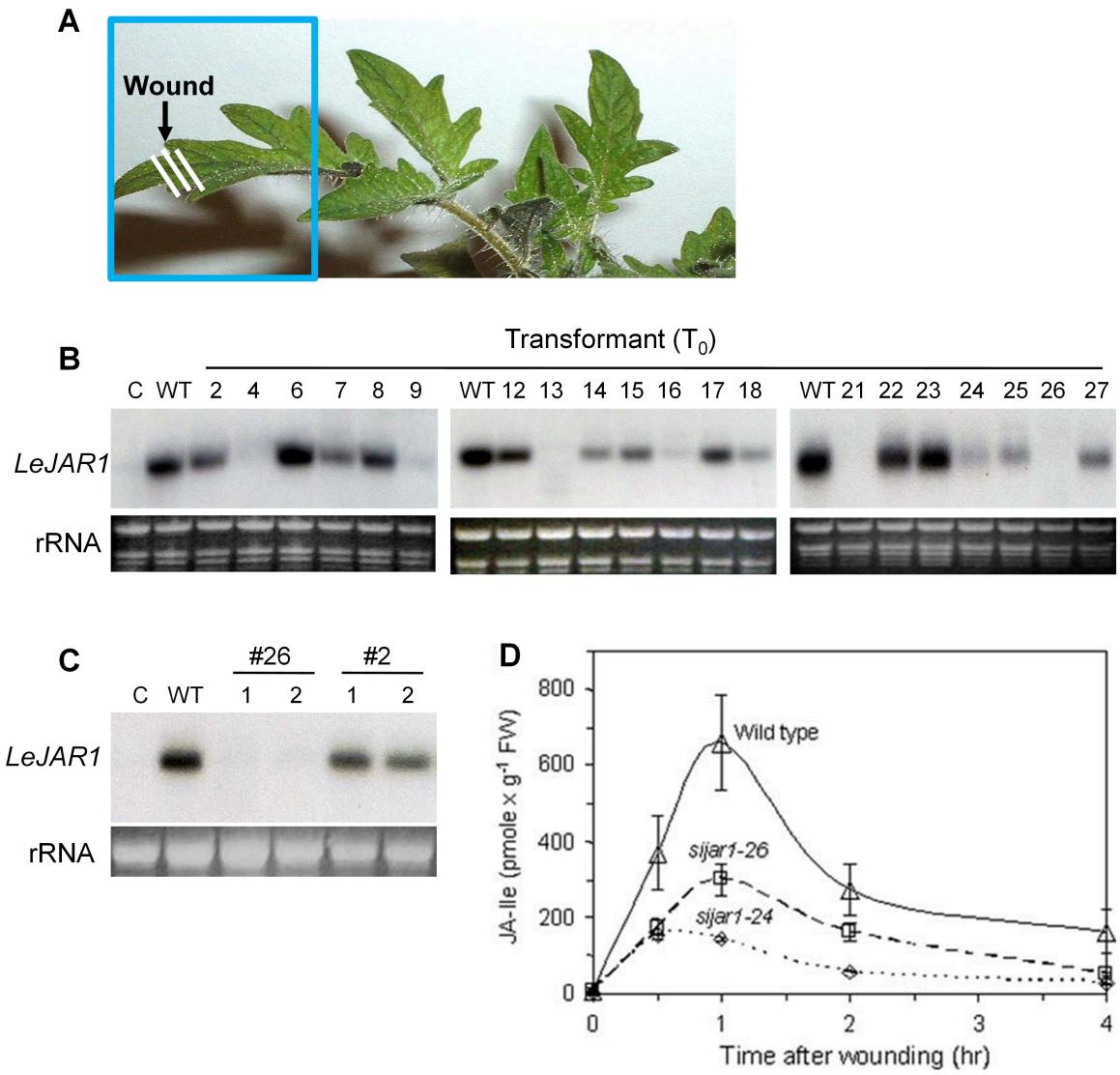


Figure 14 Transcript analysis of plants independently transformed with *pZP212:LeJAR1*-RNAi cassette **(A)** Mechanical wounding was applied as shown **(B)** Leaves of T_0 plants were wounded for 3 hr, and RNA from the wounded tip was analyzed for *LeJAR1* mRNA. Samples from unwounded WT leaves (Lane marked C) and wounded WT leaves (Lane marked WT) were loaded as negative and positive controls respectively **(C)** Leaves of T_1 progeny (two plants, labeled 1 or 2) of primary transformants number 26 and 2 were wounded for 3 hr and RNA from the wounded tip analyzed for *Le* JAR1 mRNA **(D)** Accumulation of JA-Ile in response to wounding in silenced transformants sijar1-24 and sijar1-26 **.** *Adapted from Suza et al. (2010)*.

Gene Silencing as a Tool for Trait Improvement

Consider mutation breeding described in *Plant Breeding Methods*. One can create a mutant population for selecting desirable phenotypes. Most mutations lead to loss of gene function. If the gene is a negative regulator of a particular trait, loss of that "switch" due to a mutation will lead to expression of a specific trait. One good example is the *mlo* locus in barley. *Mlo* is a negative regulator of non-race-specific powdery mildew resistance in barley. A mutation in the *Mlo* gene led to the mlo allele, which now provides broad-spectrum resistance to all races of the powdery mildew pathogen. This is one of the success

stories of mutation breeding in barley. However, the entire process of mutation breeding is labor-intensive. We can accomplish the same objective by manipulating targeted genes in transgenic crop plants.

Decreased expression of an existing gene, such as *Mlo* in barley, could be accomplished quite effectively by introducing a copy of the gene in the **antisense** (reversed) orientation. A normal transcribed region, or a portion of it, is reversed so that the RNA molecules transcribed from the engineered gene will hybridize to the RNA molecules of the endogenous native gene, leading to the degradation of the mRNA encoded by the endogenous gene. This process is highly gene-specific and leads to downregulation of the targeted gene, even though the original gene is not mutated in any way. The downregulation of the *Mlo* gene is a result of RNAi.

RNAi provides a powerful tool for genetic study and for biotechnological crop improvement. The following are examples of the application of RNAi for Crop Improvement.

Example 1: Enhancing Carotenoid and Flavonoid Content in Tomatoes

Tomato contains lycopene and β-carotene in disproportionate amounts. Lycopene is an important antioxidant that prevents heart disease and cancer in humans. β-carotene is a precursor to vitamin A. Vitamin A is essential to humans, and failure to consume adequate vitamin A is a leading cause of health problems and infant mortality in developing countries. Efforts to genetically enhance both lycopene and β-carotene had previously been unsuccessful. However, Davuluri et al. (2005) took an approach to shut down the expression of the *DE–ETIOLATED1* (*DET1)* gene in tomato fruits. The *DET1* gene is a regulatory gene that represses several signaling pathways modulated by light. A mutation in this has been shown to increase the pigment concentration in tomato. Mutants also contain increased levels of flavonoids and carotenoids. Davuluri and co-workers fused a section of the *DET1* coding sequence in both the reverse and forward orientations to a promoter (Figure 15) that only directs the expression of the "inverted repeats" in the fruit. The inverted repeats of *DET1* resulted in the formation of dsRNA. The *DET1* dsRNA was a substrate for dicer to produce siRNAs (about 21-25 nucleotides in length). These *DET1* siRNAs are complementary to the region of *DET1* that was used to produce the inverted repeats. The binding of *DET1* siRNAs to the *DET1* mRNA transcribed from the *DET1* gene in fruits resulted in patches of *DET1* siRNA-*DET1* RNA duplex. The double-stranded patches trigger the degradation of the entire *DET1* mRNA and suppression of the *DET1* gene in fruits. The result of *DET1* silencing was enhanced production of both lycopene and β-carotene in transgenic fruits.

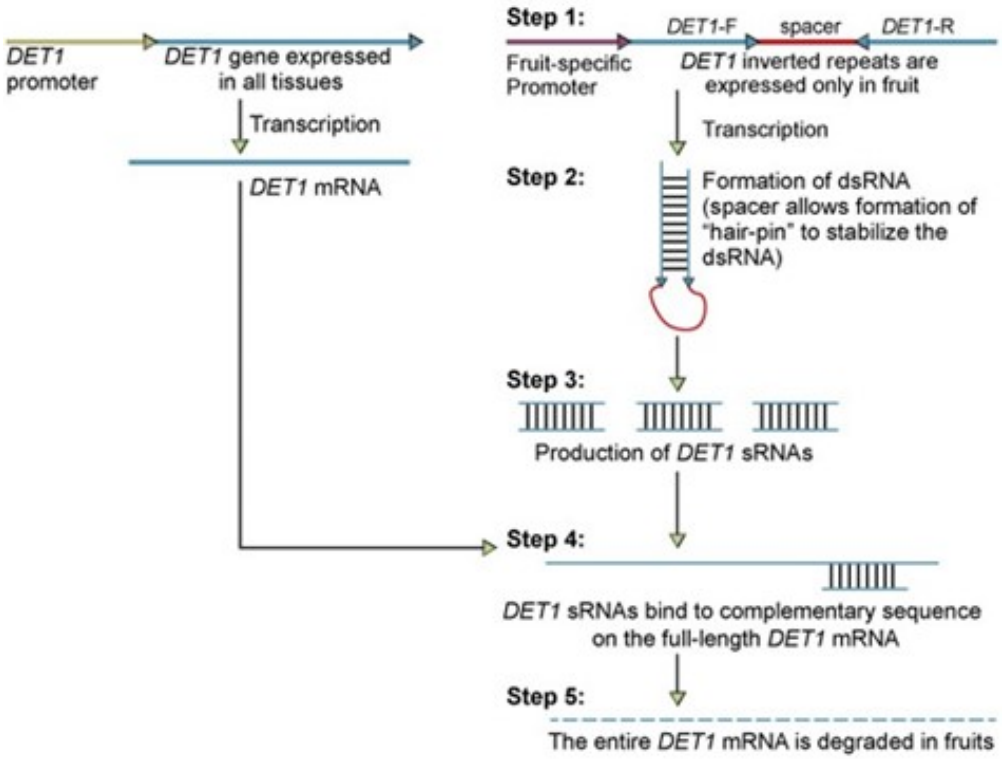


Figure 15 RNAi-mediated down-regulation of the DET1 gene in tomato fruits (Davuluri et al., 2005). The fruit-specific promoter directs the expression of the inverted repeats in fruit tissue, not leaf tissue. The spacer is a DNA sequence that provides stability and enhanced expression of the repeats. Proper expression of these inverted sequences will trigger RNAi to silence the endogenous DET1 gene.

Fruit-specific RNAi-mediated suppression of *DET1* enhances carotenoid and flavonoid content in tomatoes.

To read about this, see http://www.nature.com/nbt/journal/v23/n7/pdf/nbt1108.pdf

Example 2: Caffeine-Free Coffee

Scientists and plant breeders have sought to remove caffeine from coffee for over a century. Unfortunately, the chemical processes used currently to produce decaf coffee also destroy the flavor of the highly consumed beverage.

Caffeine is produced in coffee plant cells from a natural plant chemical called xanthosine. Caffeine biosynthesis involves three subsequent additions of methyl groups to xanthosine (Figure 16). The methylation reactions are controlled by three distinct methyltransferases, xanthosine methyltransferase (*XMT*), 7-N-methylxanthine methyltransferase (*MXMT*; theobromine synthase), and 3,7-dimethylxanthine methyltransferase (*DXMT*; caffeine synthase). Ogita and colleagues from Nara Institute of Science and Technology in Japan used RNAi to downregulate the expression of *MXMT* in coffee plants. The result

was transgenic RNAi plants with 70% less caffeine than controls (Ogita et al., 2003, 2004). The commercial application of this strategy awaits further improvement to target RNAi only to coffee beans and further to reduce caffeine content beyond the current 70% level. Other reports claim that the transgenic RNAi coffee plants show reproductive abnormalities, making it challenging to produce sufficient seed for breeding purposes (Borrell, 2012).

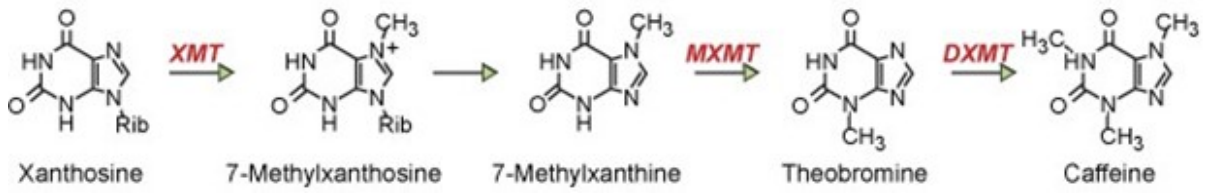


Figure 16 Proposed caffeine biosynthetic pathway in coffee plants. Key methylation steps controlled by XMT, MXMT, and DXMT are indicated. *Adapted from Ogita et al. (2004)*.

Read: Plant biotechnology: Make it a decaf

Read: Producing decaffeinated coffee plants

Read: Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties.

Example 3: Increased Resistance to Insects

Recently, it has been demonstrated that plants engineered to express double-stranded RNA (dsRNA) molecules of genes isolated from insect pests show enhanced pest resistance. For example, Baum et al. (2007) developed transgenic maize overexpressing a western corn rootworm (WCR) V-ATPase A dsRNA. The WCR V-ATPase is expressed in the midgut section, and it encodes a membrane-localized proton pump that regulates the function of numerous cellular transporters essential to the insect. Thus, WCR feeding on the transgenic plants would ingest V-ATPase dsRNA that would trigger silencing of their endogenous V-ATPase in the midgut. Consequently, the resulting transgenic maize plants overexpressing V-ATPase dsRNA were more tolerant to WCR than controls due to increased insect mortality.

Another example of the application of RNAi for insect control comes from the work of Mao et al. (2007). Mao and co-workers identified a gene called *CYP6AE14* that is highly expressed in the gut of the cotton

bollworm (CB). The expression of *CYP6AE14* is correlated with larval growth in the presence of gossypol, an insect defense compound produced by cotton, suggesting *CYP6AE14* is required for tolerance to gossypol. Cotton plants engineered to express dsRNA of *CYP6AE14* showed resistance to CB due to increased sensitivity of the pest to gossypol.

> Read: RNAi-mediated insect protection against insects.

Example 4: Increased Resistance to Nematodes

Plant parasitic nematodes cause significant crop losses worldwide. In the early 2000s, the cost of agriculture due to plant parasitic nematodes was estimated to be around US$125 billion (Chitwood, 2003). Control of nematodes relied heavily on costly and toxic chemicals. Cultural systems such as crop rotation were less effective because the nematodes can remain dormant (e.g., cyst nematode) for many years. The discovery that when short RNAs are ingested by the model worm *Caenorhabditis elegans* result in gene

silencing, inspired researchers to apply this technology to target nematodes (Figure 17). One of the targets of economic importance is the soybean cyst nematode (*Heterodera glycines*), a major pest of soybean responsible for billions of dollars each year in the US. Transgenic soybean expressing dsRNA for a gene involved in reproduction in *H. glycines* showed a significant reduction in the number of *H. glycines* eggs in root tissue (Steeves et al., 2006).

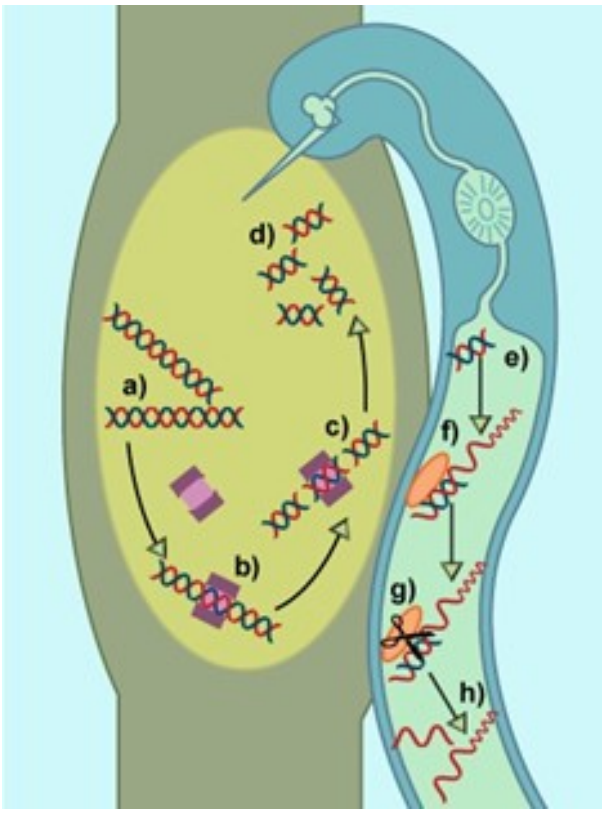


Figure 17 Double-stranded RNA produced in the plant can lead to gene silencing in plant-parasitic nematodes. A transgenic plant expressing dsRNA targeting a nematode gene is developed (a). The dsRNA is processed by the plant DICER enzyme (b) to produce small interfering (si)RNAs (c). The nematode (d) ingests the siRNAs when feeding from the plant cell. Ingested siRNAs (e) are recognized by the RISC complex, leading to their interaction with the corresponding mRNA in the nematode (f). The mRNA from the target nematode gene is cleaved (g) and degraded (h). *Adapted from (Gheysen & Vanholme, 2007)*.

Read: RNAi in plant parasitic nematodes

Read: Controlling soybean cyst nematodes by RNAi

Example 5: Engineering Male Sterility

Male sterility is an important trait in hybrid seed production. To demonstrate the usefulness of RNAi in

plant breeding, let us consider the induction of male sterility by reducing the expression of key genes involved in floral development. The maize *Zea Apetala1* (*ZAP1*) encodes a transcription factor that controls inflorescence architecture. The expression of *ZAP1* is restricted to the sterile organs of the male floret (Mena et al., 1995)—Consequently, RNAi silencing of *ZAP1* results in male sterility (Figure 18).

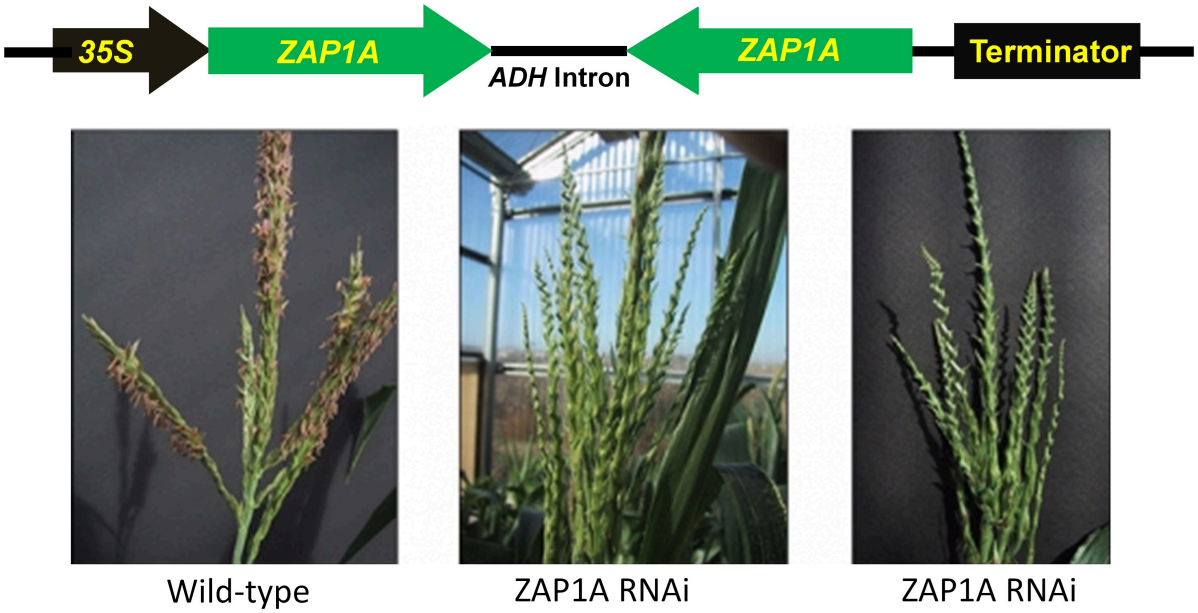


Figure 18 RNAi silencing of ZAP1. Short sequences of ZAP1 are cloned into a vector in forward (blue insert) and reverse (green insert) orientations and separated by an intron from the ADH gene. The function of the intron (spacer) is to allow formation of the ZAP1 "hairpin" structure that is conducive to triggering RNAi. The 35S promoter drives the expression of the RNAi cassette—photos by Kan Wang, Iowa State University.

Overexpression of Transgenes for Crop Improvement

Example 1: Improving Cotton Tolerance to Drought and Salt Stress

Abiotic stresses, including drought and salinity, seriously threaten agriculture worldwide. Genes controlling many steps in plant response to abiotic stress and their associated metabolic and **signaling networks** have been identified and characterized. Researchers are using this knowledge and lessons learned from plant species adapted to hostile environments to figure out ways to engineer abiotic stress tolerance.

A gene from Arabidopsis encoding a membrane-localized proton pump called AVP1 was overexpressed in several plant species, including cotton (Pasapula et al., 2011). The result was enhanced tolerance to drought and salinity stress. Also, transgenic cotton overexpressing AVP1 had 20% higher fiber yield than controls under dry conditions in the field, suggesting potential for application of this approach in other crop species.

Example 2: Improving Rice Tolerance to Salinity Stress

Sodium ion (Na^+) is toxic to plant cells when present in excessive levels. Plant cells have a mechanism to regulate their Na^+ concentration through the action of Na^+ transporters. However, the degree of tolerance to salt varies markedly among plant species. An Arabidopsis gene encoding a membrane-localized sodium (Na^+) transporter was overexpressed in rice's root cortical and epidermal cells (Plett et al., 2010). The resulting transgenic rice grown under high salt had reduced Na^+ in the shoot due to decreased root-to-shoot flux of Na^+. This genetic modification enabled the plants to gain fresh weight under salinity stress.

Read: Expression of an Arabidopsis vacuolar H+-pyrophosphatase gene (AVP1) in cotton improves drought—and salt tolerance and increases fibre yield in the field conditions.

Read: Improved Salinity Tolerance of Rice Through Cell Type-Specific Expression of *AtHKT1;1*

Other Crop Improvement Approaches

Example: Transgenic Virus-Resistant Papaya

The advent of recombinant DNA technology enabled scientists to study cross-protection by transforming plants with viral genes. Cross-protection refers to the phenomenon in which infection by a mild strain of a virus protects a plant from subsequent infection by more severe strains. This strategy has been successfully applied to control certain plant virus diseases.

Plant virus infections can devastate agriculture, as seen in papaya production in Hawaii during the early 1990s. The papaya ringspot virus (PRSV) caused systemic rotting, wilting, chlorosis, and the appearance of yellow spots. By 1997, PRSV had nearly destroyed Hawaii's papaya industry, forcing many farmers out of business. To combat PRSV, researchers built upon earlier work from Roger Beachy's laboratory (Nelson et al., 1988), which showed that transgenic plants overexpressing the tobacco mosaic virus (TMV) coat protein (CP) exhibited resistance to infection (Figure 19). Using a similar approach, Fitch et al. (1992) developed transgenic papaya plants resistant to PRSV by overexpressing the PRSV coat protein gene. This intervention ultimately rescued the papaya industry in Hawaii and demonstrated that transgenic virus resistance can be a powerful strategy for managing viral diseases in horticultural crops.

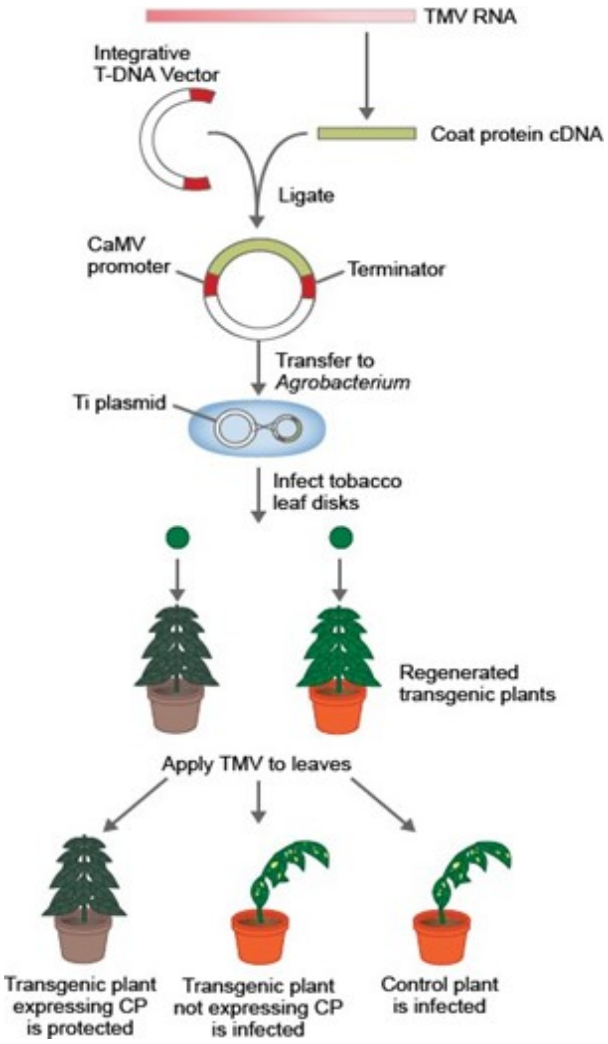


Figure 19 Protection of plants from TMV infection by transgenic expression of the TMV CP. *Adapted from Watson, J.D., Gilman, M., Witkowski, J., & Zoller, M. (1997). Recombinant DNA. New York, NY: Scientific American Books*.

Read: Virus Tolerance, Plant Performance of Transgenic Tomato Plants Expressing Coat Protein from Tobacco Mosaic Virus

Read: Virus-Resistant Papaya Plants Derived from Tissues Bombarded with the Coat Protein Gene of Papaya Ringspot Virus

Read: Transgenic Virus-Resistant Papaya: From Hope to Reality for Controlling Papaya Ringspot Virus in Hawaii

Transient Gene Expression

All examples in the previous sections involved **stable transformation** of plant cells. Stable transformation means that a plant cell has received and has stably integrated a transgene into its genome. In some cases, such as promoter analysis or to verify that a construct works properly, it isn't necessary to regenerate plants after transformation. Genes may be expressed in cell nuclei immediately or shortly after they have been treated with the appropriate DNA construct, even though new plants have not been regenerated. In this case, it is not necessary to use a selectable marker.

With **transient expression**, DNA may or may not be integrated into the genome. A reporter gene is useful for this purpose. To define the elements of a promoter required for expression of a gene, it can be dissected by creating deletions of the promoter and by mutating certain candidate nucleotides that may have a regulatory function. The modified promoters derived from either deletion or mutation are investigated for their activity by fusing them to a reported gene, the expression of which can easily be monitored. The advantage of transient expression studies is obviously that they are **much more rapid**, avoiding the many months needed for regeneration, recovery, and analysis of progeny. A disadvantage is that it may not be as accurate, since it is difficult to know how much DNA is in the nuclei of cells that are assayed. However, there are ways to standardize this by co-transforming with a second gene whose expression level does not change under the conditions being investigated. The other pitfall of transient expression study could be artifact in expression levels and patterns, significantly different from the expression patterns of the same gene in stably transformed plants. For these reasons, transient expression is often a preliminary approach, and more precise results should be obtained from stably transformed plants.

An example of a transient method of gene silencing is called **viral-induced gene silencing,** or **VIGS** for short. In the VIGS method, rather than transforming a plant, the silencing construct is incorporated into a viral genome, and the plant is then infected with the virus. The silencing DNA spreads from cell to cell in the plant along with the viral DNA or RNAs, and then the silencing portion interferes with the accumulation of the plant mRNAs corresponding to the gene construct.

Read: Virus-Induced Gene Silencing and Transient Gene Expression in Soybean (*Glycine max*) Using Bean Pod Mottle Virus Infectious Clones

Application of Next-Generation Sequencing in Crop Genetics and Breeding

Next-generation sequencing (NGS) technologies have transformed crop genetics and breeding by enabling a wide range of applications. These include the development of genomic resources, molecular marker discovery, QTL mapping, wide crosses and alien gene introgression, expression analysis, association genetics, and population biology. A more detailed discussion of this subject is provided in *Molecular Plant Breeding*.

Traditionally, map-based cloning was a lengthy and resource-intensive process, often requiring years of high-density and high-resolution mapping. With the advent of NGS, this process has been greatly simplified. Instead of constructing extensive genetic maps, researchers can now create two bulks by pooling DNA from homozygous families carrying different alleles at the target locus. This bulked segregant analysis approach, combined with NGS, dramatically reduces time and effort.

The rapid decline in sequencing costs has further accelerated genomics research. Entire genomes can now be sequenced quickly, leading to breakthroughs in functional genomics (studying gene expression patterns), metagenomics (sequencing microbial communities from specific environments), comparative genomics, molecular marker development, genome-wide association studies (GWAS), and cloning of genes underlying traits of interest.

Sequencing genomic DNA, including bacterial artificial chromosomes (BACs), reduced representation genomes (RRGs), or cDNA from reference genotypes, can generate essential genomic resources such as expressed sequence tags (ESTs), gene catalogs, and draft genome assemblies. These resources provide critical insights into genome architecture and serve as the foundation for crop genetics research.

Another powerful application of NGS lies in parental genotyping of mapping populations or wild relatives. This accelerates the development of molecular markers such as simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs). These markers enable constructing genetic maps, identifying QTLs, and tracking alien genome introgression during wide crosses. Once QTL-associated markers are identified for a trait of interest, they can be applied in marker-assisted selection (MAS) to efficiently select progeny carrying favorable alleles.

For developing functional or gene-based markers, NGS of cDNA from contrasting genotypes can identify candidate genes associated with specific traits. Expression mapping of these candidate genes and

phenotyping of segregating populations can reveal expression QTLs (eQTLs). Markers associated with eQTLs serve as highly reliable tools for MAS in crop breeding programs.

NGS also plays a major role in association genetics and population biology. Genomes or pooled PCR products from thousands of candidate genes can be sequenced across hundreds of individuals using barcoding strategies. The resulting sequence data enables the identification of SNPs and haplotypes across genes or genomes, which can then be used for association studies and population-level analyses.

The following articles describe the application of next-generation sequencing for crop improvement:

- Next-generation sequencing technologies and their implications for crop genetics and breeding
- Plant genome sequencing: applications for crop improvement
- Sequencing crop genomes: approaches and applications
- Crop genome sequencing: lessons and rationales
- Fast-forward genetics enabled by new sequencing technologies
- SHOREmap: simultaneous mapping and mutation identification by deep sequencing

Chapter Summary

Recombinant DNA technology has contributed significantly to the development of agricultural biotechnology. The transformation of cells with rDNA produces organisms called bioengineered or genetically modified organisms (GMOs) with improved traits such as pest and disease resistance, herbicide tolerance, and enhanced nutritional content. Sequencing is the determination of the order of the nucleotides on a DNA molecule. The more recent sequencing technologies are referred to as next-generation and third-generation (NGS and long-read sequencing technologies, respectively). Nextgen technologies produce many short reads compared to the earlier sequencing approaches developed by Sanger and Maxam-Gilbert. Therefore, more robust algorithms are required to assemble and analyze next-gen data (and to identify genetic variations, annotate genes, and study complex genomes). The tools for plant rDNA technology include vectors, restriction enzymes, ligation enzymes, bacterial hosts, methods to isolate and multiply nucleic acids, methods to quantify nucleic acids, and Agrobacterium as a vector to insert foreign DNA into plants, which enables stable transformation and functional analysis of genes.

References

Agilent Technologies, Inc. (2011). Lambda DASH II Undigested Vector Kit: Instruction Manual. 247201: 1–12. Available at http://www.genomics.agilent.com/files/Manual/247201.pdf

Baum, J. A., Bogaert, T., Clinton, W., Heck, G. R., Feldmann, P., Ilagan, O., Johnson, S., Plaetinck, G., Munyikwa, T., Pleau, M., Vaughn, T., & Roberts, J. (2007). Control of coleopteran insect pests through RNA interference. *Nature Biotechnology*, *25* (11), 1322–1326. https://doi.org/10.1038/nbt1359

Borrell, B. (2012). Plant biotechnology: Make it a decaf. *Nature*, *483* (7389), 264–266. https://doi.org/10.1038/483264a

Chitwood, D. J. (2003). Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture-Agricultural Research Service. *Pest Management Science*, *59* (6–7), 748–753. https://doi.org/10.1002/ps.684

Davuluri, G. R., van Tuinen, A., Fraser, P. D., Manfredonia, A., Newman, R., Burgess, D., Brummell, D. A., King, S. R., Palys, J., Uhlig, J., Bramley, P. M., Pennings, H. M. J., & Bowler, C. (2005). Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nature Biotechnology*, *23* (7), 890–895. https://doi.org/10.1038/nbt1108

Epicentre® (an Illumina® company). (2012). Available at http://www.epibio.com/item.asp?id=332 (verified 18 July 2012).

Fitch, M. M. M., Manshardt, R. M., Gonsalves, D., Slightom, J. L., & Sanford, J. C. (1992). Virus Resistant Papaya Plants Derived from Tissues Bombarded with the Coat Protein Gene of Papaya Ringspot Virus. *Bio/Technology*, *10* (11), 1466–1472. https://doi.org/10.1038/nbt1192-1466

Gheysen, G., & Vanholme, B. (2007). RNAi from plants to nematodes. *Trends in Biotechnology*, *25* (3), 89–92. https://doi.org/10.1016/j.tibtech.2007.01.007

Haemse. (2011). Single Molecule Real Time Sequencing -Pacific Biosciences [Online]. YouTube. Available at http://www.youtube.com/watch?v=v8p4ph2MAvI&feature=related

Lee, P. Y., Costumbrado, J., Hsu, C.-Y., & Kim, Y. H. (Directors). (2012, April 20). *Agarose gel electrophoresis for the separation of DNA fragments.* (Issue 62) [Video recording]. https://doi.org/10.3791/3923

Mao, Y.-B., Cai, W.-J., Wang, J.-W., Hong, G.-J., Tao, X.-Y., Wang, L.-J., Huang, Y.-P., & Chen, X.-Y. (2007). Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nature Biotechnology*, *25* (11), 1307–1313. https://doi.org/10.1038/nbt1352

Mena, M., Mandel, M. A., Lerner, D. R., Yanofsky, M. F., & Schmidt, R. J. (1995). A characterization of the MADS-box gene family in maize. *The Plant Journal*, *8* (6), 845–854.

Nelson, R. S., McCormick, S. M., Delannay, X., Dubé, P., Layton, J., Anderson, E. J., Kaniewska, M., Proksch, R. K., Horsch, R. B., Rogers, S. G., Fraley, R. T., & Beachy, R. N. (1988). Virus Tolerance, Plant Performance of Transgenic Tomato Plants Expressing Coat Protein from Tobacco Mosaic Virus. *Bio/Technology*, *6* (4), 403–409. https://doi.org/10.1038/nbt0488-403

Ogita, S., Uefuji, H., Morimoto, M., & Sano, H. (2004). Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties. *Plant Molecular Biology*, *54* (6), 931–941. https://doi.org/10.1007/s11103-004-0393-x

Ogita, S., Uefuji, H., Yamaguchi, Y., Koizumi, N., & Sano, H. (2003). Producing decaffeinated coffee plants. *Nature*, *423* (6942), 823–823. https://doi.org/10.1038/423823a

Pasapula, V., Shen, G., Kuppu, S., Paez-Valencia, J., Mendoza, M., Hou, P., Chen, J., Qiu, X., Zhu, L., Zhang, X., Auld, D., Blumwald, E., Zhang, H., Gaxiola, R., & Payton, P. (2011). Expression of an Arabidopsis vacuolar H+-pyrophosphatase gene (AVP1) in cotton improves drought- and salt-tolerance and increases fibre yield in field conditions. *Plant Biotechnology Journal*, *9* (1), 88–99. https://doi.org/10.1111/j.1467-7652.2010.00535.x

Phyla1. (2011). Illumina sequencing.mov]Online]. YouTube. Available at http://www.youtube.com/watch?v=l99aKKHcxC4&NR=1&feature=endscreen

Plett, D., Safwat, G., Gilliham, M., Skrumsager Møller, I., Roy, S., Shirley, N., Jacobs, A., Johnson, A., & Tester, M. (2010). Improved salinity tolerance of rice through cell type-specific expression of AtHKT1;1. *PloS One*, *5* (9), e12571. https://doi.org/10.1371/journal.pone.0012571

Richards, E., Reichardt, M., & Rogers, S. (1994). Preparation of genomic DNA from plant tissue. *Current Protocols in Molecular Biology*, *27* (1), 2–3.

Staswick, P. E., & Tiryaki, I. (2004). The Oxylipin Signal Jasmonic Acid Is Activated by an Enzyme That Conjugates It to Isoleucine in Arabidopsis[W]. *The Plant Cell*, *16* (8), 2117–2127. https://doi.org/10.1105/tpc.104.023549

Staswick, P. E., Tiryaki, I., & Rowe, M. L. (2002). Jasmonate Response Locus JAR1 and Several Related Arabidopsis Genes Encode Enzymes of the Firefly Luciferase Superfamily That Show Activity on Jasmonic, Salicylic, and Indole-3-Acetic Acids in an Assay for Adenylation. *The Plant Cell*, *14* (6), 1405–1415. https://doi.org/10.1105/tpc.000885

Steeves, R. M., Todd, T. C., Essig, J. S., & Trick, H. N. (2006). Transgenic soybeans expressing siRNAs

specific to a major sperm protein gene suppress Heterodera glycines reproduction. *Functional Plant Biology: FPB*, *33* (11), 991–999. https://doi.org/10.1071/FP06130

Suza, W. P., Rowe, M. L., Hamberg, M., & Staswick, P. E. (2010). A tomato enzyme synthesizes (+)-7-iso-jasmonoyl-l-isoleucine in wounded leaves. *Planta*, *231* (3), 717–728. https://doi.org/10.1007/s00425-009-1080-6

Suza, W. P., & Staswick, P. E. (2008). The role of JAR1 in Jasmonoyl-l-isoleucine production during Arabidopsis wound response. *Planta*, *227* (6), 1221–1232. https://doi.org/10.1007/s00425-008-0694-4

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408* (6814), 796–815. https://doi.org/10.1038/35048692

Vasil, I. K. (2008a). A history of plant biotechnology: From the Cell Theory of Schleiden and Schwann to biotech crops. *Plant Cell Reports*, *27* (9), 1423–1440. https://doi.org/10.1007/s00299-008-0571-4

Vasil, I. K. (2008b). History and evolution of the International Association for Plant Biotechnology (IAPB): 1963–2008. *In Vitro Cellular & Developmental Biology – Plant*, *44* (5), 365–372. https://doi.org/10.1007/s11627-008-9148-8

CHAPTER 7.

PLANT TRANSFORMATION

Faizo Kasule; Walter P. Suza; and Madan K. Bhattacharyya

Introduction

Plant transformation is cornerstone of modern biotechnology. It involves introducing foreign genes into plant cells to produce gene-edited or genetically modified plants with stable, integrated, and heritable traits. This process includes two main steps: (1) transient transformation, where foreign DNA is introduced into plant cells but not yet integrated into the genome, and (2) stable transformation, where the foreign DNA is integrated into the plant genome. Both steps are important in plant biotechnology, with the latter being essential for producing transgenic plants that pass desired traits to subsequent generations. Unlike traditional breeding methods that rely on crossing related species, plant transformation enables the transfer of a desired gene across species, genera, families, and kingdoms, which was previously impossible. In a matter of weeks, plant transformation produces millions of cells with regeneration capacity, which dramatically expanded the scope of plant genetic improvement, revolutionizing agriculture and plant science (Figure 1).

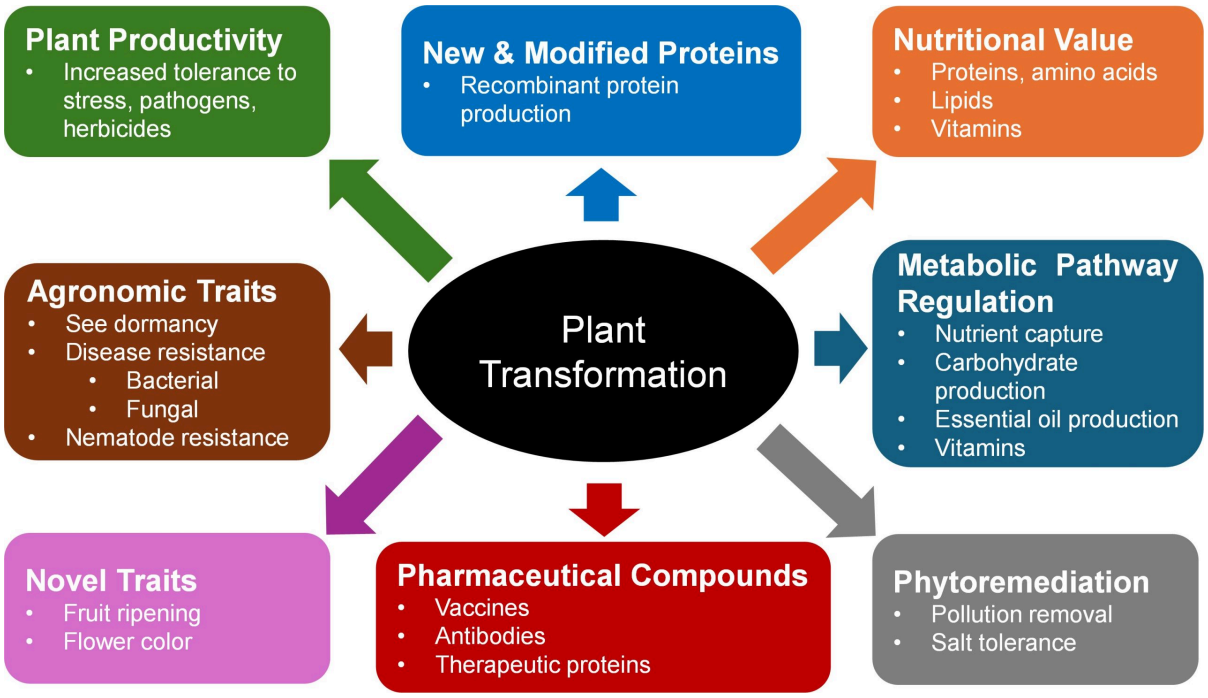


Figure 1 Applications of plant transformation. *Adapted from Newell (2000)*.

Through plant transformation, scientists have developed crops with enhanced resistance to abiotic stresses, including drought, salinity, and flooding, as well as biotic stresses such as pests and diseases. Additionally, the process has resulted in increased yields, improved nutritional quality, and enhanced pharmaceutical compounds in plants. Furthermore, scientists utilize plant transformation to study functional genomics and unravel the roles of specific genes in plant growth, development, and physiology.

History of plant transformation

The history of plant transformation traces back to early studies of plant totipotency and tissue culture (Figure 2). The notion that plant cells could regenerate into whole plants originated in the 18th century, when **polyembryony** was first observed in citrus seeds, which inspired the artificial regeneration of plants from somatic tissues. The first reports detailing in vitro cell culture date back to 1902, when Haberlandt proposed that isolated plant cells, under appropriate conditions, could grow and develop into whole plants, introducing the concept of **totipotency**. Although unsuccessful, his idea was later adopted when Gautheret (1934) achieved continuous in vitro culture of vascular plant tissues in the 1930s (Gautheret, 1939). By the 1950s, Straus and LaRue (1954) established the first maize callus cultures from immature endosperm, marking an early step toward cereal tissue culture (Straus & LaRue, 1954). Later, Steward et al. (1958) demonstrated somatic embryogenesis in carrots, confirming regeneration from single somatic cells (Steward et al., 1958).

In the 1970s, Green and Phillips established stable embryogenic maize callus cultures and identified genotype A188 for its high embryogenic potential (Green & Phillips, 1975). Around the same time,

Agrobacterium tumefaciens, first identified in 1907 as the cause of crown gall disease, was shown to transfer a DNA segment (T-DNA) into plant genomes. This discovery revealed a natural mechanism for genetic transformation, and its subsequent development led to the stable transformation of transgenic plants, including tobacco, petunia, and sunflower, in the 1980s. However, despite its success in dicots, the transformation of monocots such as rice and maize remained elusive due to host incompatibility, which prompted the development of alternative DNA delivery methods.

Maize transformation was achieved by Armstrong and Green (1985) through friable Type II callus cultures from the inbred line A188, enabling efficient and stable plant regeneration. This system became a cornerstone of modern maize transformation. In the 1980s, Sanford and colleagues developed the biolistic **gene gun**, which enabled the creation of the first fertile transgenic maize (Klein et al., 1987).

By the 1990s, *Agrobacterium*-mediated transformation was successfully extended to rice and other maize genotypes, helping to bridge the gap between dicots and monocots. In the 2000s, transformation research using morphogenic regulators such as *WUSCHEL* (*WUS*) and *BABY BOOM* (*Bbm*) (Lowe et al., 2016), enabled the reprogramming of somatic cells into an embryonic state, thereby improving regeneration efficiency and expanding the range of transformable genotypes.

Building on these advances, the 2010s witnessed the integration of CRISPR/Cas genome editing into transformation pipelines, enabling the precise and efficient modification of plant genomes within regenerable tissues. Current innovations leverage the use *of GROWTH-REGULATING FACTOR5, a chimeric fusion of GRF4 with its cofactor GRF-interacting Factor1 (GRF4-GIFI), Wuschel-related homeobox 5 (WOX5), and WOX2a,* among others, to enhance transformation efficiency (Lee & Wang, 2023) (Figure 2).

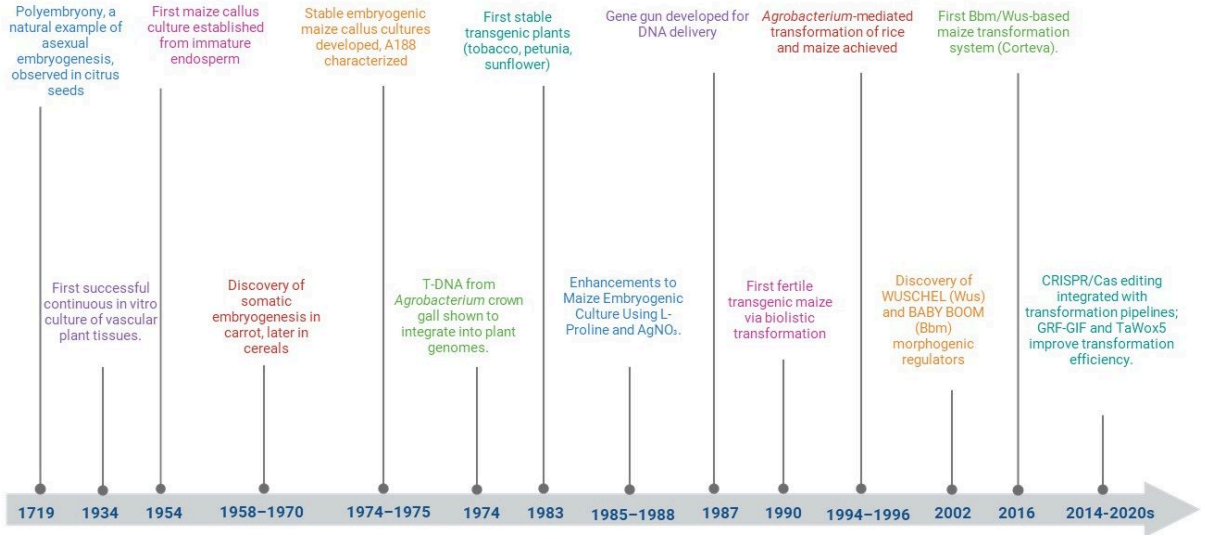


Figure 2 Timeline of the history of plant transformation.

Incorporation of superior genes

The goal of plant transformation is to introduce specific, well-characterized genes into plants to confer desirable traits, such as pest resistance, disease tolerance, nutritional enhancement, and stress adaptation (Figure 1). As mentioned earlier, unlike conventional breeding, which transfers thousands of genes at once, transformation enables the precise insertion of individual genes into plant genomes, including those from unrelated species. It also allows the use of genes from organisms that cannot be introduced naturally into a plant due to crossing barriers that prevent interspecific hybridization. These features make plant transformation a valuable tool for developing male-sterile parents, which are essential for producing hybrid cultivars. Beyond enhancing yields, improved nutritional quality, and enhanced pharmaceutical compounds, plant transformation also facilitates functional genomics studies to identify the roles of specific genes in plant growth and development.

Transformation system

A transformation system refers to the entire process and a set of tools used to introduce and integrate foreign DNA (transgene) into the plant genome, followed by regeneration of whole, fertile, and stable transgenic plants. A successful transformation system combines the following.

1. A suitable explant (tissue that regenerates),
2. An efficient DNA delivery method, and
3. A reliable selection system to identify transformed cells

Choice of explant

In preparing plant tissue for transformation, the cells must be **competent to integrate** the construct into their genome. The transformed cells must also be competent for **regeneration** into a plant (see *Plant Breeding Methods*). However, not all tissues are competent for transformation and regeneration, and different plant species vary considerably in their competence for cell transformation and regeneration potential. Typically, only a fraction of cells or tissues can regenerate, and within that subset, only a small portion is capable of transformation.

In certain instances, tissues that are less competent can be induced to become competent by specific treatments (such as exposure to plant growth regulators or stress signals). A cell's response to wounding can also influence its capacity for regeneration. In some dicot species, wounding triggers the formation of **dedifferentiated cells** around the wound site, which can then be induced to regenerate under appropriate conditions. However, monocot species respond differently to wounding, often resulting in cell death near the wound site. This cell death limits regeneration due to the lack of dedifferentiated

cells. Consequently, transformation of many monocots relies on callus tissue or immature embryos, as embryogenic tissue can regenerate without undergoing dedifferentiation.

Explant or tissue types used for the transformation of plants

An explant refers to any part of a plant used as starting material for tissue culture. It is often confused with callus tissue, which consists of a mass of undifferentiated cells. Callus cells are **totipotent**, meaning a single cell can divide, differentiate into various specialized cell types, and ultimately regenerate into a whole plant.

Protoplasts from whole tissue or suspension cells

Protoplasts are plant cells that have had their cell walls removed using enzymatic or mechanical methods and are surrounded only by plasma membrane. Protoplasts are generally isolated from differentiated plant tissues and can regenerate new cell walls, forming callus with totipotent potential. These properties enable protoplasts to regenerate into whole plants. Additionally, the absence of cell walls, which normally resist foreign DNA uptake, enables protoplasts to readily take up foreign DNA for genetic transformation. Protoplasts have been utilized to study cellular processes, including hormone signaling, protein function, DNA transfection, and viral infection, among others. Crops such as potatoes and rice can be transformed from protoplasts.

Immature embryo

This explant is widely used for many crop transformations, such as maize, due to its high regenerative capacity and presence of many totipotent cells. Immature embryos are developmentally active and responsive to in vitro culture, making them ideal for inducing embryonic calli and regenerating gene-edited or genetically modified plants. In maize, for example, immature embryos can be harvested 8 to 20 days after pollination. Transformation efficiency using immature embryos is influenced by various factors, including genotype, embryo size, and hormone combinations, and must often be optimized for consistent results. Other crops transformed using immature embryos include Barley, wheat, sorghum, and rice.

Intact cells from suspension cultures

These offer a homogenous and rapidly growing source of explants for plant transformation. Intact cells from suspension cultures enable efficient metabolite production and analysis at the cellular level. Furthermore, their accelerated metabolic and biosynthetic activity, resulting from uniform exposure to nutrients and growth regulators, helps reduce time and labor compared to solid media. Common grape vine, *Citrus* spp, and rice are among the crops transformed using Intact cells from suspension cultures.

Leaf discs

These are widely used in *Agrobacterium tumefaciens*-mediated plant transformation due to several rea-

sons, such as their accessibility and responsiveness. Tobacco and other model plants exhibit higher transformation efficiencies (TE) than recalcitrant species, such as potato and soybean, when using leaf discs. This is because TE varies depending on species, genotypes, and developmental stages of the explant. In addition, factors such as explant age and culture conditions can influence TE in leaf disc-based plant transformation.

Cotyledon sections

Embryonic leaves within a seed can regenerate into whole plants, especially in dicots. Juvenile cotyledon sections often regenerate at a higher rate than mature tissues, making them ideal for transformation. Crops transformed using cotyledon sections include tomatoes and cucurbit species, such as watermelons and cucumbers.

Embryogenic tissue

The ability of embryonic tissues to rapidly proliferate and produce multiple transformed lines makes them ideal explants for plant transformation. Embryogenic tissues originate from somatic cells that undergo dedifferentiation under specific hormonal conditions, allowing them to develop into whole plants through the process of somatic embryogenesis. Crops transformed using embryogenic tissues include Barley, wheat, maize, sorghum, and rice.

Seedling leaf whorls

Seedling leaf whorls are the central, tightly wrapped leaves at the shoot apex of young seedlings. Recently, leaf whorls have emerged as a promising explant for regenerating recalcitrant monocot genotypes in combination with morphogenic transcription factors *BABY BOOM* (*Bbm*) and *WUSCHEL2* (*Wus2*). During the transformation process, three to four outer leaf sheaths are removed to expose the inner, greener, and softer leaf tissues that are more responsive to transformation and regeneration. Examples of crops that have been transformed using seedling leaf whorls include maize and sorghum, rice (both indica and japonica), teff, switchgrass, pearl millet, foxtail millet, barley, and rye

Introduction of DNA into the Cells of the Target Tissue

A requirement of a transformation system is that the DNA of interest must be able to integrate into the genome of target plant cells that can regenerate. It is also important that the process of integrating DNA into a cell does not render the cells unable to regenerate a new plant in tissue culture. Several methods are available to introduce DNA into the host. The method used depends on the host, its capacity to receive and incorporate the foreign DNA via the method, and the host's ability to regenerate if transformation occurs in culture. The following methods can be used for plant transformation, with the choice depending on the recipient species and the type of transformation: (i) stable transformation or (ii) transient expression of genes.

1. **Direct DNA uptake**—DNA cannot be taken directly into cells because of the cell wall; therefore, protoplasts are used. DNA is imported into the protoplasts with polyethylene glycol.
2. **Microinjection**—Microcapillaries inject DNA directly into the host nucleus.
3. ***Agrobacterium***—mediated gene transfer—*Agrobacterium tumefaciens* or *A. rhizogenes* is used to deliver DNA into the recipient plant cells.
4. **Particle bombardment or biolistic transformation**—a physical means of transfecting cells by bombarding tissue with high-velocity DNA-coated particles. This method can be applied to protoplasts, intact cells isolated from a suspension culture, or whole tissues or organs.
5. **Electroporation**—DNA is electroporated into protoplasts.

Here, we discuss two commonly applied plant transformation procedures:

1. *Agrobacterium*-mediated gene transfer
2. Particle bombardment or biolistic transformation

Agrobacterium-mediated gene transfer

The genus *Agrobacterium* has been divided into several species based on the disease symptoms and host range. For example, *A. tumefaciens* causes crown gall disease , *A. rhizogenes* induces hairy root disease, *A. rubi* causes cane gall disease, and *A. Vitis*, which has been most recently proposed to cause galls on grapes and a few other plants. These soil-pathogenic bacteria are valued by scientists in plant biotechnology and genetic engineering for their ability to transfer DNA into plants via **horizontal gene transfer**. This unique ability has earned *Agrobacterium* the title of "nature's genetic engineer". *Agrobacterium*-mediated gene transfer leverages the bacterium's capability to introduce a gene of interest into plant cells to create transgenic plants. Of the several species, *A. tumefaciens* is the most used species for generating stable transgenic plants, while *A. rhizogenes* produces transformed adventitious "hairy roots", often applied in rapid functional analyses of root-expressed traits. In the following section, you will learn about the crown gall disease to some extent and the application of *Agrobacterium* in mediating stable plant transformation.

The Molecular Basis for the Crown Gall Disease

Agrobacterium tumefaciens infects wounded plant tissues by transferring a segment of its tumor-inducing (Ti) plasmid, a region called T-DNA, into the plant genome. Upon infection, this leads to crown **gall disease**, characterized by the formation of tumors on plant stems at or near the soil surface. This bacterium belongs to the family *Rhizobiaceae*, which includes Rhizobium, a genus involved in the formation of symbiotic nodules in leguminous plants. The Ti plasmid is exchanged between *Agrobacterium* species to facilitate the spread of this genetic element. *A. tumefaciens* tumors result from the rapid proliferation

of plant cells after bacterial infection. Remarkably, tumor cells isolated from an infected plant can continue to grow in culture, even after the culture is cured of *Agrobacterium*, and in the absence of added external plant hormones, such as **auxins** and **cytokinins.** For many plant species, these hormones are normally needed to stimulate growth in culture, and the ratio of auxins to cytokinins determines whether roots, shoots, or undifferentiated tissue will form.

Agrobacterium-infected cells undergo natural transformation, resulting in undifferentiated growth and ultimately leading to the formation of tumors. The promoters driving these genes are constitutive and highly active in plant tissues, resulting in continuous, upregulated production of auxins and cytokinins. Examples of these genes are *aux, cyt1*; also called *iaaM and iaaH, iptZ*. Enzymes encoded by these genes convert existing plant compounds to **indole-3-acetic acid (IAA)** and **isopentenyladenosine-5-monophosphate**, naturally occurring auxin and cytokinin, respectively. As a result, infected plant cells are in a continuous state of undifferentiated proliferation, leading to tumor/gall formation. **Opines** are novel compounds derived from amino acids, rich in carbon and nitrogen. **The most common opines are Octopine (carboxylarginine) and nopaline (dicarboxypropylarginine)**. Opines cannot be metabolized by plant cells or other microorganisms, but *Agrobacteria* can utilize *them* because the Ti plasmid contains additional genes for opine metabolism. Opine synthesis genes are located on the T-DNA that gets integrated into the plant genome, whereas opine metabolism genes remain in the *Agrobacterium* on the Ti plasmid. This creates a favorable environment for bacterial growth because plant cells are programmed to produce N- and C-rich compounds that can only be utilized by *Agrobacterium*.

Scientists harness this mechanism by replacing tumor-inducing genes with genes of interest to produce transgenic plants. The transformation process can be divided into four steps: T-DNA processing, T-DNA strand transfer, T-DNA integration, and transgene expression.

T-DNA processing and integration into the plant genome

The Ti plasmid, typically 200-800 kb in size, contains genes responsible for virulence, host specificity, T-DNA transfer, phytohormone synthesis, opine synthesis, and opine catabolism. A 25-bp border sequence flanks the T-DNA (Gelvin, 2003).

Agrobacterium-mediated transformation, a critical early step in the activation of the T-DNA transfer pathway, involves the induction of bacterial virulence (vir) genes through a signal transduction pathway (Figure 3). Wounded plant tissue produces phenolic signal molecules, such as acetosyringone, which are detected by sensors produced by *Agrobacterium*, specifically VirA and VirG proteins (Winans, 1992). VirA, a membrane-bound sensor kinase, is autophosphorylated upon detection of plant phenolics with the help of ChvE, a sugar-binding protein, and this in turn activates VirG by transferring the phosphate group. This causes the phosphorylated VirG to upregulate the expression of *vir* genes, initiating T-DNA processing and transfer cascade pathways.

The VirD1/VirD2 complex is involved in T-DNA processing from the Ti plasmid by nicking the bottom strand of the DNA at the 25bp border sequences that flank the T-DNA. However, VirD2 remains covalently attached to the 5' end of the resultant single-stranded T-DNA strand, forming the core of the complex, which is transferable. The T-DNA strand is then escorted by VirD2 and later by VirE2, a single-stranded DNA-binding protein, to create the "T-complex". VirE2 associates with the T-strand within the bacterium or upon entry into the plant cytoplasm to facilitate nuclear targeting and protection of the T-strand from nucleases. The export of T-DNA occurs via a type IV secretion system, composed of 11 VirB proteins and VirD4, which acts as a coupling protein that guides the T-complex to the VirB translocation channel. Some VirB proteins form the channel, while others, such as VirB2 and VirB5, form a T-pilus that anchors *Agrobacterium* to the plant cell surface and acts as a conduit for DNA/protein passage (Gelvin, 2017).

VirD2 possesses nuclear localization signal sequences that facilitate nuclear import and direct the T-complex to the plant nucleus. Studies have shown that VirD2 can direct small T-DNA segments into the plant nucleus independently, but for larger DNA molecules, VirE2 is required. These proteins together direct the T-strand to the plant nucleus for integration into the plant genome, facilitated by VirD2. Inside the nucleus, T-DNA integration occurs with the assistance of host DNA repair pathways, which are often activated by double-stranded breaks (DSBs) (Gelvin, 2003). VirD2 tethers the T-strand to the host's chromatin by interacting with plant transcription factors and DNA repair components. Microhomology-mediated end joining (MMEJ), facilitated by the action of theta polymerase, also promotes the integration of T-DNA.

> Read: Gelvin, S.B. (2003) for a detailed explanation of the molecular basis of *Agrobacterium*-mediated transformation and how T-DNA moves into plant cells.

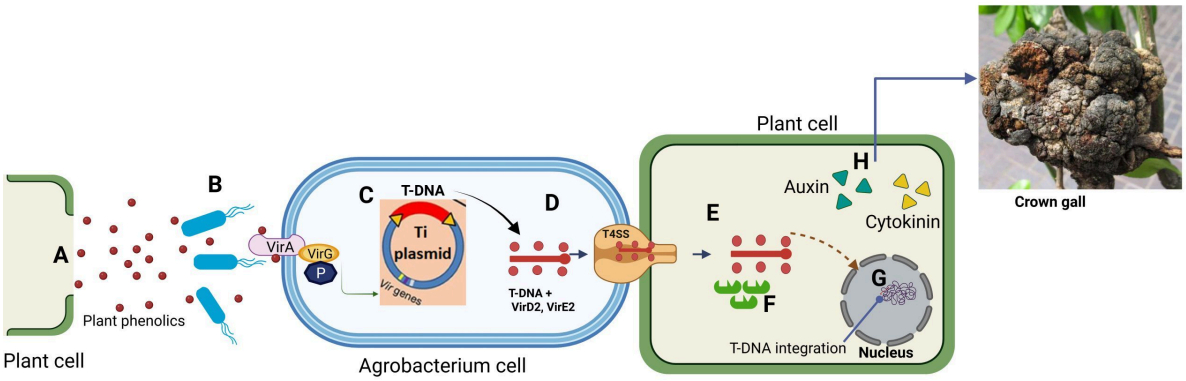


Figure 3 Transfer of T-DNA from *Agrobacterium* into a plant cell. **A)** A wounded plant cell releases phenolic signals like acetosyringone. **B)** Agrobacteria sense the plant-derived signals and move towards the wounded plant cell. VirA protein on the membrane of A. tumefaciens cell recognizes the wound-triggered plant signals and phosphorylates the sequence-specific DNA-binding protein VirG. **C)** Virulence gene induction, required for infection, **D)** T-DNA is cut from the *Agrobacterium*'s DNA plasmid and injected into the plant cell. T-complex consists of T-DNA and various Vir proteins, and it enters plant cells through a type IV secretion system (T4SS). **E)** T-DNA is transported toward the plant cell nucleus with the help of VirD2 and VirE2. **F)**. Plant enzymes and other host factors. **G)** Nuclear import of T-DNA and effector proteins in the plant cells. **H)** After T-DNA integration and expression in the plant genome, plant hormones such as auxins and cytokinins are produced, leading to tumor/gall formation (Weir & Dalzell, 2020). *Image by Faizo Kasule*.

Use of A. Tumefaciens in Plant Transformation

The steps in *Agrobacterium*-mediated plant transformation were described in *Plant Breeding Methods*. The process, however, is complex and relies on precise interactions between *Agrobacterium* virulence (Vir) genes and plant cellular components. A key event is the transfer of **T-DNA** from the bacterium into the plant genome, which is mediated by a **type IV secretion system** and guided by Vir proteins. Understanding the molecular mechanisms of T-DNA processing, transfer, and the plant's response is essential for enhancing the efficiency of *Agrobacterium* in biotechnology applications.

In the early days of genetic engineering, scientists became interested in *Agrobacterium*'s ability to insert its DNA into plant cells. They subsequently modified the *Agrobacterium* by removing its ability to cause disease but retaining its ability to transfer foreign DNA into plant cells. The transferred DNA (T-DNA) has a **left (L) and a right (R) border** sequence, and all DNA in between these two borders is transferred to the plant genomes (Figure 4).

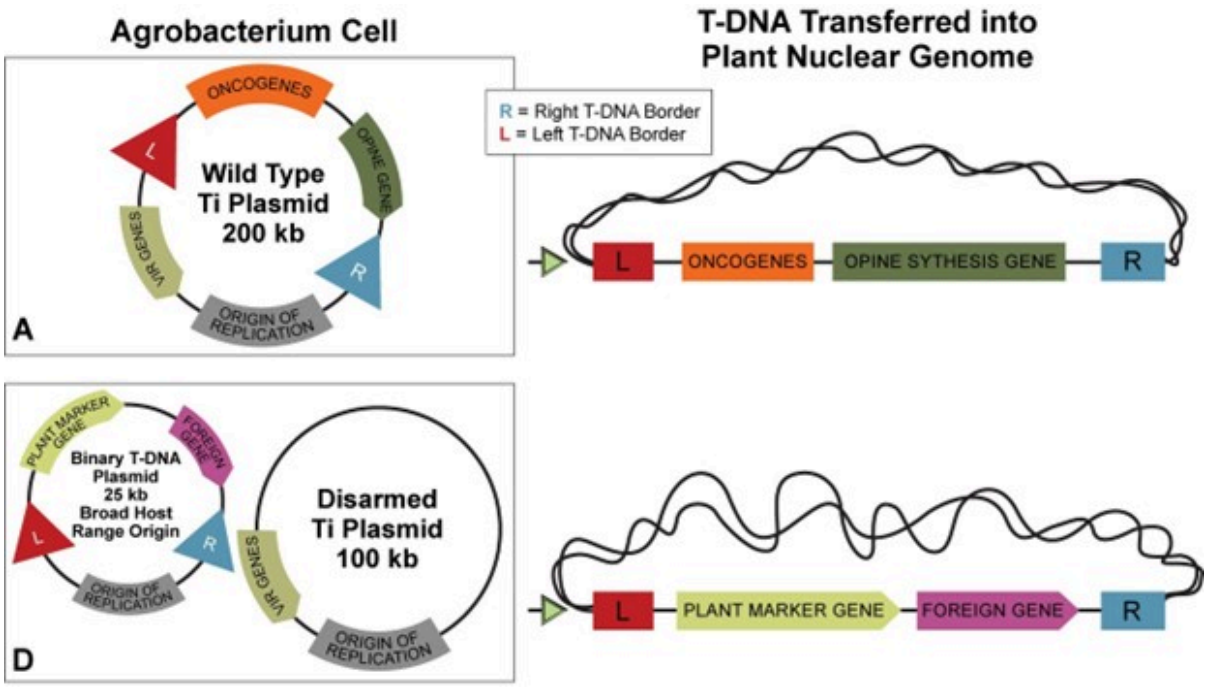


Figure 4 Ti plasmid-mediated T-DNA transfer. (A) Natural wild-type *Agrobacterium* Ti plasmid. (B) Binary T-DNA plasmid in conjunction with a "disarmed" Ti plasmid. Since the "disarmed" Ti plasmid does not contain a T-DNA region, the vir genes can only act to transfer the T-DNA on the binary T-DNA plasmid. Note: The wavy lines represent plant DNA.

The Ti plasmid has been engineered to enhance its usefulness for plant transformation. For instance, genes responsible for altered plant hormone production (oncogenes) and other genes (such as those involved in opine synthesis) that are unnecessary for DNA transfer have been eliminated, thereby preventing aberrant plant growth and metabolism. A binary T-DNA plasmid vector can be constructed to contain the left and right T-DNA borders and **restriction endonuclease sites** between the borders to enable foreign DNA (any transgene you want to transfer into the plant genomes) to be cloned into the T-DNA region. A **plant selectable marker** (often an antibiotic resistance gene or herbicide tolerance gene) is also placed in the T-DNA. The right border initially enters the plant cells, and the subsequent selection of T-DNA molecules using a selectable marker adjacent to the right border results in partial integration of the T-DNA molecules. Placing the selectable marker next to the left border can enhance the complete integration of T-DNA and the expression of transgenes (Bhattacharyya et al., 1994).

The binary plasmid also has an origin of replication that enables its replication in both *Escherichia coli* and *Agrobacterium*. This principle enables the cloning of foreign genes into the T-DNA molecule in E. coli using a binary plasmid, which is significantly smaller than the Ti plasmid and, therefore, easier to manipulate. The **binary vector** does not contain the virulence genes necessary for infection and transfer of T-DNA. The binary plasmid is introduced into an *Agrobacterium* strain that carries a Ti plasmid, lacking tumor-inducing genes as well as the left and right border sequences. It is called a disarmed Ti plasmid, but it contains *vir* genes, which allow the *Agrobacterium* cells to infect plant cells and transfer

T-DNA from the binary plasmid. An assembly of the DNA elements used for transformation is called a construct (see *Plant Breeding Methods*). A construct for T-DNA transformation must include the T-DNA border sequences, a marker for selection, such as a gene for herbicide resistance, and one or more genes intended for transfer into the recipient genome. *Agrobacterium*-mediated transformation offers high efficiency and throughput, reliably generating single-copy DNA insertion events.

As stated earlier in the chapter, only a fraction of the cells competent for regeneration are transformed. The use of selectable markers enables the selection of rare, transformed cells. Therefore, after *Agrobacterium* infection, the recipient cells carrying the inserted T-DNA continue to grow, whereas the antibiotics or herbicides used in selection kill non-transformed cells.

Use of A. Rhizogenes in Plant Transformation

A. rhizogenes is another species that also induces abnormal plant growth, in this case, adventitious roots on shoots of dicotyledonous plants. Like *A. tumefaciens*, *A. rhizogenes* infects plants through wound sites, followed by the transfer, integration, and expression of T-DNA from the root-inducing (Ri) plasmid. The expression of T-DNA from the Ri plasmid results in the formation of fine fibrous structures referred to as hairy roots. Hairy roots are formed due to changes in the cells' physiology, rendering them more sensitive to auxins. Hairy roots are useful for producing high-value metabolites in medicinal plants and for phytoremediation studies (Suza et al., 2008).

A. rhizogenes is particularly useful for transforming legume species that are difficult to transform by the *A. tumefaciens* method (Collier et al., 2005). The transgenic roots enable the study of the function of genes required for root-specific processes (Kereszt et al., 2007). In *A. rhizogenes*-mediated transformation, the T-DNA from the binary vector is co-transferred with the T-DNA from the Ri plasmid into the host genome. Not all hairy roots may contain the transgene, so the hairy root phenotype is often used to select transformants in the absence of a selectable marker. However, molecular confirmation of transgene presence in individual hairy roots is typically done using PCR.

> Read: Ex vitro composite plants: an inexpensive, rapid method for root biology *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology

Particle Bombardment or Biolistic Transformation

Particle bombardment, also known as biolistics, is a versatile and powerful method for plant transformation that often overcomes the limitations associated with *Agrobacterium*-mediated approaches. Unlike *Agrobacterium*, biolistics methods have no biological host constraints, allowing the transformation of a wide range of plant species and tissue types, including those recalcitrant to *Agrobacterium* infection.

Biolistic transformation utilizes high-velocity gold or tungsten particles to deliver DNA molecules into living cells, resulting in stable transformation. The method can be applied to protoplasts, intact cells from a cell suspension culture, or tissues or organs, and is the most effective approach for organelle transformation, such as in chloroplasts and mitochondria. The method also supports the delivery of non-DNA-based Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) reagents, for example, RNA and ribonucleoproteins (RNPs). These reduce off-target effects and often lead to the production of transgene-free edited plants (Chapter 8), which is essential and valuable in precision plant breeding and regulatory approval (Thorpe et al., 2025).

How biolistics particle delivery works:

Microparticles of tungsten or gold coated with DNA are accelerated toward the target tissues. The particles penetrate the cell wall and carry DNA into the cells and even into the nuclei. The original biolistic instruments used gunpowder explosions to accelerate DNA-coated particles into target cells. Other instruments discharge sudden gas pressure to accelerate the particles (Figure 5).

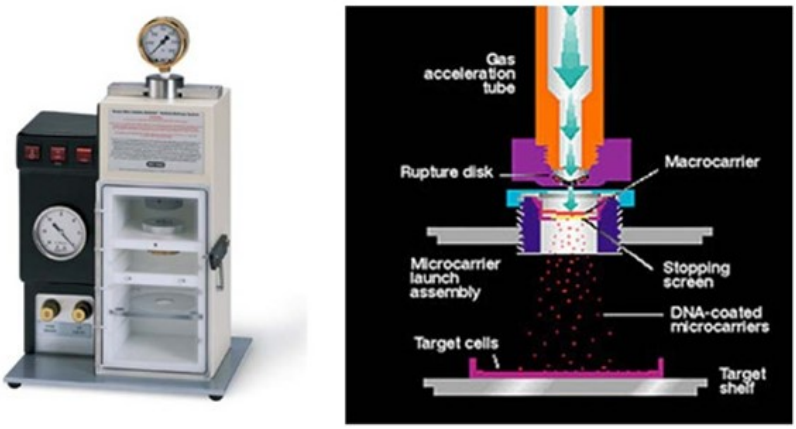


Figure 5 Biolistic PDS-1000/He System from Bio-Rad.

The PDS-1000/He system (Figure 5) accelerates nucleic acid-coated gold or tungsten microparticles (0.6–1.6 µm) to velocities necessary to penetrate cells, tissues, or organelles (Figure 6). The system utilizes a burst of high-pressure helium gas to accelerate a plastic macrocarrier disk, which carries microparticles, toward the target cells. The helium pressure used to propel the macrocarrier is determined by the choice of rupture disk, a plastic seal designed to burst at a specific pressure. A stopping screen retains the macrocarrier while allowing the microparticles to pass through and penetrate the target cells. To increase the efficiency of the process, the chamber may be evacuated to subatmospheric pressures, reducing the frictional drag on the microparticles as they travel toward the target cells.

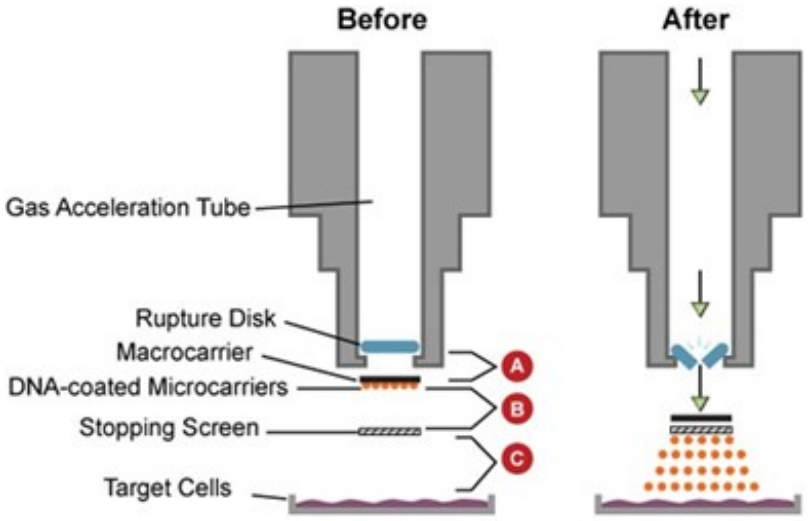


Figure 6 The biolistic transformation process.

However, the biolistic transformation process has some associated challenges, like tissue damage, lower efficiency, and the risk of multiple or fragmented gene insertions. Despite these challenges, the method remains an indispensable tool for plant transformation, especially for species or applications where other transformation methods are inadequate.

> A new biolistic enhancement that utilizes a flow guiding barrel (FGB) addresses the limitations associated with gas and particle flow. FGB significantly improves stable transformation in maize by over 10-fold and boosts transient expression and genome editing across crop species, addressing key delivery barriers. This method enables the delivery of DNA, RNA, and proteins, regardless of tissue type, genotype, or crop. This breakthrough makes biolistic genome editing more reliable and broadly applicable across plant species.

Requirements for a successful plant transformation

For the process of plant transformation to be successful, several essential components must be in place, including the facilities (tissue culture facility, greenhouse, and molecular laboratory), transformation protocols and systems, gene selection, and construct designs. Regeneration systems and Molecular Analysis of transgenics., divided into three major units, is required for a

Laboratory Facilities and Infrastructure

A specialized infrastructure divided into three major units is required for successful plant transformation. These include a tissue culture facility, a molecular laboratory, and a greenhouse.

Plant tissue culture (PTC) facility

Plant transformation relies on tissue culture for regenerating whole plants from explants. Typically, a standard PTC facility should have the following equipment: tissue culture incubators, laminar airflow hoods, multiple temperature-controlled shaking incubators, a –80°C freezer, multiple dissecting microscopes, a fluorescence microscope, a Bio-Rad PDS 1000/HE biolistic particle delivery system, a Bio-Rad Gene Pulser, and other necessary equipment for tissue culture. Additionally, the PTC facility should be equipped with autoclaves for sterilizing media, waste, and tools, a media preparation area featuring pH meters and a weighing balance, as well as a supply of lab consumables. This facility enables explant sterilization and establishes a transformation system under sterile conditions.

Molecular Biology Laboratory

This facility is important for DNA extraction, cloning, plasmid construction, PCR validation, gel electrophoresis, and bacterial transformation, among other applications. This facility should also have sterile conditions to prevent contamination during bacterial transformation.

Greenhouse Facility

After regenerating healthy and robust-looking plantlets, these can be transferred to greenhouse conditions under biosafety regulations. For proper acclimatization, temperature, humidity, and light must be regulated.

Crop Bioengineering Laboratory (CBL), Iowa State University

The CBL offers plant tissue culture and transformation services for maize and soybeans, utilizing a 1,352 ft2 lab space and a 4,100 ft2 greenhouse. Key equipment includes tissue culture incubators, laminar flow hoods, microscopes, and biolistic and *Agrobacterium* transformation systems.

Services:

Maize: *Agrobacterium*-mediated transformation of B104 immature embryos.

Soybean: *Agrobacterium*-mediated transformation of Williams 82 half-seeds.

DNA Construct Validation & Cloning: Transform plasmids into *Agrobacterium*, verify constructs, and assemble vectors for overexpression or CRISPR editing. For more information, read about the CBL and the services it offers.

Gene selection and construct design

For successful plant transformation, a well-designed gene construct is fundamental. Construct design

involves careful selection of the gene of interest, regulatory elements, selectable markers, and vector backbone to maximize transgene expression and stability.

Gene of interest (GOI)

The GOI design should be well-characterized and associated with a defined target trait, such as pest resistance, disease resistance, drought tolerance, or improved nutritional quality, among others. Efficient expression requires codon optimization to match host plant usage, especially when introducing bacterial or other foreign genes into plants.

Expression Cassette Components

The T-DNA regions typically include promoters, introns, untranslated regions, the GOI, and terminators (These are discussed in Chapter 1). Promoters can be constitutive to drive expression in all tissues or tissue-specific for targeted expressions. Inclusion of introns can enhance mRNA processing and increase transgene expression, while terminators ensure proper transcription termination.

Selectable markers and reporter genes

Selection markers identify transformed cells during plant tissue culture, while reporter genes provide visual confirmation of transgene expression and protein localization (Table 1).

Table 1 *Common selectable markers and reporter genes used in plant transformation*

Category	Gene	Function / Use	Detection / Selection method	Reference
Selectable Marker Genes	nptII	Confers resistance to Neomycin, Kanamycin, Paramomycin, Ribostamycin, Butirosin, and Geneticin (G418)	Transformed cells survive on kanamycin-containing medium	(Bevan et al., 1983)
	hpt (hptII)	Confers resistance to hygromycin	Survives on hygromycin-containing medium	(Waldron et al., 1985)
	bar/pat	Confers resistance to phosphinothricin (glufosinate herbicide)	Culturing on media with glufosinate	(Thompson et al., 1987)
	aadA	Provides resistance to streptomycin/spectinomycin	Growth on antibiotic medium	(Svab & Maliga, 1993)
Reporter Genes	GUS (uidA)	Produces β-glucuronidase enzyme	Blue staining with X-Gluc substrate	(Jefferson et al., 1987)
	GFP (Green Fluorescent Protein)	Emits green fluorescence; real-time visualization	Fluorescence microscopy	(Chalfie et al., 1994)
	RFP / mCherry	Red fluorescent protein for protein localization	Fluorescence microscopy (red channel)	(Shaner et al., 2004)
	Luciferase (luc)	Produces bioluminescence in the presence of luciferin	Light emission measured with a luminometer	(Ow et al., 1986)
	LacZ	β-galactosidase enzyme; blue/white screening	X-gal staining (less common in plants)	(Casadaban & Cohen, 1980)
	RUBY	Produces red betalain pigment, visible without a microscope	Visibly red tissues in transformed cells	(He et al., 2020)

Selection of Transformed Cells

Because only a few cells that can be regenerated will be transformed, one must have a method to eliminate the non-transformed cells to identify the transformed ones. A variety of antibiotics have been used to kill cells that are not transformed, such as kanamycin, spectinomycin, and hygromycin. Genes conferring resistance to the appropriate antibiotics are included in the T-DNA for *A. tumefaciens*-mediated transformation, and the gene of interest is in a plasmid for biolistic transformation. Other common selectable markers include herbicide resistance genes.

As mentioned earlier, certain plant species respond to wounding by producing **dedifferentiated cells** around the wound site, and these can be induced to regenerate under suitable conditions. Dedifferen-

tiated cells at the periphery lead to the development of multiple transformants when the leaf discs are placed on a plate containing the appropriate antibiotics to select transformed cells. The leaf disk transformation-regeneration process is illustrated in Figure 7.

There are also transformation protocols that do not require tissue cultures. For example, the floral dip method is extensively applied *in Arabidopsis*, where flowers are immersed in a solution containing *Agrobacterium* that carries the Ti and binary plasmid constructs.

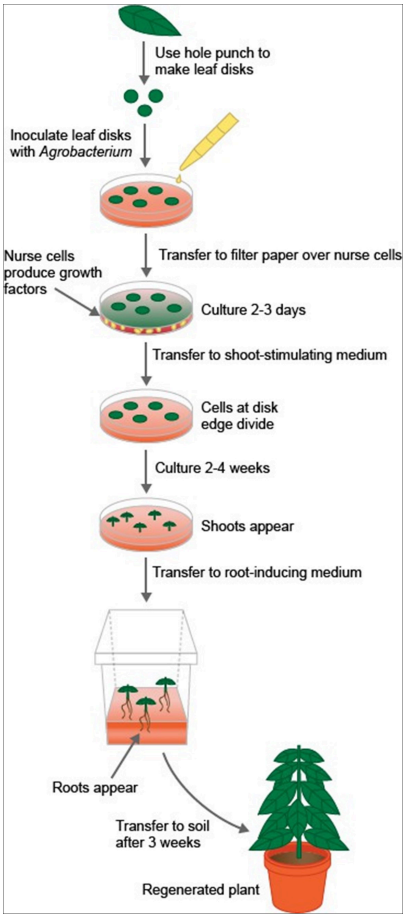


Figure 7 *Agrobacterium*-mediated Transformation and regeneration using leaf discs.

Early **meristematic tissue** of the embryo is transformed. The transformed seeds can be easily screened by growing them on media containing the appropriate selective compounds. This procedure is composed of the following steps:

1. Select a parent plant at flowering.
2. Dip flowers in a solution of *Agrobacterium* containing a binary plasmid.
3. Grow infected plants to maturity and harvest seeds.

4. Plant seeds on selective media, usually kanamycin. All seeds germinate, but only those that are transformed yield seedlings that grow vigorously. These are then transferred to the soil and grown to maturity for seed harvest. The collected seed (T2) is analogous to an F_2 segregating 1:2:1 for the transgene, assuming a single transgene in the original T1. Note that the T1 is **hemizygous** for the transgene, **not heterozygous**. Heterozygosity is when there are *two different alleles* at a single locus. The T1 is hemizygous because one chromosome carries the transgene, and the other homologous chromosome carries nothing (i.e., no transgene).

Advances in Plant Transformation

Plant transformation using ternary vector systems and auxotrophic *Agrobacterium tumefaciens*

Recent advances in plant transformation, especially using ternary vector systems and auxotrophic *Agrobacterium tumefaciens* strains, have enhanced transformation efficiency and biosafety in crop genetic engineering. The ternary vector system consists of four components: (1) an *Agrobacterium* genome, which includes C1, C2 chromosomes, plus pAT plasmid, (2) disarmed Ti plasmid, which lacks tumor-inducing genes, (3) a T-DNA binary vector carrying the gene of interest, and (4) an additional virulence (*vir*) gene helper plasmid, which boosts expression of virulence genes crucial for T-DNA transfer into plant cells (Aliu et al., 2024; Lee et al., 2023). This ternary vector system significantly enhances transformation efficiencies in recalcitrant crops, such as maize. The Wang Lab at Iowa State University has developed helper plasmids, pKL2299 (Addgene # 186332) and its enhanced version, pKL2299A (Addgene # 222105) (Figure 8), with an additional *virA* to boost *Agrobacterium*-mediated plant transformation by providing additional *vir* genes essential for T-DNA transfer (Aliu et al., 2024; Lee et al., 2023).

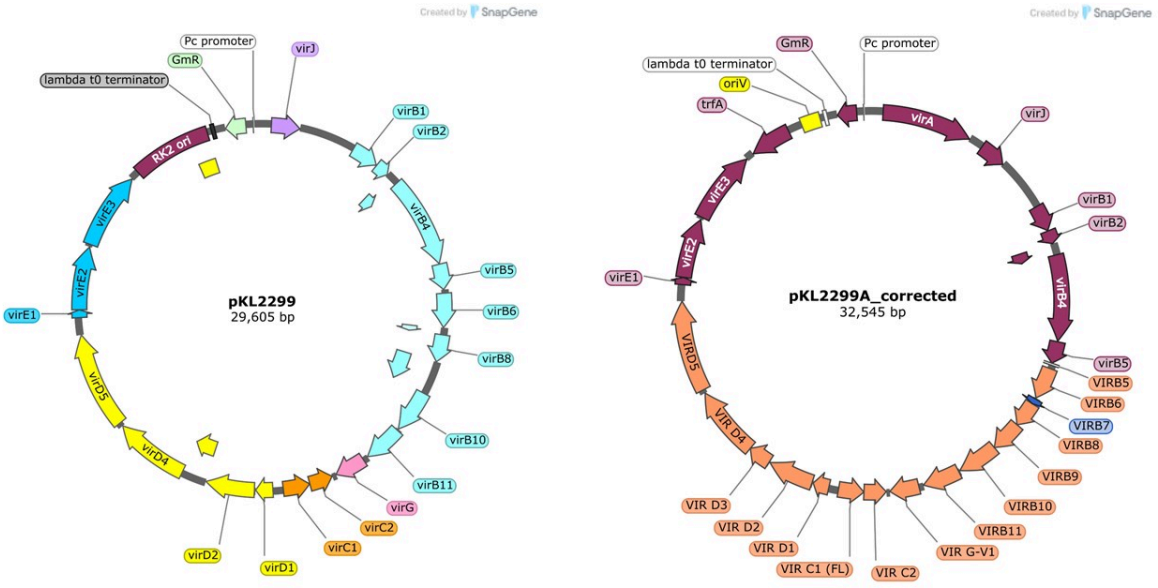


Figure 8 Maps of the vir helper plasmids pKL2299 (left) and pKL2299A (right). GmR, gentamicin resistance gene; RK2 ori, origin of replication from the RK2 plasmid; virulence genes and operons (virA, virJ, virB1-11, virG, virC1-2, virD1-5, and virE1-3) were originally from the Ti plasmid of *Agrobacterium* tumefaciens Bo542 (pTiBo542).

Auxotrophic *A. tumefaciens* strains are genetically modified bacteria that cannot synthesize essential nutrients such as thymidine, methionine, and cysteine, and thus depend on external supplementation of such nutrients for survival (Aliu et al., 2024; Prías-Blanco et al., 2022). Unlike traditional strains that may persist and damage plant tissues despite antibiotic treatments, auxotrophic strains can be easily removed by simply withdrawing the required nutrient, reducing the need for antibiotics and minimizing phytotoxicity. This makes these strains valuable in plant transformation, as their nutrient dependence helps control bacterial overgrowth during cocultivation, a significant challenge in plant tissue culture. For example, a common *Agrobacterium* strain, LBA44404, a thymidine auxotrophic mutant, engineered by deleting a *thymidylate synthase* gene *thyA*, requires external thymidine for growth, aiding the control of the bacterial overgrowth during plant transformation (Aliu et al., 2024). When combined, ternary vector systems and auxotrophic *Agrobacterium* strains offer a robust and reliable platform for enhancing and advancing plant genetic transformation and gene editing.

The concept of combined ternary vector systems and auxotrophic *Agrobacterium* strains has been applied to modify the quick corn protocol, shortening the B104 transformation timeline from 160 to approximately 50 days. The traditional B104 transformation protocols were lengthy and had low transformation frequencies. Therefore, the new protocol developed by Kang et al. (2022) shortened the regeneration time for B104 and increased the average transformation frequency to over 20%. This rapid and efficient protocol is suitable for routine genome editing using B104, without the use of morphogenic transcription factors *Bbm* and *Wus2, as is the case for the Quick Corn method* developed by Masters

et al. (2020). Building on this background and research, in 2022, Iowa State University established the Crop Bioengineering Laboratory (CBL) to meet the growing demand for plant transformation services in the research community. Using the Kang et al. (2022) Protocol, combined with ternary vector and plant selection systems, CBL achieves B104 maize transformation in approximately two months (Figure 9).

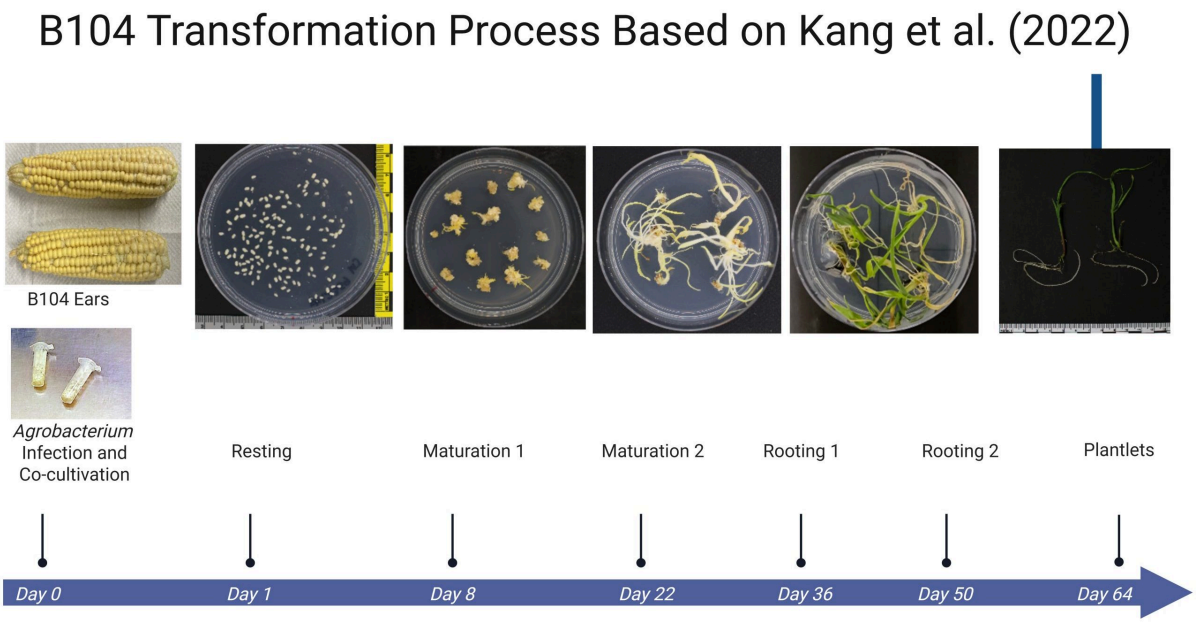


Figure 9 Improved B104 maize transformation protocol that combines a ternary helper plasmid and an autotrophic *Agrobacterium* strain. *Image by Faizo Kasule. Picture Courtesy: Qing (Jessica) Ji and Faizo Kasule*.

Morphogene-Assisted Transformation (MAT): A genotype-flexible breakthrough in Plant Biotechnology

A new advancement in plant biotechnology that uses morphogenic transcription factors (MTFs) has greatly improved the efficiency and flexibility of genetic transformation. These morphogenic genes, which are key regulators of plant development, enhance both regeneration and the transformability of a wider range of genotypes, including those previously considered recalcitrant. The **ectopic expression** of MTFs induces somatic embryogenesis and shoot formation directly from explant tissues. This approach effectively overcomes genotype dependence, a long-standing challenge in monocot transformation.

The MTF breakthrough was spearheaded by scientists from Corteva Agriscience, and in 2016, they showed that co-expression of maize *BABY BOOM* (*Bbm*) and *WUSCHEL2* (*Wus2*) genes produced high transformation frequencies in maize inbred lines that were previously non-transformable using traditional biolistic and *Agrobacterium* transformation protocols (Lowe et al., 2016). Ectopic expression of *Bbm* and *Wus2* increased transformation frequencies in maize, sorghum, rice, among other crops (Lowe et al., 2016). *Bbm*, one of the AP2/ERF family of transcriptional factors, is involved in plant growth and development processes, particularly in embryogenesis and cell regeneration , while *Wuschel* encodes a

homeodomain transcription factor that regulates stem cell fate in plants, particularly in the shoot apical meristem.

The MAT technology was expanded in 2023 by Corteva Agriscience, and this time, the team used seedling leaf whorls as a new explant for monocot transformation (Wang et al., 2023). Given the difficulty in isolating immature embryos and seed from some grasses like teff and small millets, the use of these seedling leaf whorls as alternative explants for *Agrobacterium*-mediated transformation employing *Bbm* and *Wus2* has expanded the list of recalcitrant monocot crop species (Fig. 6), which are now transformable using this leaf transformation method, including maize, sorghum, rice, switch grass, pearl millet, teff, foxtail millet, barley, wheat, and rye (Wang et al., 2023). This leaf transformation method utilizes seedling leaf whorls as alternative explants, and MTFs provide a faster, more accessible, and cost-effective process, which eliminates the need to maintain a state-of-the-art greenhouse required for providing immature embryos (Azanu et al., 2025).

> Traditional transformation of grasses like maize relies on immature embryos, which limits efficiency and depends on greenhouse-grown ears that are both costly and seasonally constrained. A breakthrough approach now uses seedling leaf-whorl explants together with the morphogenic genes *WUSCHEL2* (*Wus2*) and *BABY BOOM* (*Bbm*) in an *Agrobacterium*-mediated vector system. This innovation significantly improves transformation and genome editing efficiency, providing a practical alternative to immature embryo–based methods. For more information, read: Leaf transformation for efficient random integration and targeted genome modification in maize and sorghum.

Regeneration of New Plants from Tissue Culture

The tissue selected for transformation is usually chosen for its strong regenerative capacity. The strategies and tissue culture techniques used to regenerate plants vary widely among species and depend greatly on the type of tissue used. In some cases, the process is as much an art as it is a science. Factors influencing regeneration include the composition of the culture medium, particularly the types and concentrations of hormones, as well as light, temperature, and incubation time. **Auxins** promote root formation, while **cytokinins** stimulate shoot development. Regenerated transgenic plants may require six months or longer to mature in soil, depending on the species. Moreover, reproducing a successful regeneration protocol in another laboratory can be challenging because subtle and often unrecognized factors affect the outcome.

Verification of the Success and Stability of Transformation

It is important to screen progeny from regenerated plants to verify that the transgene is integrated into the genome and genetically transmitted. The following methods are commonly used for transgene analysis:

Detection of DNA

Southern hybridization and quantitative PCR (see Chapter 6 on Polymerase Chain Reaction) can be used to detect the presence and determine the copy number of the transgene.

Detection of mRNA

Detection of the transgene is necessary but insufficient to guarantee the expression of the transgene. The integration of transgenes at specific regions of the genome suppresses their expression. For many transformations, there is a broad variation in the level of expression among **independent transformants.** Independent transformants are individuals transformed with the same construct, but they arose from independent transformation events. Thus, they are not clonal products of a single transformed cell; therefore, the T-DNA is likely to have integrated into different regions in the genome in independent transformants. Detection of mRNA is considered to determine the extent of transgene expression levels. Northern blot hybridization, RT-PCR, and qRT-PCR are commonly used methods for detecting mRNA.

Detection of Protein

For example, the mRNA product may be present but not translated efficiently if a mutation occurs during the cloning of the gene into a binary vector. Therefore, detecting the protein product of the gene is a crucial step in validating the success of transformation. A common method used to detect specific proteins is Western blotting. Western blotting utilizes gel electrophoresis to separate proteins based on their size or structure, followed by the transfer of these proteins onto a membrane, where they are detected using antibodies specific to the protein of interest.

Phenotype of Biochemistry

Depending on the function of the transgene, one might expect an altered phenotype. However, an altered phenotype may result from mutations in other genes during tissue culture. The process of T-DNA integration is random, and the transgene may insert within a functional gene, rendering it nonfunctional. A study of several independent transformants and/or co-segregation of transgenes with phenotypes ruled out such artifacts. Biochemical analysis, such as enzyme activity or metabolite detection, may also be used to test the expression of transgenes in transgenic plants.

Transformation application in different main crops

Plant transformation is preferred because it enables precise, efficient, and targeted improvement of crops, surpassing the capabilities of traditional breeding. Plant transformation technologies have been widely applied to enhance economically important traits across major crops (Table 2)

Table 2 *Application of Plant transformation in major crops. Adopted from USDA Agricultural Marketing Service (AMS) List of Bioengineered Foods)*

Crop	Transformed Trait(s)	Application / Purpose	Reference
Maize	Bt insect resistance	Protect against insect pests	(Abbas, 2018)
	Herbicide tolerance	Roundup resistance	(Brookes, 2022)
	Drought tolerance	Survive water stress	(Edge et al., 2017)
Rice	Provitamin A (Golden Rice)	Nutritional enhancement	(Potrykus, 2001)
	Bacterial blight resistance (Xa21)	Disease resistance	(Fiyaz et al., 2022)
	Drought tolerance	Climate adaptability	(Ravikumar et al., 2014)
Wheat	Fusarium resistance	Disease reduction	(Mao et al., 2023)
	Herbicide tolerance	Weed management	(Nakka et al., 2019)
	Salt/drought tolerance	Abiotic stress resilience	(Mao et al., 2023)
Soybean	Herbicide tolerance (Roundup Ready)	Efficient weed management	(Lundry et al., 2008)
	Insect resistance	Protect against insect pests	(Parrott et al., 1994)
	High oleic acid content	Improved oil quality	(Buhr et al., 2002)
Cotton	Bt insect resistance	Reduce bollworm damage	(James, 2007)
	Herbicide tolerance	Glyphosate resistance	(Kishore et al., 1992)
	Fiber quality improvement	Better textile quality	(Constable et al., 2015)
Tomato	Delayed ripening	Shelf-life extension	(Gupta et al., 2013)
	Virus resistance	Disease resistance	(Raj et al., 2005)
	Higher lycopene	Nutrition improvement	(Li et al., 2018)
Potato	Late blight resistance	Reduce post-harvest losses	(Bubolz et al., 2022)
	Vitamin B6	Nutritional enhancement	(Bagri et al., 2018)
	Cold storage tolerance	Extend storage life	(Chen et al., 2008)
Cassava	Virus resistance (CMD, CBSD)	Disease control	(Lin et al., 2019)
	Biofortification (vitamin A)	Nutrition improvement	(Beyene et al., 2018)
	Fusarium wilt resistance	Protect yield	(Harding et al., 2025)
Banana	Bacterial wilt resistance	Disease protection	(Tripathi et al., 2017)
	Provitamin A biofortification	Nutrition improvement	(Paul et al., 2017)
	Striga resistance	Improve yield in Striga-infested areas	(Kaniganti et al., 2025)

Crop	Transformed Trait(s)	Application / Purpose	Reference
Sorghum	Drought tolerance	Adaptation to dry conditions	(Ferguson et al., 2024)
	Cold tolerance	Enhanced cold stress tolerance	(Gierczik et al., 2019)
	Salt tolerance	Stress resilience	(Mian et al., 2011)
Barley	Drought tolerance	Adaptation to dry conditions	(Feng et al., 2020)

Examples of accomplishments in plant transformation

Table 3 highlights key achievements in plant transformation that have significantly influenced agriculture, food security, and biotechnology.

Table 3 *Famous Accomplishments in Plant Transformation*Adopted from USDA Agricultural Marketing Service (AMS)AMS List of Bioengineered Foods)

Bioengineered Crop	Purpose
Alfalfa	Engineered mainly for herbicide tolerance (e.g., glyphosate-resistant).
Apple (ArcticTM varieties)	Engineered to reduce enzymatic browning by silencing PPO genes.
Canola	Modified for herbicide tolerance and high oleic acid profiles.
Corn (Maize)	Includes Bt insect-resistant and herbicide-tolerant varieties.
Cotton	Modified for insect resistance (Bt) and herbicide tolerance. Used for cottonseed oil.
Eggplant (BARI Bt Begun)	Bt eggplant grown mainly in Bangladesh to resist fruit and shoot borer pests.
Papaya	Engineered for resistance to Papaya Ringspot Virus (PRSV); saved Hawaii's papaya industry.
Pineapple (Pink-flesh)	Engineered by inserting a tangerine gene to increase lycopene → pink flesh.
Potato	Non-browning, reduced bruising, low acrylamide, late blight-resistant (Innate® Potato).
Soybean	Herbicide tolerance (e.g., Roundup Ready); high oleic, disease-resistant varieties.
Squash (Summer squash)	Engineered with viral coat protein genes for resistance to CMV, WMV, and ZYMV.
Sugarbeet	Modified mainly for herbicide tolerance.
Sugarcane (Bt)	Engineered for insect resistance (mainly in Brazil).

Concerns about GMOs

Genetically modified (GM) crops offer numerous benefits that can enhance agricultural productivity and

food security. They can provide improved nutrition, increased resistance to pests and diseases, higher yields, and greater tolerance to environmental stresses such as drought and salinity. Advances in gene editing now allow for precise modifications, including the development of transgene-free plants, which can improve public acceptance and facilitate regulatory compliance.

Although concerns have been raised regarding potential health and environmental impacts, evidence indicates that, when properly regulated and responsibly managed, these risks are minimal compared to the benefits of GM crops. A major advancement addressing GMO concerns is the use of gene editing technologies, such as CRISPR/Cas9, which enable precise modifications of plant genomes without introducing foreign DNA. These transgene-free edited crops mimic natural mutations and are increasingly exempt from strict GMO regulations in many countries. As a result, they are more publicly acceptable while still offering advantages such as disease resistance, enhanced nutrition, and improved climate resilience.

Overall, while traditional GMOs sparked ethical and ecological debates, modern gene editing represents a more precise, safer, and publicly acceptable alternative. A detailed exploration of gene editing technologies, their regulatory status, and applications is presented in Chapter 8.

Chapter Summary

Plant transformation enables the introduction of new genes for desirable traits into existing crop varieties. With plant transformation, sexual compatibility becomes irrelevant, and the process of developing a new variety becomes faster because transformed cells expressing genes can be selected directly. Successful transformation requires a well-coordinated transformation system, which integrates three key components: a suitable explant (tissue that regenerates), an efficient DNA delivery method, and a reliable selection system to identify transformed cells. In monocots, embryonic tissues such as immature embryos, calli, or seedling leaf whorls are commonly used, while in dicots, leaf discs, cotyledon sections, and embryonic calli are preferred. Several methods are available to introduce DNA into the host. These methods include direct DNA uptake, microinjection, *Agrobacterium*-mediated gene transfer, and particle bombardment, also known as biolistic transformation. In most cases, only a few cells that can be regenerated are transformed, requiring the elimination of non-transformed cells using selectable markers to identify those that have undergone transformation. It is also important to screen progenies of regenerated plants to verify that the transgene is integrated into the genome and is genetically transmissible. The use of morphogenic regulators, such as *BABY BOOM* (*Bbm*) and *WUSCHEL2* (*Wus2*), enhances cell competence and regeneration, thereby broadening the range of genotypes amenable to transformation. Additionally, optimizing culture conditions, hormone conditions, and wound responses is often necessary to maximize transformation efficiency.

References

Abbas, M. S. T. (2018). Genetically engineered (modified) crops (Bacillus thuringiensis crops) and the world controversy on their safety. *Egyptian Journal of Biological Pest Control*, *28* (1), 52. https://doi.org/10.1186/s41938-018-0051-2

Aliu, E., Ji, Q., Wlazlo, A., Grosic, S., Azanu, M. K., Wang, K., & Lee, K. (2024). Enhancing *Agrobacterium*-mediated plant transformation efficiency through improved ternary vector systems and auxotrophic strains. *Frontiers in Plant Science*, *15*, 1429353. https://doi.org/10.3389/fpls.2024.1429353

Azanu, M. K., Kang, M., Lee, K., & Wang, K. (2025). *Agrobacterium*-Mediated Transformation of Tropical Maize Using Seedling Leaf Whorl Explants. *Cold Spring Harbor Protocols*. https://doi.org/10.1101/pdb.prot108596

Bagri, D. S., Upadhyaya, D. C., Kumar, A., & Upadhyaya, C. P. (2018). Overexpression of PDX-II gene in potato (Solanum tuberosum L.) leads to the enhanced accumulation of vitamin B6 in tuber tissues and tolerance to abiotic stresses. *Plant Science*, *272*, 267–275.

Bevan, M. W., Flavell, R. B., & Chilton, M.-D. (1983). A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature*, *304* (5922), 184–187. https://doi.org/10.1038/304184a0

Beyene, G., Solomon, F. R., Chauhan, R. D., Gaitán-Solis, E., Narayanan, N., Gehan, J., Siritunga, D., Stevens, R. L., Jifon, J., Van Eck, J., Linsler, E., Gehan, M., Ilyas, M., Fregene, M., Sayre, R. T., Anderson, P., Taylor, N. J., & Cahoon, E. B. (2018). Provitamin A biofortification of cassava enhances shelf life but reduces dry matter content of storage roots due to altered carbon partitioning into starch. *Plant Biotechnology Journal*, *16* (6), 1186–1200. https://doi.org/10.1111/pbi.12862

Brookes, G. (2022). Genetically modified (GM) crop use 1996–2020: Environmental impacts associated with pesticide use change. *GM Crops & Food*, *13* (1), 262–289.

Bubolz, J., Sleboda, P., Lehrman, A., Hansson, S.-O., Johan Lagerkvist, C., Andersson, B., Lenman, M., Resjö, S., Ghislain, M., Zahid, M. A., Kieu, N. P., & Andreasson, E. (2022). Genetically modified (GM) late blight-resistant potato and consumer attitudes before and after a field visit. *GM Crops & Food*, *13* (1), 290–298. https://doi.org/10.1080/21645698.2022.2133396

Buhr, T., Sato, S., Ebrahim, F., Xing, A., Zhou, Y., Mathiesen, M., Schweiger, B., Kinney, A., & Staswick, P. (2002). Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. *The Plant Journal: For Cell and Molecular Biology*, *30* (2), 155–163. https://doi.org/10.1046/j.1365-313x.2002.01283.x

Casadaban, M. J., & Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. *Journal of Molecular Biology*, *138* (2), 179–207. https://doi.org/10.1016/0022-2836(80)90283-1

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., & Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science*, *263* (5148), 802–805. https://doi.org/10.1126/science.8303295

Chen, S., Hajirezaei, M. R., ZANOR, M.-I., Hornyik, C., Debast, S., Lacomme, C., Fernie, A. R., Sonnewald,

U., & Boernke, F. (2008). RNA interference-mediated repression of sucrose-phosphatase in transgenic potato tubers (Solanum tuberosum) strongly affects the hexose-to-sucrose ratio upon cold storage with only minor effects on total soluble carbohydrate accumulation. *Plant, Cell & Environment*, *31* (1), 165–176.
Collier, R., Fuchs, B., Walter, N., Kevin Lutke, W., & Taylor, C. G. (2005). Ex vitro composite plants: An inexpensive, rapid method for root biology. *The Plant Journal*, *43* (3), 449–457.
Constable, G., Llewellyn, D., Walford, S. A., & Clement, J. D. (2015). Cotton Breeding for Fiber Quality Improvement. In V. M. V. Cruz & D. A. Dierig (Eds.), *Industrial Crops: Breeding for BioEnergy and Bioproducts* (pp. 191–232). Springer New York. https://doi.org/10.1007/978-1-4939-1447-0_10
Edge, M., Oikeh, S. O., Kyetere, D., Mugo, S., & Mashingaidze, K. (2017). Water efficient maize for Africa: A public-private partnership in technology transfer to smallholder farmers in sub-Saharan Africa. In *From agriscience to agribusiness: Theories, policies and practices in technology transfer and commercialization* (pp. 391–412). Springer.
Feng, X., Liu, W., Cao, F., Wang, Y., Zhang, G., Chen, Z.-H., & Wu, F. (2020). Overexpression of HvAKT1 improves drought tolerance in barley by regulating root ion homeostasis and ROS and NO signaling. *Journal of Experimental Botany*, *71* (20), 6587–6600. https://doi.org/10.1093/jxb/eraa354
Ferguson, J. N., Schmuker, P., Dmitrieva, A., Quach, T., Zhang, T., Ge, Z., Nersesian, N., Sato, S. J., Clemente, T. E., & Leakey, A. D. B. (2024). Reducing stomatal density by expression of a synthetic epidermal patterning factor increases leaf intrinsic water use efficiency and reduces plant water use in a C4 crop. *Journal of Experimental Botany*, *75* (21), 6823–6836. https://doi.org/10.1093/jxb/erae289
Fiyaz, R. A., Shivani, D., Chaithanya, K., Mounika, K., Chiranjeevi, M., Laha, G. S., Viraktamath, B. C., Rao, L. V. S., & Sundaram, R. M. (2022). Genetic Improvement of Rice for Bacterial Blight Resistance: Present Status and Future Prospects. *Rice Science*, *29* (2), 118–132. https://doi.org/10.1016/j.rsci.2021.08.002
Gautheret, R. (1939). Sur la possibilité de réaliser la culture indéfinie des tissus de tubercules de carotte. *CR Hebd. Seances Acad. Sc*, *208*, 118–120.
Gelvin, S. B. (2003). *Agrobacterium*-mediated plant transformation: The biology behind the "gene-jockeying" tool. *Microbiology and Molecular Biology Reviews*, *67* (1), 16–37.
Gelvin, S. B. (2017). Integration of *Agrobacterium* T-DNA into the plant genome. *Annual Review of Genetics*, *51*, 195–217.
Gierczik, K., Székely, A., Ahres, M., Marozsán-Tóth, Z., Vashegyi, I., Harwood, W., Tóth, B., Galiba, G., Soltész, A., & Vágújfalvi, A. (2019). Overexpression of Two Upstream Phospholipid Signaling Genes Improves Cold Stress Response and Hypoxia Tolerance, but Leads to Developmental Abnormalities in Barley. *Plant Molecular Biology Reporter*, *37* (4), 314–326. https://doi.org/10.1007/s11105-019-01154-5
Green, C., & Phillips, R. (1975). Plant regeneration from tissue cultures of maize 1. *Crop Science*, *15* (3), 417–421.
Gupta, A., Pal, R. K., & Rajam, M. V. (2013). Delayed ripening and improved fruit processing quality in tomato by RNAi-mediated silencing of three homologs of 1-aminopropane-1-carboxylate synthase gene. *Journal of Plant Physiology*, *170* (11), 987–995. https://doi.org/10.1016/j.jplph.2013.02.003

Harding, R., Paul, J.-Y., James, A., Smith, M., Kleidon, J., Shekhawat, U., Phillips, A., Kidanemariam, D., Dawit, A., & Dale, J. (2025). QCAV-4, the first genetically modified Cavendish (cv. Grand Nain) banana resistant to Fusarium wilt tropical race 4 approved for commercial production and consumption. *Plant Biotechnology Journal*, *23* (9), 3628–3637. https://doi.org/10.1111/pbi.70178

He, Y., Zhang, T., Sun, H., Zhan, H., & Zhao, Y. (2020). A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research*, *7* (1), 152. https://doi.org/10.1038/s41438-020-00390-1

James, C. (2007). *Global status of commercialized biotech/GM crops, 2007* (Vol. 37). isaaa Ithaca, NY.

Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO Journal*, *6* (13), 3901–3907. https://doi.org/10.1002/j.1460-2075.1987.tb02730.x

Kang, M., Lee, K., Finley, T., Chappell, H., Veena, V., & Wang, K. (2022). An Improved *Agrobacterium*-Mediated Transformation and Genome-Editing Method for Maize Inbred B104 Using a Ternary Vector System and Immature Embryos. *Frontiers in Plant Science*, *13*, 860971. https://doi.org/10.3389/fpls.2022.860971

Kaniganti, S., Palakolanu, S. R., Thiombiano, B., Damarasingh, J., Bommineni, P. R., Che, P., Sharma, K. K., Jones, T., Bouwmeester, H., & Bhatnagar-Mathur, P. (2025). Developing Striga resistance in sorghum by modulating host cues through CRISPR/Cas9 gene editing. *Plant Cell Reports*, *44* (4), 90. https://doi.org/10.1007/s00299-025-03474-1

Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C. D., Nontachaiyapoom, S., Kinkema, M., & Gresshoff, P. M. (2007). *Agrobacterium* rhizogenes-mediated transformation of soybean to study root biology. *Nature Protocols*, *2* (4), 948–952.

Kishore, G. M., Padgette, S. R., & Fraley, R. T. (1992). History of herbicide-tolerant crops, methods of development and current state of the art–emphasis on glyphosate tolerance. *Weed Technology*, *6* (3), 626–634. https://doi.org/10.1017/S0890037X00035934

Klein, T. M., Wolf, E. D., Wu, R., & Sanford, J. C. (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature*, *327* (6117), 70–73.

Lee, K., Kang, M., Ji, Q., Grosic, S., & Wang, K. (2023). New T-DNA binary vectors with NptII selection and RUBY reporter for efficient maize transformation and targeted mutagenesis. *Plant Physiology*, *192* (4), 2598–2603. https://doi.org/10.1093/plphys/kiad231

Lee, K., & Wang, K. (2023). Strategies for genotype-flexible plant transformation. *Current Opinion in Biotechnology*, *79*, 102848. https://doi.org/10.1016/j.copbio.2022.102848

Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., Luo, Y., & Zhu, H. (2018). Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Frontiers in Plant Science*, *9*, 559. https://doi.org/10.3389/fpls.2018.00559

Lin, Z. J. D., Taylor, N. J., & Bart, R. (2019). Engineering Disease-Resistant Cassava. *Cold Spring Harbor Perspectives in Biology*, *11* (11). https://doi.org/10.1101/cshperspect.a034595

Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M.-J., Scelonge, C., Lenderts, B., Chamberlin,

M., Cushatt, J., Wang, L., Ryan, L., Khan, T., Chow-Yiu, J., Hua, W., Yu, M., Banh, J., Bao, Z., Brink, K.,… Gordon-Kamm, W. (2016). Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation. *The Plant Cell*, *28* (9), 1998–2015. https://doi.org/10.1105/tpc.16.00124

Lundry, D. R., Ridley, W. P., Meyer, J. J., Riordan, S. G., Nemeth, M. A., Trujillo, W. A., Breeze, M. L., & Sorbet, R. (2008). Composition of grain, forage, and processed fractions from second-generation glyphosate-tolerant soybean, MON 89788, is equivalent to that of conventional soybean (Glycine max L.). *Journal of Agricultural and Food Chemistry*, *56* (12), 4611–4622.

Mao, H., Jiang, C., Tang, C., Nie, X., Du, L., Liu, Y., Cheng, P., Wu, Y., Liu, H., Kang, Z., & Wang, X. (2023). Wheat adaptation to environmental stresses under climate change: Molecular basis and genetic improvement. *Special Issue on Climate Change and Food Security: Plant Science Roles*, *16* (10), 1564–1589. https://doi.org/10.1016/j.molp.2023.09.001

Masters, A., Kang, M., McCaw, M., Zobrist, J. D., Gordon-Kamm, W., Jones, T., & Wang, K. (2020). *Agrobacterium*-Mediated Immature Embryo Transformation of Recalcitrant Maize Inbred Lines Using Morphogenic Genes. *JoVE*, *156*, e60782. https://doi.org/10.3791/60782

Mian, A., Oomen, R. J., Isayenkov, S., Sentenac, H., Maathuis, F. J., & Véry, A.-A. (2011). Over-expression of an Na+-and K+-permeable HKT transporter in barley improves salt tolerance. *The Plant Journal*, *68* (3), 468–479.

Nakka, S., Jugulam, M., Peterson, D., & Asif, M. (2019). Herbicide resistance: Development of wheat production systems and current status of resistant weeds in wheat cropping systems. *Breeding Wheat for the Global North: China, the USA and Canada*, *7* (6), 750–760. https://doi.org/10.1016/j.cj.2019.09.004

Ow, D. W., Wood, K. V., DeLuca, M., De Wet, J. R., Helinski, D. R., & Howell, S. H. (1986). Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science*, *234* (4778), 856–859. https://doi.org/10.1126/science.234.4778.856

Parrott, W., All, J., Adang, M., Bailey, M., Boerma, H., & Stewart Jr, C. (1994). Recovery and evaluation of soybean plants transgenic for a Bacillus thuringiensis var. Kurstaki insecticidal gene. *In Vitro Cellular & Developmental Biology-Plant*, *30* (3), 144–149.

Paul, J.-Y., Khanna, H., Kleidon, J., Hoang, P., Geijskes, J., Daniells, J., Zaplin, E., Rosenberg, Y., James, A., Mlalazi, B., Deo, P., Arinaitwe, G., Namanya, P., Becker, D., Tindamanyire, J., Tushemereirwe, W., Harding, R., & Dale, J. (2017). Golden bananas in the field: Elevated fruit pro-vitamin A from the expression of a single banana transgene. *Plant Biotechnology Journal*, *15* (4), 520–532. https://doi.org/10.1111/pbi.12650

Potrykus, I. (2001). Golden rice and beyond. *Plant Physiology*, *125* (3), 1157–1161. https://doi.org/10.1104/pp.125.3.1157

Prías-Blanco, M., Chappell, T. M., Freed, E. F., Illa-Berenguer, E., Eckert, C. A., & Parrott, W. A. (2022). An *Agrobacterium* strain auxotrophic for methionine is useful for switchgrass transformation. *Transgenic Research*, *31* (6), 661–676. https://doi.org/10.1007/s11248-022-00328-4

Raj, S., Singh, R., Pandey, S., & Singh, B. (2005). *Agrobacterium*-mediated tomato transformation and regeneration of transgenic lines expressing Tomato leaf curl virus coat protein gene for resistance

against TLCV infection. *Current Science*, 1674–1679.
Ravikumar, G., Manimaran, P., Voleti, S. R., Subrahmanyam, D., Sundaram, R. M., Bansal, K. C., Viraktamath, B. C., & Balachandran, S. M. (2014). Stress-inducible expression of AtDREB1A transcription factor greatly improves drought stress tolerance in transgenic indica rice. *Transgenic Research*, *23* (3), 421–439. https://doi.org/10.1007/s11248-013-9776-6
Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., & Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. Red fluorescent protein. *Nature Biotechnology*, *22* (12), 1567–1572. https://doi.org/10.1038/nbt1037
Steward, F., Mapes, M. O., & Mears, K. (1958). Growth and organized development of cultured cells. II. Organization in cultures grown from freely suspended cells. *American Journal of Botany*, 705–708.
Straus, J., & LaRue, C. D. (1954). Maize endosperm tissue grown in vitro I. Culture requirements. *American Journal of Botany*, *41* (8), 687–694.
Suza, W., Harris, R. S., & Lorence, A. (2008). Hairy roots: From high-value metabolite production to phytoremediation. *Electron J Integr Biosci*, *3* (1), 57–65.
Svab, Z., & Maliga, P. (1993). High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proceedings of the National Academy of Sciences of the United States of America*, *90* (3), 913–917. https://doi.org/10.1073/pnas.90.3.913
Thompson, C. J., Movva, N. R., Tizard, R., Crameri, R., Davies, J. E., Lauwereys, M., & Botterman, J. (1987). Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus. *The EMBO Journal*, *6* (9), 2519–2523. https://doi.org/10.1002/j.1460-2075.1987.tb02538.x
Thorpe, C., Luo, W., Ji, Q., Eggenberger, A. L., Chicowski, A. S., Xu, W., Sandhu, R., Lee, K., Whitham, S. A., Qi, Y., Wang, K., & Jiang, S. (2025). Enhancing biolistic plant transformation and genome editing with a flow guiding barrel. *Nature Communications*, *16* (1), 5624. https://doi.org/10.1038/s41467-025-60761-x
Tripathi, L., Atkinson, H., Roderick, H., Kubiriba, J., & Tripathi, J. N. (2017). Genetically engineered bananas resistant to Xanthomonas wilt disease and nematodes. *Food and Energy Security*, *6* (2), 37–47. https://doi.org/10.1002/fes3.101
Waldron, C., Murphy, E. B., Roberts, J. L., Gustafson, G. D., Armour, S. L., & Malcolm, S. K. (1985). Resistance to hygromycin B. *Plant Molecular Biology*, *5* (2), 103–108. https://doi.org/10.1007/BF00020092
Wang, N., Ryan, L., Sardesai, N., Wu, E., Lenderts, B., Lowe, K., Che, P., Anand, A., Worden, A., van Dyk, D., Barone, P., Svitashev, S., Jones, T., & Gordon-Kamm, W. (2023). Leaf transformation for efficient random integration and targeted genome modification in maize and sorghum. *Nature Plants*, *9* (2), 255–270. https://doi.org/10.1038/s41477-022-01338-0
Weir, R. T., & Dalzell, J. J. (2020). *Agrobacterium*: Soil Microbe, Plant Pathogen, and Natural Genetic Engineer. *Frontiers for Young Minds*, *8*, 64. https://doi.org/10.3389/frym.2020.00064
Winans, S. C. (1992). Two-way chemical signaling in *Agrobacterium*-plant interactions. *Microbiological Reviews*, *56* (1), 12–31.

CHAPTER 8.

GENE FUNCTION AND BIOENGINEERING

Faizo Kasule; Walter P. Suza; and Madan K. Bhattacharyya

Introduction

A major advantage of using plant transformation to modify plants is the ability to transfer a single well-defined gene, most likely without altering the functions of other genes in the recipient plant. In contrast, plant breeding involves the simultaneous transfer of thousands of genes, and the favorable ones must then be separated from the unfavorable ones by segregation among the progenies. Transformation also allows the use of genes from other organisms that could not be introduced naturally to a plant because of **crossing barriers** preventing interspecific hybridization. In *Plant Breeding Methods*, several examples are provided that involve the use of novel genes developed through genetic engineering to confer resistance against insects and herbicides in commercial transgenic crops. It was discussed in Chapter 7 that transformation is a valuable tool for plant breeding and biotechnology. This chapter discusses the history of genome editing, as well as the various types of genome editing technologies utilized in crop biotechnology, ranging from meganucleases to zinc finger proteins, and CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein) systems, as well as prime editing.

History of Genome Editing Technology

For half a century, gene editing technologies that specifically modify genomic sequences have revolutionized various aspects of plant sciences and crop biotechnology through the elucidation of gene functions, the understanding of complex traits, and enhanced crop improvement. This has progressed through ranks from early transgenesis to the current use of precise, programmable tools (Figure 1). During the 1970s and 1980s, significant breakthroughs occurred primarily in random DNA integration, while progress in homologous recombination was limited (Mansour et al., 1988). In the 1980s, meganucleases, also known as homing endonucleases, were discovered. These create double-stranded breaks (DSBs) at rare 14-40 bp recognition sites (Jacquier & Dujon, 1985). In the same era, zinc finger proteins were discovered in 1985 in the African clawed frog (J. Miller et al., 1985). These finger-like DNA-binding motifs played a vital role in gene regulation and were applied in gene manipulation studies.

Zinc finger nucleases (ZFNs) were created in 1996 by fusing two domains: a zinc-finger-based DNA recognition domain and a non-specific *Fok* I endonuclease domain to cleave the targeted DNA (Kim et

al., 1996a). ZFNs were widely used for gene editing in the late 1990s and a major part of the 2000s. In 2010, transcription activator-like effector nucleases (TALENs) were developed (Miller et al., 2011). Like ZFNs, TALENs utilize the *Fok 1* endonuclease domain to cut the targeted DNA, and this system offers advantages for designing the cleavage of desired target sequences. This advantage over ZFNs made it a method of choice over ZFNs for a short period, until around 2010-2011.

In 1987, clustered regularly interspaced short palindromic repeats (CRISPRs) were described in the *Escherichia coli* genome during an analysis of genes involved in phosphate metabolism (Ishino et al., 1987). In 2012, a revolutionary gene editing tool, CRISPR/Cas9, was discovered (Jinek et al., 2012). Unlike ZFNs and TALENs, which require complex protein engineering, CRISPR/Cas9 utilizes a 20-nucleotide guide RNA to target a CRISPR-associated (Cas) nuclease corresponding site in the genome, adjacent to a protospacer adjacent motif (PAM) with the NGG sequence, for generating mutations. This makes CRISPR/Cas9 a simple and highly efficient tool for genome editing applications, making it the most widely used genome engineering system. For this reason, Jennifer Doudna and Emmanuelle Charpentier were awarded the 2020 Nobel Prize in Chemistry for their discovery of CRISPR/Cas9 as a genome editing tool.

In 2016, base editors were created that utilize a deactivated or nickase Cas9 enzyme with a deaminase activity to introduce single-nucleotide polymorphisms (SNPs) by chemically altering the targeted DNA sequence without inducing double-stranded DNA breaks (DSBs) (Komor et al., 2016). Prime editing was unveiled in 2019, which enables precise small insertions, deletions, and base swaps without introducing DSBs (Anzalone et al., 2019). Prime editors also overcome the limitations of base editors, which are restricted to only four possible base substitutions.

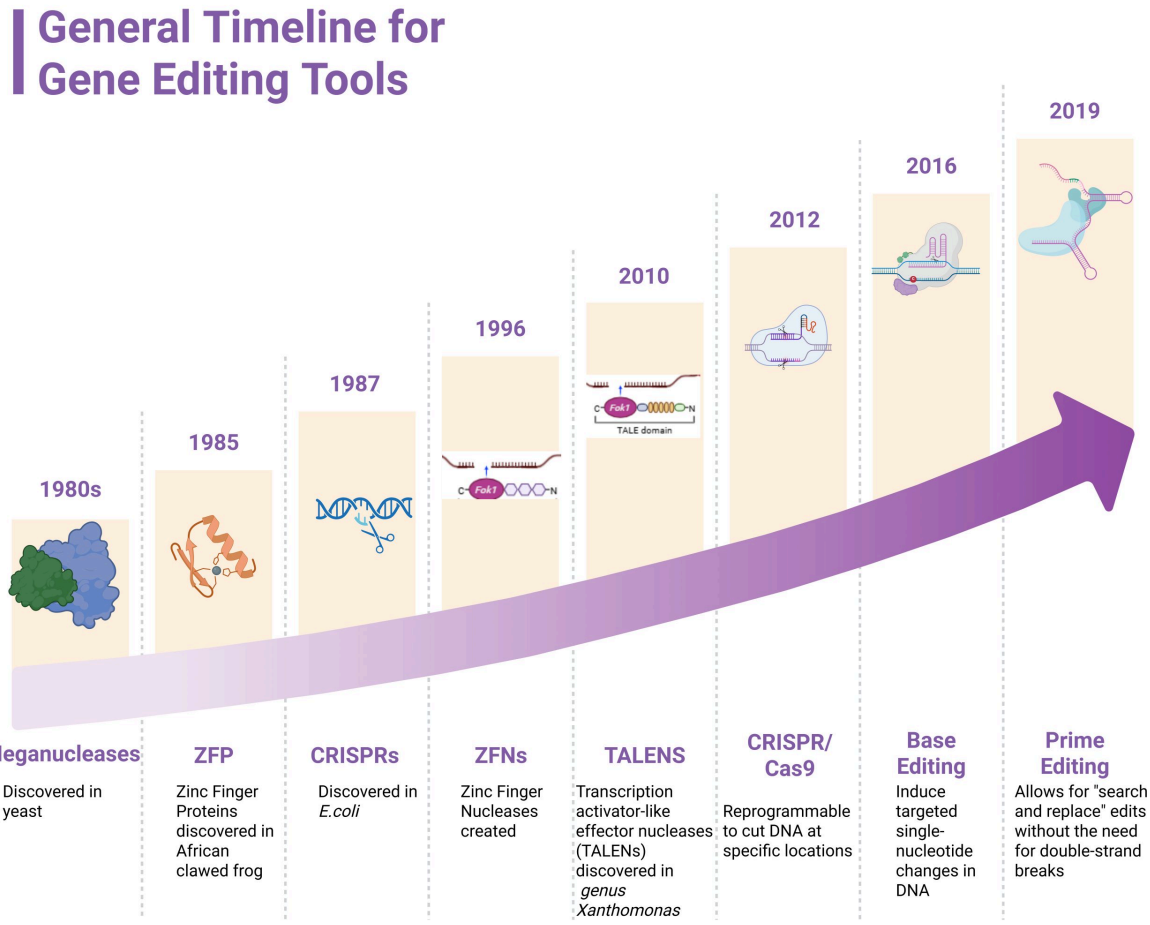


Figure 1 History of genome editing technologies. *Image created by Faizo Kasule*.

Types of Genome Editing Technologies in Crop Plants

Targeted gene editing is gaining popularity for crop improvement of traits for which contributing genes have been identified and characterized. There are two main bottlenecks in gene editing in crop plants: (i) lack of information about the genes that govern most traits; (ii) absence of stable transformation methods for some crop species. The main advantage of gene editing is the improvement of cultivars for one or more targeted genetic loci without changing the rest of the genome. In the backcrossing method, deleterious genes are often introduced along with the desirable trait loci. Such genes can reduce the performance of the cultivar and could result in yield reduction. This phenomenon is known as linkage drag. The precise genome editing systems enable us to replace an undesirable gene with a favorable allele, thereby avoiding linkage drag, as there is no need to introgress the favorable allele through backcrossing. In this section, we will discuss the following major genome editing technologies: (i) Zinc finger nucleases technology; (ii) TALEN technology; and (iii) CRISPR-Cas9 technology.

DNA Repair Mechanisms

Double-stranded DNA breaks (DSBs) are intentionally created during gene editing using primary genome editing tools, such as zinc finger nucleases, TALENs, and CRISPR-Cas9. DSBs are a form of DNA damage that occurs on both strands of the DNA double helix, where the DNA is cut at the same location or in its proximity. These breaks can be harmful and lead to genome instability, aberrant growth, or cell death. Therefore, the organism must employ its repair mechanisms to repair these DSBs. During the process of DNA repair, the DNA sequence at the site of the break may be altered. Many DNA repair mechanisms produce genome editing by joining the cut ends, which include Base Excision Repair, Nucleotide Excision Repair, Mismatch Repair, Homology-Directed Repair (HDR), Non-Homologous End Joining (NHEJ), and other minor repair mechanisms. By far, DSBs effect genome editing and undergo two major repair pathways in the cell: NHEJ and HDR (Figure 2).

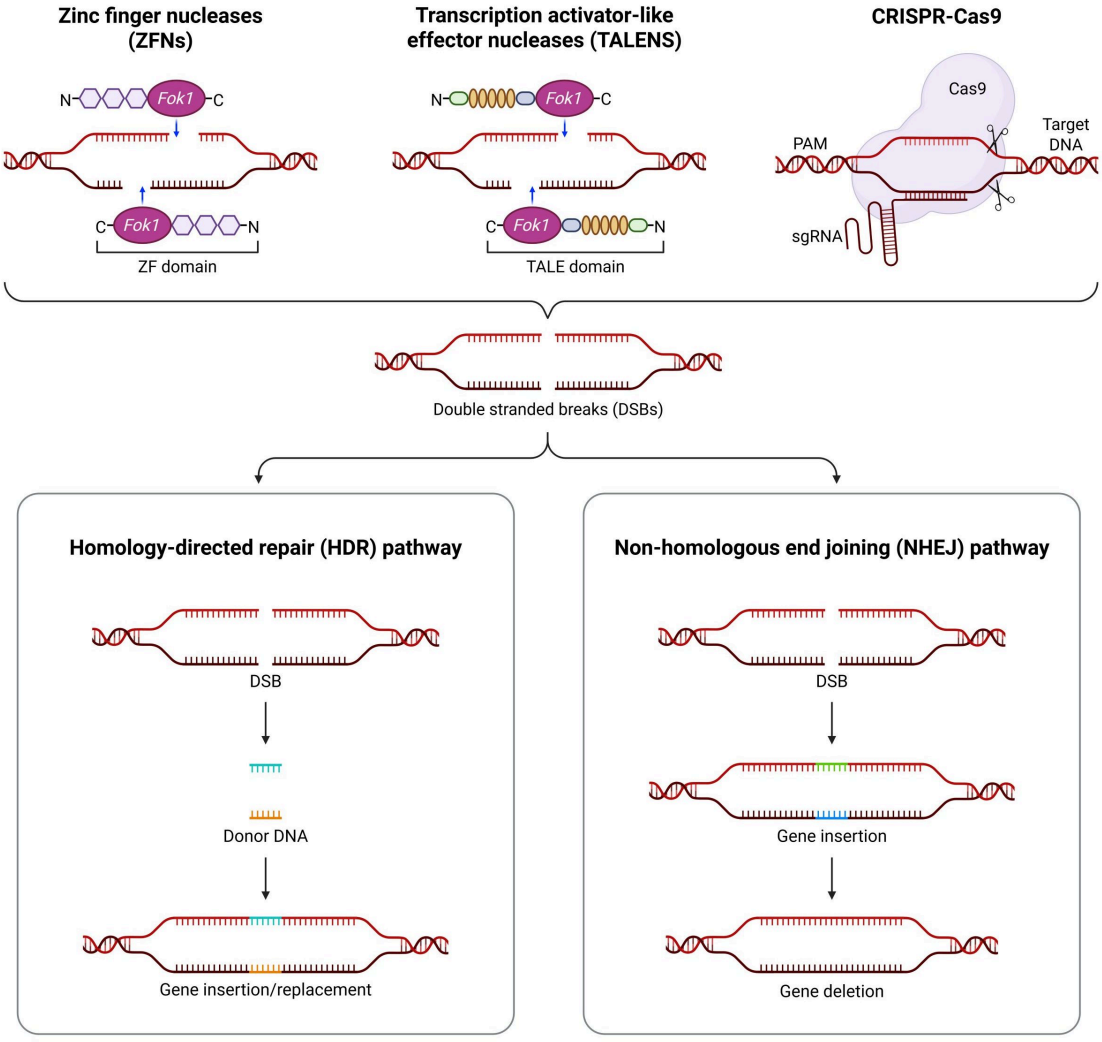


Figure 2 Generation of double-stranded breaks (DSBs) and the two main pathways for DNA DSBs repair: homologous recombination (HR) and non-homologous end joining (NHEJ). *Modified image of Wendy Jiang from BioRender. Adapted from Hussain et al. (2021).*

HDR is a precise DNA repair mechanism that utilizes a homologous donor template to repair DSBs (Hussain et al., 2021). In the native HDR pathway, a homologous sequence of sister chromatids is used as a repair template to repair DSB in the S and G2 phases of the cell cycle. In contrast, an external repair template must be provided during genome editing to introduce an insertion or replacement at the targeted site. Using a DNA template to guide the repair of DSBs makes HDR highly accurate and nearly error-free. Scientists leverage HDR to introduce specific and accurate gene modifications by designing a donor template wherein the DNA sequence intended for insertion is flanked by arms homologous to the 5' and 3' sites of the DSB (Puchta, 2005). However, HDR efficiency is low compared to other repair pathways in plants.

NHEJ is the predominant form of plant DNA repair mechanism to join broken pieces of DNA successfully (Puchta, 2005). NHEJ might lead, in a certain fraction of cases, to genomic changes such as deletions, insertions, or various kinds of genomic rearrangements (Puchta, 2005). However, NHEJ is error-prone, often leading to insertions and deletions (indels) at the break site, which can disrupt gene function or cause mutations. This is because the NHEJ process does not rely on a homologous template for repair.

Zinc Finger Nucleases Technology

Zinc finger nuclease (ZFN) (Kim et al., 1996b) was created by fusing two proteins: (i) a zinc finger protein that binds to DNA, e.g., a transcription factor; (ii) *Fok* I, a type IIS restriction enzyme that recognizes asymmetric DNA sequences and cleaves next to the binding site.

The Cys $_2$ His $_2$ zinc finger proteins used for designing ZFN contain (Tyr, Phe)-Xaa-Cys-Xaa $_{2\text{-}4}$ -Cys-Xaa $_3$ -Phe-Xaa5 – Leu-Xaa $_2$ -His-Xaa $_{3\text{-}5}$ -His, amino acids, and they usually occur in tandem. Each of the Cys $_2$ His $_2$ proteins interacts with a zinc ion to form the structure termed a zinc finger that binds to the DNA double helix in sequence sequence-specific manner (Kim et al., 1996b). As shown below, we can design a Zn finger protein (Figure 3) that binds to the target site, generating a double-stranded break (Osakabe et al., 2010).

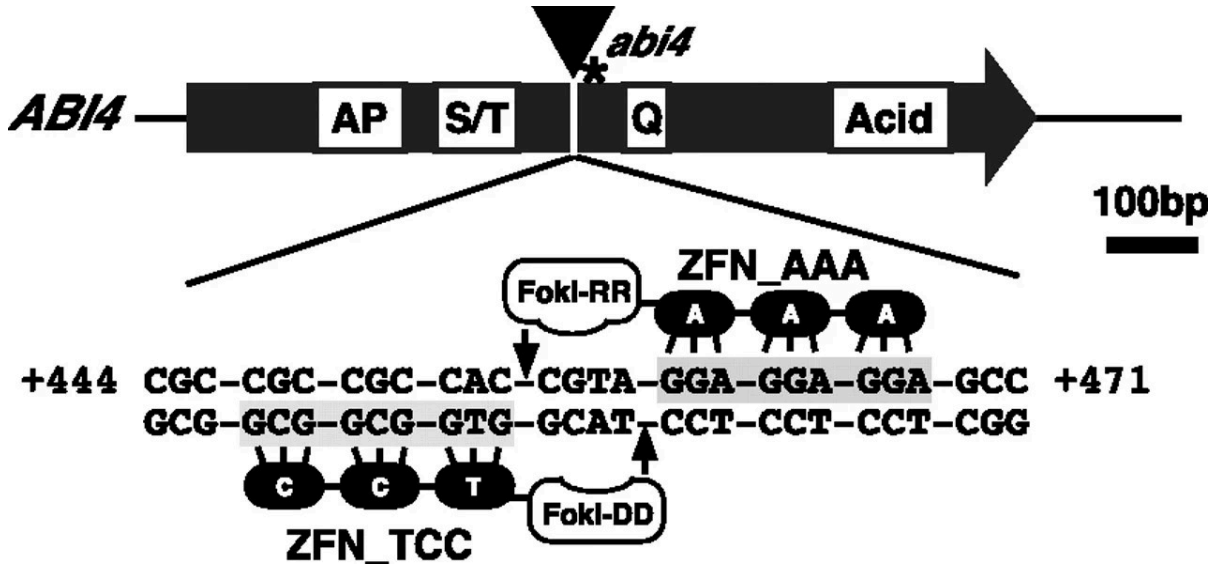


Figure 3 Consensus ZFN target sites in the Arabidopsis ABA-INSENSITIVE4 (ABI4) gene. The schematic representation of the ABI4 gene is shown at the top. An asterisk indicates the position of the mutation in the abi4 mutant. AP, AP2 domain; S/T, serine- and threonine-rich domain; Q, glutamine-rich domain; Acid, activation domain. Target sites of ZFN monomers are highlighted with gray bars. Arrows show the putative cleavage sites. FokI-DD, mutated Fok I nuclease domain (R487D); FokI-RR, mutated Fok I nuclease domain (D483R). *Adapted from Osakabe et al. 2010*.

The outcomes of a double-stranded break (DBS) caused by ZFN can be repaired by two mechanisms: (1) via non-homologous end joining (NHEJ) and (2) homology-directed repair (HDR), as shown below (Urnov et al., 2010). NHEJ generates knockout mutations while HDR can lead to homologous recombination and allelic substitution (Figure 4). We can replace a bad gene with a good one.

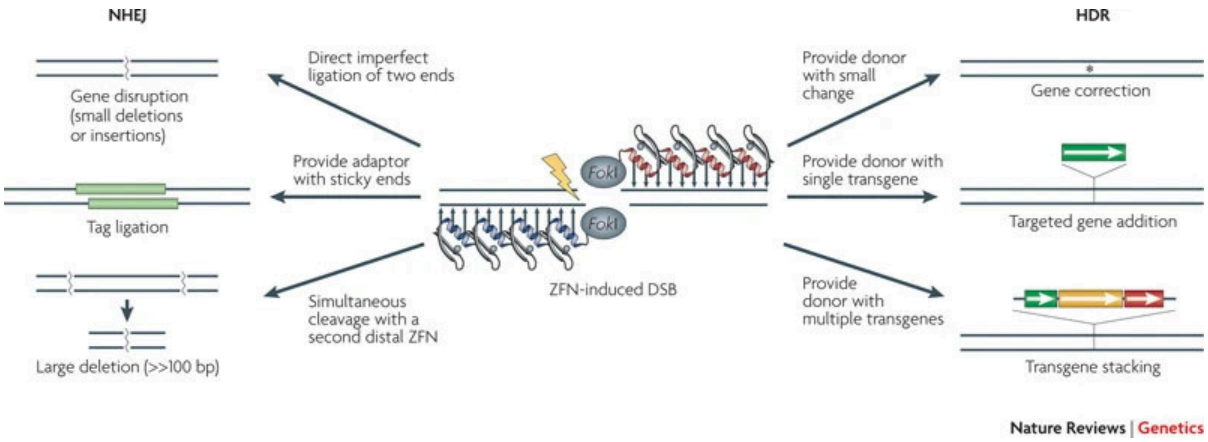


Figure 4 A zinc finger nuclease (ZFN)-induced double-strand break (DSB) allows a range of alleles to be generated at endogenous loci, as specified by the investigator. The diagram shows the different outcomes that can result from the introduction of a site-specific DNA break. A ZFN pair is shown bound to a genomic target site (the two different DNA-binding domains are shown in red and blue). The DSB generated by ZFN cleavage induces DNA repair processes that may be influenced by the addition of an investigator-designed donor DNA. *Adapted from Urnov et al. (2010)*.

TALEN Technology

TALEN (**Transcription activator-like effector nuclease)** is similar to ZFN. Concepts used in TALEN and ZFN are the same. In TALEN (Figure 5), ZF is substituted with a synthetic protein designed based on the DNA-binding properties of a class of plant pathogen effector proteins known as transcription activator-like (TAL) effector proteins (Bogdanove & Voytas, 2011). TAL proteins of a bacterial rice pathogen of the genus *Xanthomonas* are targeted to the nucleus of rice plants to reprogram transcription of specific genes for the pathogen's benefit. TAL factors carry tandem, polymorphic amino acid repeats that individually specify contiguous nucleotides in DNA as targets for binding. One can generate synthetic TAL factors for sequences of the target genes based on the codes learned from the investigation of TAL factors. As shown below, the FokI nuclease is added or fused to designed TAL factors to generate the TALEN molecule for mutating a gene.

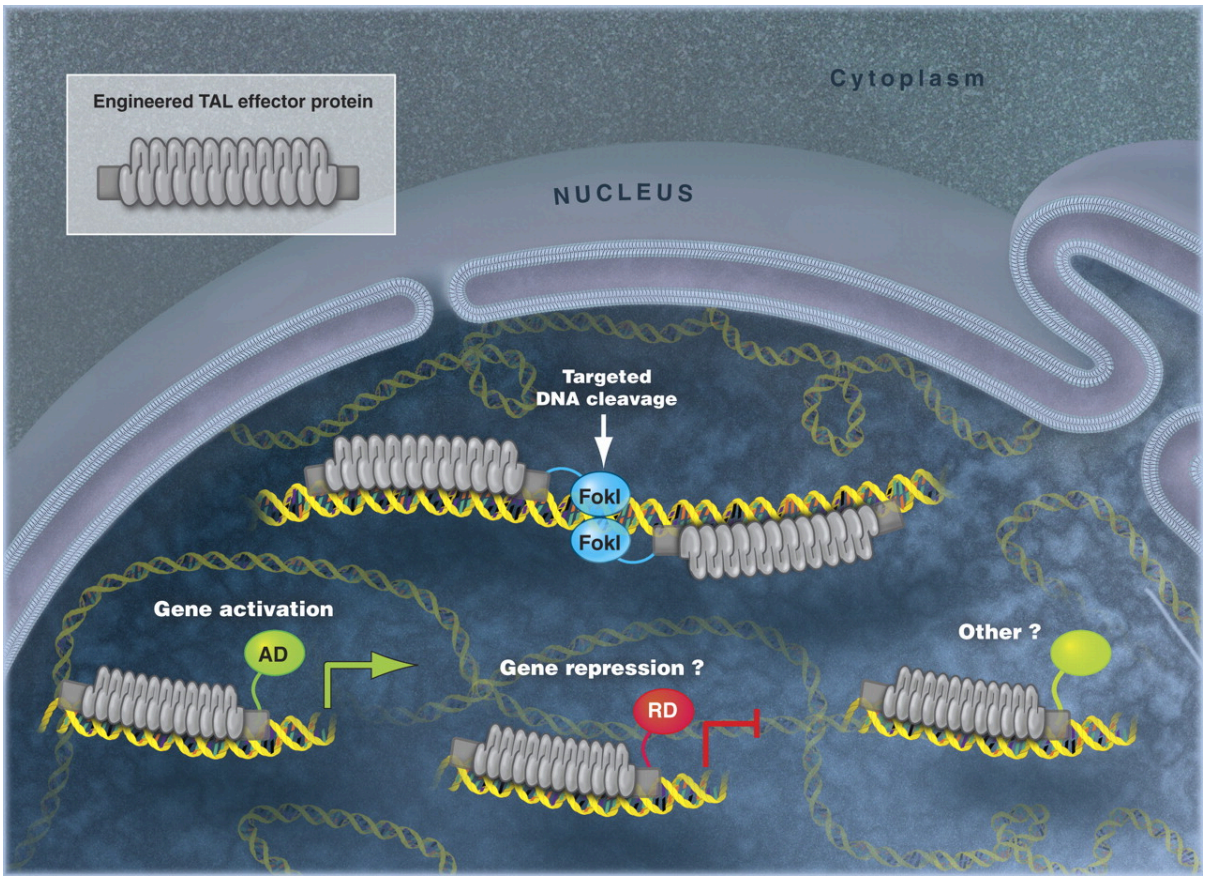


Figure 5 Genomic control enabled by engineered TAL effector proteins. Fusion of TAL effector proteins to FokI creates sequence-specific nucleases that enable targeted DNA cleavage for gene knockouts and genome editing. TAL effector proteins fused to transcriptional activation domains (AD) and putatively to repression domains (RD) provide artificial switches for gene regulation in vivo. TAL effector-based sequence-specific mutagens or chromatin-modifying proteins created by fusing TAL effectors to domains such as cytidine deaminases, histone acetyltransferases or deacetylases, or DNA methyltransferases can also be envisioned. *Adapted from Bogdanove and Voytas (2011)*.

As shown above, fusion of TAL effector proteins to *Fok* I creates sequence-specific nucleases that enable targeted DNA cleavage for gene knockouts and genome editing. TALEN technology has been widely utilized for targeted gene mutagenesis, particularly for gene inactivation, in various organisms, including agriculturally important plants such as rice, wheat, tomato, and barley. TALENs were employed to generate stable, heritable mutations at the maize *glossy2* (*gl2*) locus (Char et al., 2015). Transgenic lines containing mono- or di-allelic mutations were obtained from the maize genotype Hi-II at a frequency of about 10% (nine mutated events in 91 transgenic events) (Figure 6). In addition, three of the novel alleles were tested for function in progeny seedlings, where they were able to confer the glossy phenotype. Char et al. (2015) demonstrated that TALENs are an effective tool for genome mutagenesis in maize, empowering the discovery of gene function and the development of trait improvement.

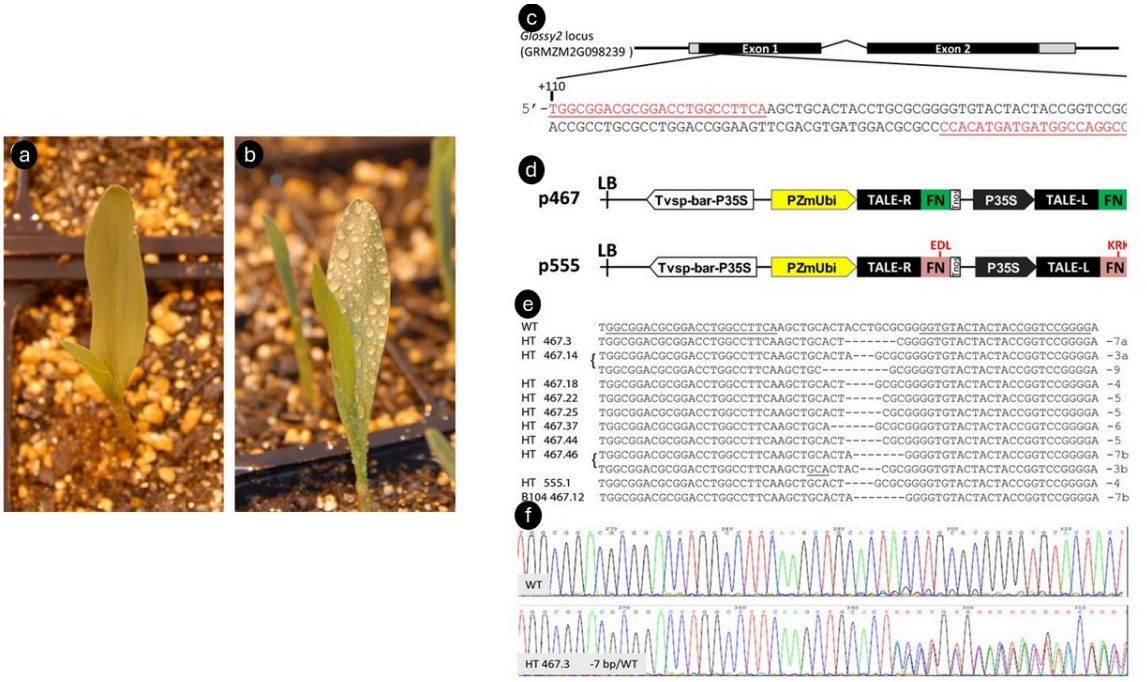


Figure 6 Glossy phenotype caused by TALEN-induced gl2 mutation. Glossy phenotype in maize homozygous seedlings for a TALEN-induced 7-bp deletion in the gl2 gene. (a) wild-type plants compared to the gl2 mutant leaf (b) due to reduced epicuticular wax caused by the loss of function of the gl2 gene. Gene structure of glossy2 containing the TALEN target sequences, (c) The target site is located between 110 and 176 bp downstream of the translation initiation site of glossy2, and (d) Schematics of T-DNA construct for gene editing containing the selection marker bar gene under control of the CaMV 35S promoter, TALEN gene (TALEN-R) under control of the maize ubiquitin 1 promoter, and the CaMV 35S promoter linked with another TALEN gene (TALEN-L). (e) TALEN-induced deletion mutations at the gl2 target site. Genomic DNA sequences at the relevant regions from each mutant are aligned with the corresponding wild-type sequence. (f) Example of chromatographs showing the sequencing traces of two alleles, the wild type (WT) and 7-bp deletion (−7) at the gl2 locus from a callus line. *Adapted from (Char et al., 2015)*.

CRISPR-Cas9 Technology

The **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (CRISPR) was first discovered in 1987 (as shown in Figure 1). The significance of CRISPR was unknown until 2007, when its immunity function

was discovered. The mechanism used by the element to confer immunity became known only after another five years (2012), when a detailed understanding of this adaptive defense system was achieved (Jinek et al., 2012).

The bacteria and archaea use CRISPR to defend against the invading viruses, termed as bacteriophages (Figure 7). Following viral infection, bacteria employ a special CRISPR-associated nuclease 9 (Cas9) to generate a double-strand break (DSB) in the target bacteriophage DNA molecules. Thus, viruses become ineffective (inactive). Cas9 is guided to the target sequence by a short RNA fragment known as a guide RNA (gRNA), which is complementary to a viral genome segment, thereby generating a DSB (Le Rhun et al., 2019).

Parallel to viral DNA cleavage, a short viral DNA fragment is captured and integrated into the host genome at one end of the CRISPR locus, at the end of the leader end, to form what is known as a "spacer". The spacers flanked by short palindromic repeats (20–50 bp) serve as a genetic memory for past infections and form a growing library of immunity. Therefore, the sequence is used to generate the gRNA, which rapidly activates this defense mechanism against any subsequent infection by the same bacteriophage (Figure 7).

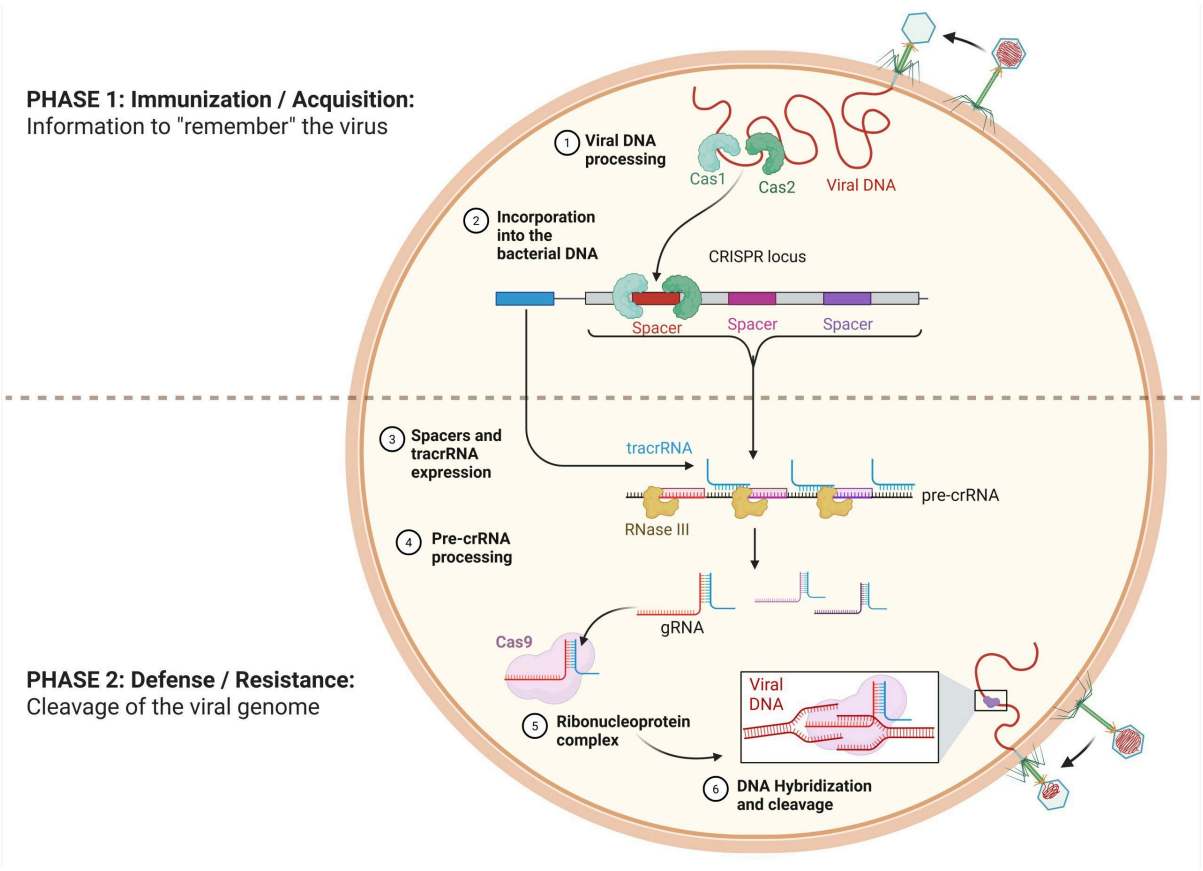


Figure 7 CRISPR-Cas9 Adaptive Immune System of Streptococcus pyogenes Against Bacteriophages. (Phase1) Immunization process: After insertion of exogenous DNA from viruses or plasmids, a Cas complex recognizes foreign DNA and integrates a novel repeat-spacer unit at the leader end of the CRISPR locus. (Phase 2) Immunity process: The CRISPR repeat-spacer array is transcribed into a pre-crRNA that is processed into mature crRNAs, which are subsequently used as a guide by a Cas complex to interfere with the corresponding invading nucleic acid. Adapted from BioRender, created by Sebastián Felipe González Moraga.

CRISPRs are a diverse family of DNA repeats with a common architecture. Each CRISPR locus consists of a series of short repeat sequences (20–50 bp) separated by unique spacer sequences of a similar length. The number of repeat sequences varies; some bacteria may have up to 100 repeat segments. Spacers are highly variable and derived from **foreign DNA.**

Invading DNA is fragmented and is selected for integration based on the PAM (protospacer adjacent motif) sequence. The CRISPR locus acquires new DNA at the Leader end of the locus and provides an immunity "library" of foreign DNAs. The CRISPR-Cas immune response, a bacterial defense mechanism, operates in three stages: Acquisition, expression, and Interference.

In ***the acquisition phase,*** when a virus infects a bacterium, bacterial *Cas1* and *Cas2* proteins recognize foreign phage DNA containing protospacers adjacent to PAM sequences and insert them into the CRISPR array as spacers. In the ***expression*** phase, the CRISPR array is transcribed into a pre-crRNA, which is then processed into individual CRISPR RNAs (crRNAs) by RNase III, assisted by tracrRNA in type

II systems. Each crRNA carries a unique spacer and guides the Cas nuclease. crRNAs pair with tracrRNA or exist as synthetic single guide RNAs (sgRNAs) (Figure 7). During the **interference phase**, the crRNA-Cas complex (crRNP) scans for matching sequences in the incoming foreign DNA and directs Cas9 to cleave and degrade the target phage DNA, thereby establishing immunity against those viruses.

This system is similar to antibody production in humans. Once an antibody is produced against a pathogen following infection or exposure to a vaccine, the information is retained in the cells' memory, allowing the rapid production of the antibody against the pathogen in subsequent infections.

CRISPR-Cas9, a tool for genome editing, was discovered by Dr. Jennifer Doudna and Dr. Emmanuelle Charpentier (Figure 8).

Figure 8 Jennifer Doudna (left) and Emmanuelle Charpentier (right) were awarded the Nobel Prize in Chemistry in 2020 for their discovery and development of the CRISPR-Cas9 genetic editing tool. *Credit: Alexander Heinel/Picture Alliance/DPA*.

Two steps to consider in designing a CRISPR-Cas9 system for editing a gene are (Asmamaw & Zawdie, 2021): (i) the design of gRNA, and (ii) the nuclease to be used should be specific to the application (Figure 9).

The CRISPR-Cas9 Genetic Scissors

When researchers edit a genome using the CRISPR-Cas9 genetic scissors, they artificially construct a **single guide RNA (sgRNA)**, which matches the DNA code where the cut is to be made.

The scissor protein, **Cas9**, forms a complex with the **sgRNA**, and directs the scissors to the precise location in the genome where the cut will be made.

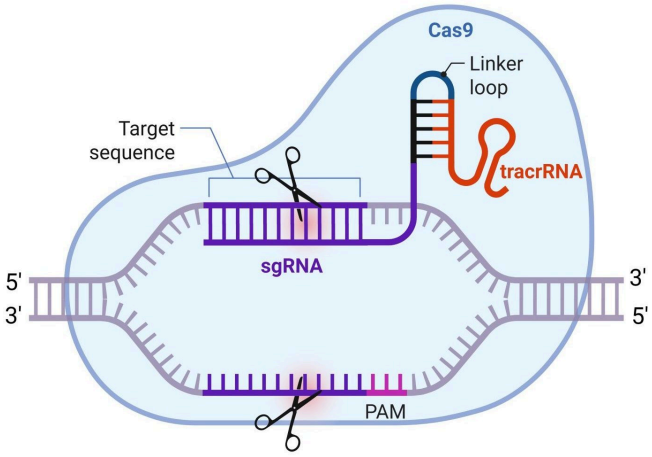


Figure 9 The CRISPR-Cas9 system. The CRISPR-Cas system comprises a guide RNA (gRNA) and Cas9 nuclease, which together form a ribonucleoprotein (RNP) complex. The presence of a specific protospacer adjacent motif (PAM) in the genomic DNA is required for the gRNA to bind to the target sequence. The Cas9 nuclease then makes a double-strand break (DSB) in the DNA (denoted by scissors). Endogenous repair mechanisms triggered by the DSB may result in gene knockout via a frameshift mutation or knock-in of a desired sequence. *Image modified by Faizo Kasule using BioRender, adapted from the template of Brady Cress (Asmamaw & Zawdie, 2021)*.

The gRNA that targets specifically the gene of interest should be designed. The ideal scenario would be zero non-specific cuts, with the gRNA being highly specific to the target site in an exon. You can edit members of a gene family simultaneously. Following DSB, cells repair the DNA using the non-homologous end-joining (NHEJ) mechanism (Figure 10). During this process, indels are created, leading to frameshift mutations and gene knockouts (KO) *(Addgene, 2017)*.

We can also conduct a knock-in (KI) mutation by repairing the DSB through the homology-directed repair (HDR) mechanism. To facilitate or induce HDR-mediated repair, copies of homologous DNA molecules with a desirable mutation are provided as a template. The mutation can be a single-point mutation to change an amino acid. It can correct disease-causing mutations in humans and generate herbicide resistance in plants. For example, we can replace an amino acid in a plant gene encoding acetolactate synthase (ALS), which is involved in the biosynthesis of branched-chain amino acids, to make the enzyme resistant to several herbicides, including imidazolinones and sulfonylureas. In KI experiments, we must carefully check the templates for PAM sequences to ensure that they are not treated as targets.

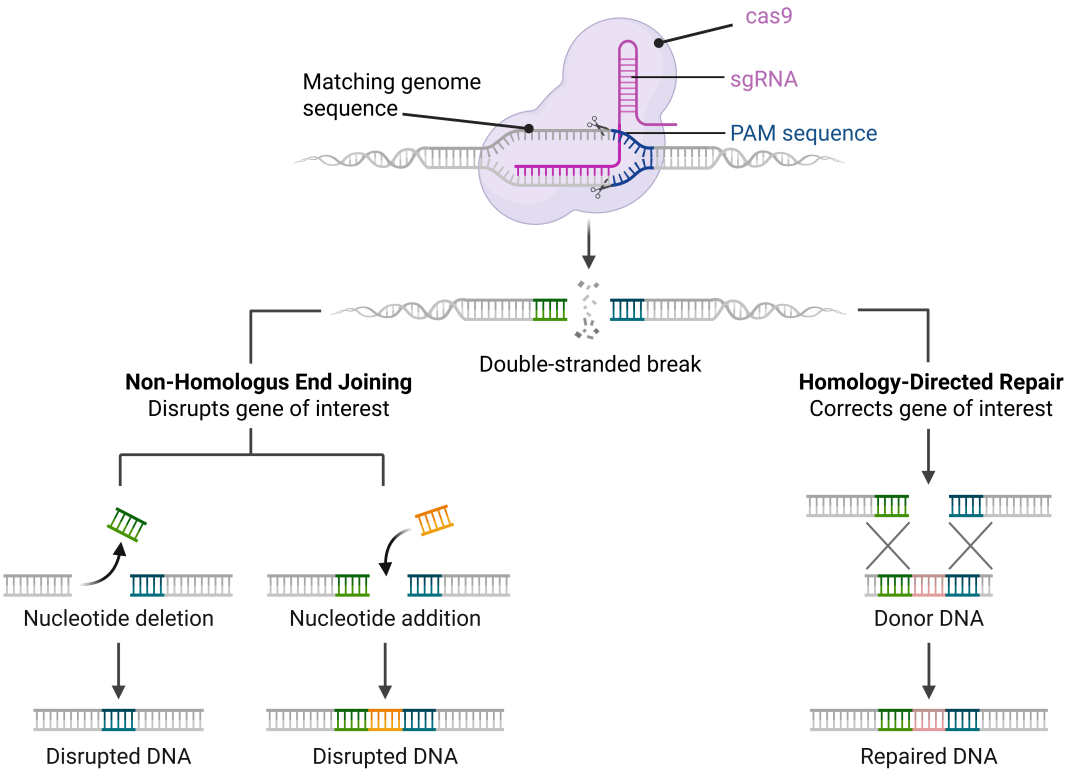


Figure 10 CRISPR/Cas9 for genome editing. The Cas9 complex, along with an sgRNA, recognizes a specific sequence, known as the protospacer. This is only possible if this sequence is followed by a Protospacer Adjacent Motif (PAM). When Cas9 binds, a dsDNA break is generated. Subsequently, non-homologous end joining, or homology-directed repair, can occur, leading to KO or KI mutations, respectively. *Image modified by Faizo Kasule using BioRender, adapted from the template of Esmée Dragt (Addgene, 2017)*.

Point or insertion-deletion mutations through CRISPR/Cas9 are predictable. The only weakness is the mutations that may be caused by the system in non-targeted genomic regions. Many sequences are similar to the 16 bp target sequence in a complex genome. To minimize this, careful analysis of the genome is essential to confirm that the target sequence is unique and not located within coding regions. One can always breed the desirable CRISPR/Cas9-induced mutations in crop plants through backcrossing.

Another way to eliminate non-specific mutation is to use gRNA and Cas9 protein in test tubes and apply the two as ribonucleoprotein (RNP) complexes to the cells, thereby generating a mutation. This approach has been reported to be an effective strategy (Liang et al., 2017). The complex is unstable and does not continue to function, thereby preventing additional off-target mutations from occurring. At the same time, the regenerated gene-edited lines are free from integration of any foreign DNA and selectable antibiotic resistance genes (Liang et al., 2017).

Strategies for delivering single guide RNA (sgRNA) and Cas9 ribonucleoprotein complex CRISPR/Cas9 genome editing have recently been reviewed by Zhang et al. (2021). Svitashev et al. (2016) demonstrated

that Cas9 and gRNA in RNP complexes can be delivered into maize embryo cells via particle bombardment. The resulting regenerated plants had specific targeted gene mutations and gene edits at high frequencies (Figure 11). They showed that this completely DNA- and selectable marker-free method generated mutant plants at high frequencies. Therefore, RNP delivery provides a precise **and efficient tool** for plant breeders to create improved crop varieties more quickly and reliably than traditional breeding methods.

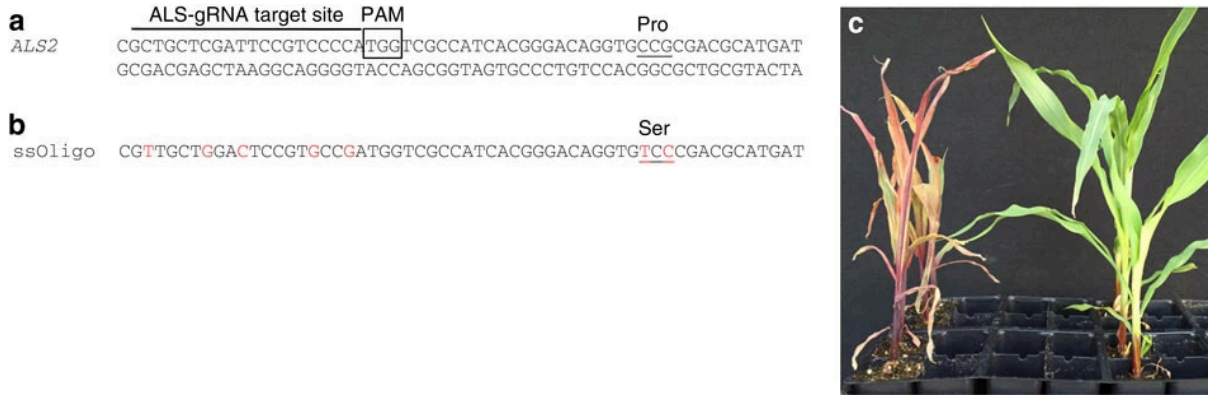
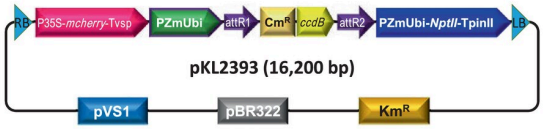


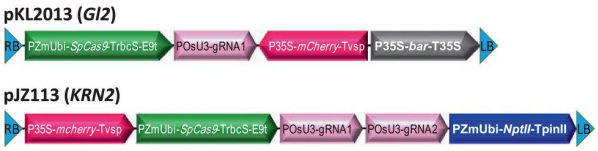
Figure 11 A) Partial ALS2 sequence with ALS–gRNA target site, protospacer adjacent motif (PAM) is boxed, and the proline codon (amino acid position 165) to be edited is underlined. **B**) Partial sequence of the ssDNA oligo used as a repair DNA template; modified nucleotides are shown in red font, and the serine codon is underlined. **C**) Wild type (left) and ALS2 edited (right) plants tested for resistance to chlorsulfuron (200 mg l^{-1}); shown at 10 days after spraying. Adopted from Svitashev et al. (2016).

CRISPR/Cas9 has been utilized to generate targeted gene knockouts in teosinte (*Zea mays ssp. parviglumis*) using a biolistic delivery method (Zobrist et al., 2023). Leaf-derived callus explants contained short indel mutations generated by a single guide RNA (sgRNA) and precise sequence deletion using two sgRNAs (Figure 12). Selectable markers, including mutant ALS and bialaphos resistance, enabled the recovery of edited callus with an average transformation frequency of 4.17%. Visual markers (tdTomato, GUS) confirmed expression in T1 and T2 progeny. Genes such as *Glossy2* (Gl2) and *Kernel Row Number2* (KRN2) were successfully edited, demonstrating that CRISPR/Cas9 can introduce heritable, beneficial traits. This approach allows plant breeders to harness teosinte's genetic diversity for yield improvement, trait discovery, and de novo domestication of maize.

(a) **New Gateway destination vector for CRISPR/Cas9**



(b) **Schematic illustration of T-DNA regions**



(c) **Sequence conservation in *Glossy2* target region**

Genome	Sequence
B73	CTTTGGTC--ACAGATCACAAACTTCAAATGCGGTGGGCTGGCGCTG
RS	CTTTGGTC--ACAGATCACAAACTTCAAATGCGGTGGGCTGGCGCTG
parvi TIL01	CTTTGGTC--ACAGATCACAAACTTCAAATGCGGTGGGCTGGCGCTG
parvi TIL11	CTTTGGTC--ACAGATCACAAACTTCAAATGCGGTGGGCTGGCGCTG
diploperennis	CTTTGGTC--ACAGATCACAAACTTCAAATGCGGTGGGCTTGCGCTG
luxurians	CTTTGGTC--ACAGATCACAAACTTCAAATGCGGTGGGCTGGCGCTG
mexicana TIL18	CTTTGGTCACACAGATCACAAACTTCAAATGCGGTGGGCTGGCGCTG
mexicana TIL25	CTTTGGTCACACAGATCACAAACTTCAAATGCGGTGGGCTGGCGCTG

*PAM (blue) and protospacer (red) sequences are highlighted.

(d) **Teosinte transformation**

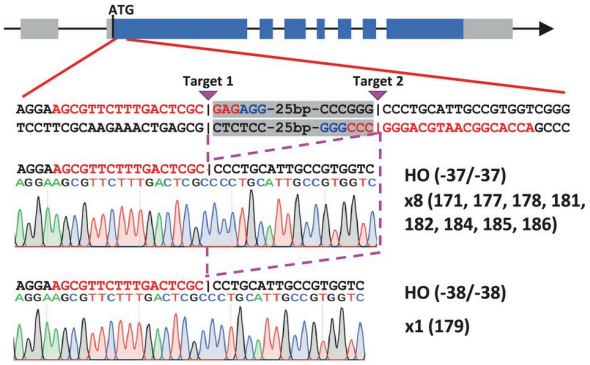
(e) **Targeted mutagenesis of *Glossy2***

T0 genotype	*Glossy2* sequence	Indel mutation	# T0 plants
WT	Allele 1: TGGTCACAGATCACAAACTTCAAATGCGGTGGGCTG	0 bp	0
	Allele 2: TGGTCACAGATCACAAACTTCAAATGCGGTGGGCTG	0 bp	
BI	Allele 1: TGGTCACAGATCACAAACTTCAA-TGCGGTGGGCTG	-1 bp	27
	Allele 2: TGGTCACAGATCACAAACTTCAAATGCGGTGGGCTG	+1 bp	
BI	Allele 1: TGGTCACAGATCACAAACTTCAAATGCGGTGGGCTG	+1 bp	10
	Allele 2: TGGTCACAGATCACAAACTTCAAAATGCGGTGGGCTG	+2 bp	

*PAM (blue) and protospacer (red) sequences are highlighted. Inserted bases (A or AA) are underlined.

(f) **Phenotype of *glossy2* knockout mutant**

(g) **Targeted mutagenesis of *KRN2***



*PAM (blue) and protospacer (red) sequences are highlighted.
**Double strand DNA break sites are indicated by "|" and purple triangle (▼).

Figure 12 CRISPR/Cas9-mediated targeted mutagenesis in teosinte. **(a)** Gateway destination vector pKL2393. **(b)** T-DNA regions of pKL2013 (Gl2) and pJZ113 (KRN2). **(c)** Sequence conservation in the Gl2 target region among Zea mays subspecies. B73, maize inbred B73; RS, parviglumis from Restoration Seeds (https://www.restorationseeds.com/products/teosinte); parvi, parviglumis; **(d)** teosinte transformation. (d-1 & d-2) Teosinte seeds before and after removal of fruit case. (d-3) Germinating seedlings. (d-4) Leaf-derived embryogenic callus. (d-5) Transgenic callus expressing red fluorescent protein (RFP, mCherry). (d-6) Regenerating transgenic shoots. (d-7) T1 seeds expressing RFP. **(e)** Summary of targeted mutagenesis of Gl2. **F)** Phenotype of the gl2 mutant on a young leaf. Scale bar = 1 cm. **(e)** Targeted mutagenesis of KRN2. Deleted sequences are shaded in grey. *Adopted from Zobrist et al. (2023)*.

Marker-free HDR-mediated insertion of the golden rice trait

Figure 13 Ye et al. (2000); transferred a plant phytoene synthase (psy) and a bacterial phytoene desaturase (crtI) to rice, producing beta-carotene, which is responsible for the yellow color in rice kernels. Yellow rice is healthy because it provides beta-carotene, the precursor of Vitamin A, the deficiency of which causes night blindness, a major problem among the children of underdeveloped countries of Asia and Africa.

In the original golden rice reported in 2000 by Ye et al. (2000) the phytoene synthase (*psy*) was from the daffodil flower. In a subsequent version, Paine et al. (2005) demonstrated that replacing the daffodil gene with the maize psy gene increased carotenoid content up to 23-fold (maximum 37 µg/g) compared to the original Golden Rice. The genes used by Paine et al. (2005) were used in the CRISPR-mediated generation of the marker-free golden rice by Dong et al. (2020), shown below. They delivered the gRNA-Cas9 construct and a selectable marker in a second plasmid, along with the plasmid carrying the carotenoid cassette that contained the two beta-carotenoid pathway enzyme genes. They targeted the gene cassette to a genomic safe harbor (where genes can be incorporated without causing any damage to the genome) (Figures 14 and 15).

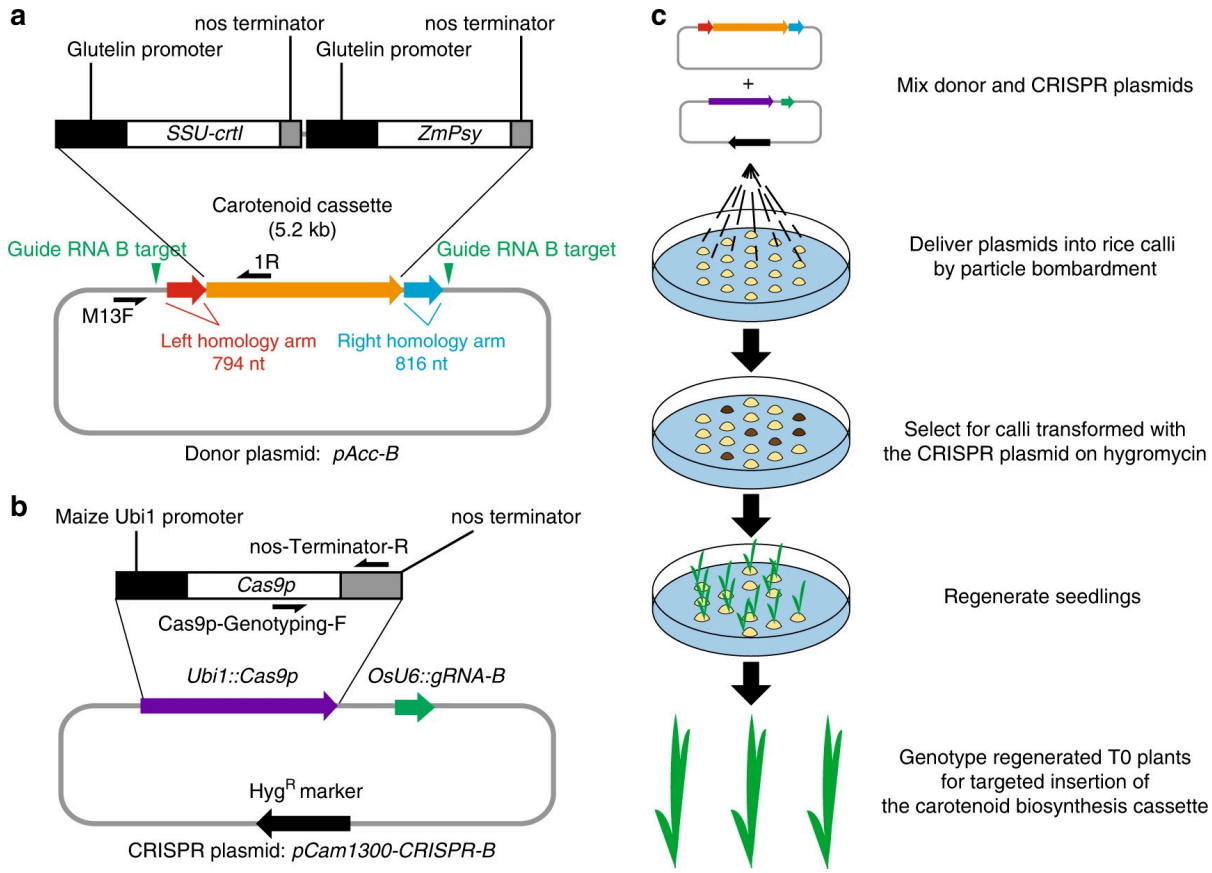


Figure 14 a) Map of the donor plasmid pAcc-B with details of the carotenoid cassette (orange arrow). Red and blue arrows represent the homology arms. The two vertical green triangles mark the positions of the guide RNA B target sites. The nucleotide sequence of the donor plasmid is presented in Supplementary Data 1. Primers used for genotyping the donor plasmid are indicated on the map. **b)** Map of the plasmid pCam1300-CRISPR-B. Genes encoding Cas9p, gRNA-B, and hygromycin resistance (Hyg ^{R}) are represented by purple, green, and black arrows, respectively. The Cas9p module is shown in detail. Primers used to genotype Cas9p are marked on the map. **c)**, Scheme for transformation, selection, and regeneration. *Source:* Dong et al. (2020).

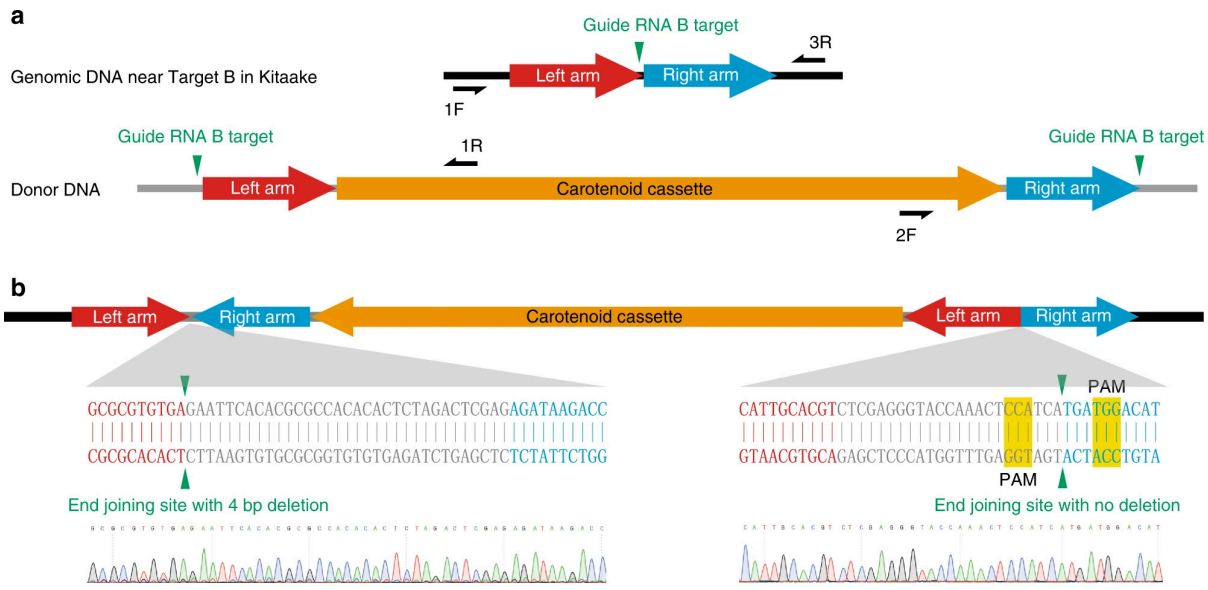


Figure 15 a) illustrates diagrams showing the genomic region near Target B in Kitaake rice and the donor DNA. Gray lines represent plasmid backbone DNA, while black lines represent Kitaake genomic DNA. The vertical green triangles mark the positions of the guide RNA B target sites. **b)** Diagram of the inserted carotenoid cassette at Target B in T0 plants #11, 16, 17, 24, 28, 48, and 50. The junction sequences in all seven plants are identical, as shown in the diagram. For convenience, only the sequencing chromatograms for T0 #48 are shown. The protospacer adjacent motif (PAM) of the original guide RNA B targets is highlighted in yellow ; *Source: Dong et al. (2020)*.

In the Golden rice study, embryonic calli was bombarded, and 55 T_0 plants were regenerated (Dong et al., 2020). They performed genetic segregation analysis on the T1 generation to obtain rice plants homozygous for the carotenoid cassette at Target B, which lack the Cas9-gRNA module. identified T1 individual 48A-7 as being homozygous for the inserted carotenoid cassette. Interestingly, this integration did not follow perfect homology-directed repair; the cassette was inserted via non-homologous end joining and was flipped in orientation. This was an unexpected outcome, but it still produced the desired phenotype. In plant breeding, such work demonstrates the practical challenges and potential of precision gene insertion, showing that even rare, imperfect recombination events can produce stable traits for crop improvement.

Cas12 nucleases

These form several variants of the type-V CRISPR systems and are often characterized by the presence of a single RuvC-like nuclease domain that mediates the cleavage of both target DNA strands, unlike Cas9, which requires an HNH domain (cleaving the DNA strand complementary to the crRNA) and a RuvC domain (cleaving the complementary strand). Cas 12 nucleases like Cas12a(Cpf1), Cas12b, Cas12e, and Cas12f recognize T-rich PAM sequences, and they produce staggered DNA breaks with 5´ overhangs, which improves insertion (Xin et al., 2022). This allows for the targeting of genomic sequences that are not targetable by G-rich PAM-dependent nucleases, such as Cas9 (NGG). Furthermore, Cas12 nucleases can process multiple crRNAs simultaneously from an array, thereby simplifying multiplex gene editing

compared to Cas9. However, overall, Cas9 achieves a higher editing frequency in plants for biallelic or homozygous mutations, while Cas12a is efficient at specific target site editing (Lee et al., 2019).

> Read: Scientists have harnessed TnpB, an RNA-guided DNA endonuclease from the radiation-resistant bacterium Deinococcus radiodurans, which can cleave targeted DNA sequences with 34% efficiency in both dicots and monocots. The system has been utilized to develop resilient rice varieties and holds promise for future applications of the TnpB system in crop improvement for trait enhancement, including climate resilience, disease and pest resistance, yield improvement, and nutritional enrichment (Karmakar et al., 2024).

For more information, read: A miniature alternative to Cas9 and Cas12: Transposon-associated TnpB mediates targeted genome editing in plants.

Base editing

This technology, developed by Davis Liu and Akihiko Kondo in 2016, enables the direct, irreversible conversion of one base of DNA into another at a specific target site without introducing double-stranded breaks (DSBs) (Komor et al., 2016). Base editing combines two components, including Cas9 nickase (nCas9), a catalytically inactive form of the Cas9 enzyme, or a transcription activator-like effector (TALE) repeat array fused to a nucleoside deaminase enzyme that removes and converts one base to a different base. DNA base editors are categorized as cytosine base editors (CBEs), which substitute a cytosine with a Thymine (C→T), and adenosine base editors (ABEs), which introduce a substitution of adenine with guanine (A→G) (Figure 16). New base editors, which can introduce cytosine-to-guanine (CGBEs) and adenine-to-cytosine (ACBEs) substitutions, have been developed, although they are still of limited use.

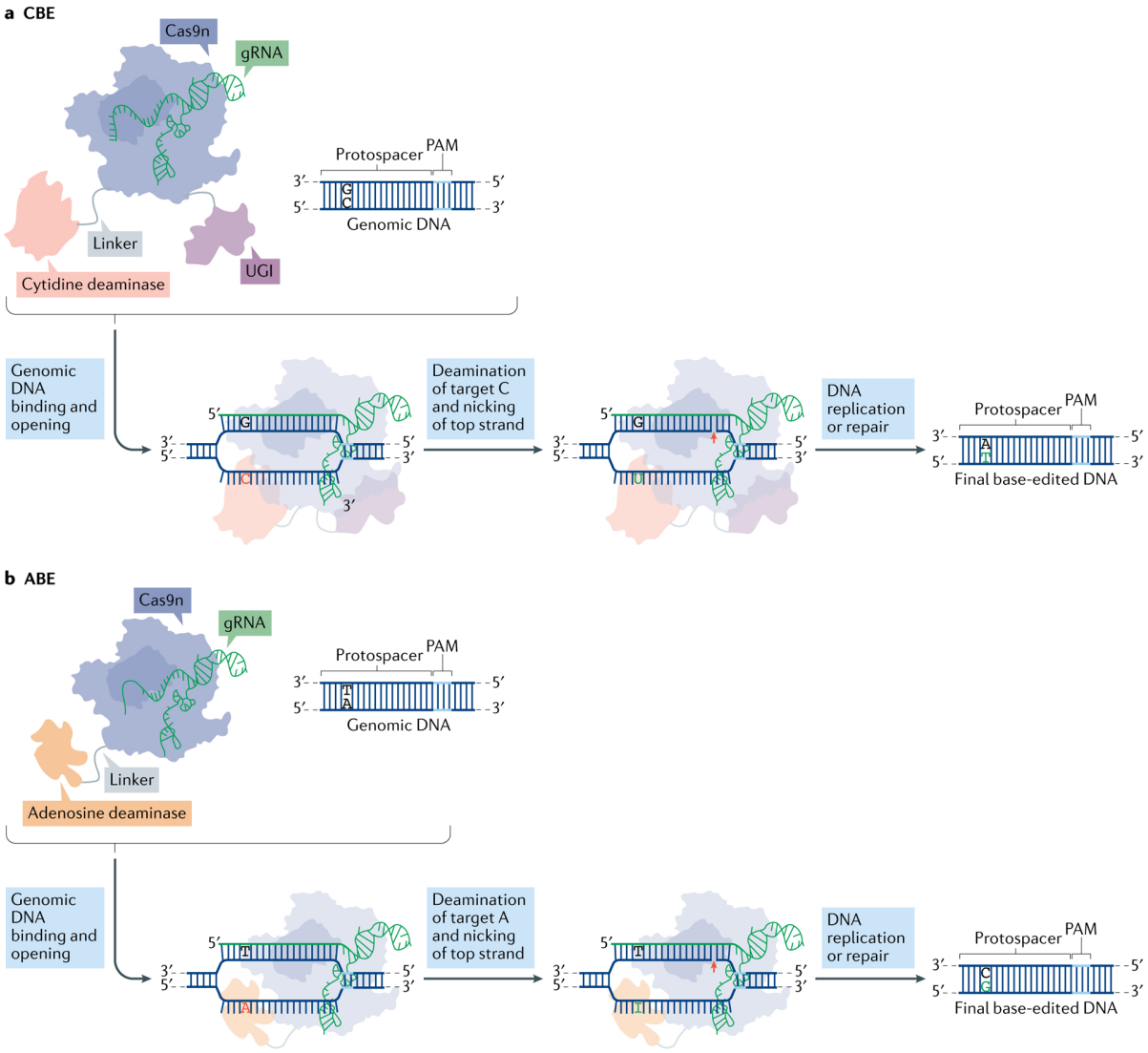


Figure 16 a) Cytosine Base Editor (CBE) mechanism; Key components of the CBE are shown in colored text boxes. The inclusion of uracil glycosylase inhibitor (UGI) (which is optional) 'shields' the U•G intermediate from being excised by uracil DNA glycosylase (UDG), enhancing the efficiency of the final base-edited DNA result. The Cas9 nickase variant (Cas9n) makes a cut in the top strand (red arrow), while the cytidine deaminase transforms cytosine (red) into uracil (green). The complete transformation of a C•G to a T•A base pair is accomplished via the specified steps. **b)** The adenine base editor (ABE) mechanism functions similarly to CBE, but it does not include a UGI domain within the ABE design. ABE-mediated editing enables the conversion of an A•T base pair to a G•C base pair through an inosine-containing intermediate. gRNA = guide RNA; PAM = protospacer adjacent motif; target A = ABE's desired base substrate; and target C = CBE's desired base substrate. *Adapted from Porto et al. (2020)*.

A guide RNA (gRNA) directs the base editor to the matching DNA sequence for CRISPR base editors. The Cas domain then unwinds the DNA sequence, thereby exposing a short stretch of single-stranded DNA where the tethered deaminase chemically changes one base to another. The base editing technology can revert a single base change or SNP with high accuracy, without disrupting the gene, thereby minimizing insertions and deletions (Indels). The approach has been applied to rice, wheat, maize, and tomatoes.Because base editors produce fewer indels and prevent harmful outcomes associated with DSB

repair, they are more efficient than standard Cas nucleases. However, their precision is limited by: (1) base editors deaminate within a small window of 4–5 nucleotides (nt), which causes C or A nucleotides very close to the target C or A can also undergo conversion, resulting in 'bystander editing', (2) also base editing is restricted by Cas domain which requires a PAM located about 15 ± 2 nucleotides from the intended base change, (3) Some base editors can induce off-target mutations in DNA and RNA, (4) Current base editor designs only perform six of the 12 possible types of point mutation, hence cannot make large insertions, deletions, and most base transversions. Gaps still exist in base editing capabilities, despite engineering efforts that have reduced off-target activity and expanded the compatibility of PAMs.

Prime editing

Prime editing enables precise small insertions, deletions, and base swaps in a unique and precise manner. Problems associated with the use of base editing motivated researchers to develop prime editing. The latter differs from traditional Cas9 in that this system performs targeted DNA modification for all 12 possible base substitutions without inducing double-stranded breaks (Anzalone et al., 2019).This Technology was developed by Andrew Anzalone, a postdoctoral fellow in David Liu's lab in 2019, and it utilizes the fusion of a Cas9 nickase (nCas9) to an engineered reverse transcriptase (RT) and guided by a prime editing guide RNA (pegRNA), which specifies the target site for the edit and contains a programmable RNA template for the desired DNA sequence change (Figure 17).

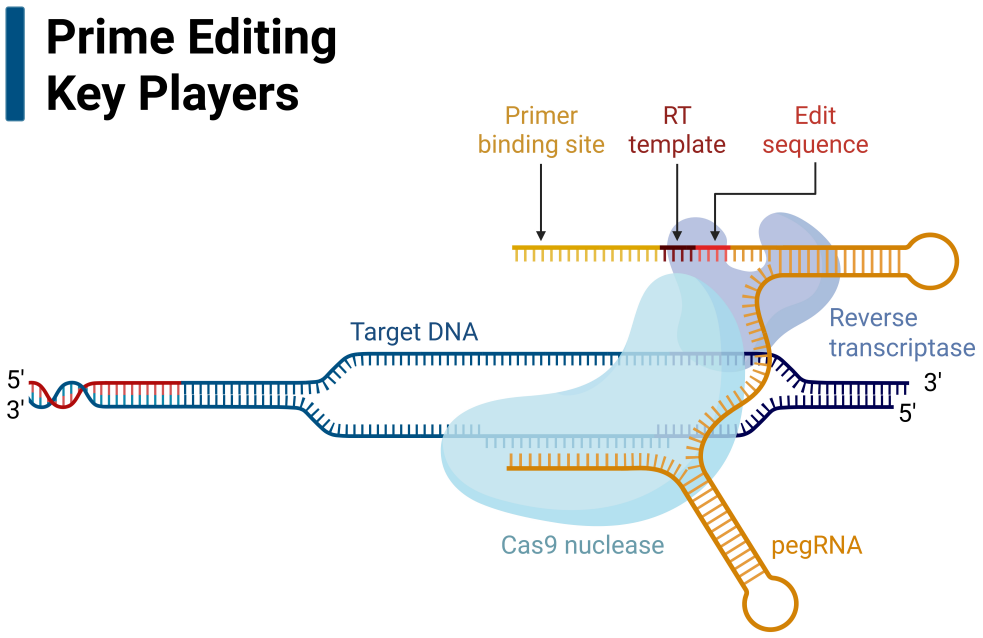


Figure 17 Prime editing. A genome editing technique that utilizes three main components: a modified Cas9 (nCas9), a prime editing guide RNA (pegRNA), and a reverse transcriptase (RT) enzyme. *Image adapted from the BioRender template created by Joanne Rose (Anzalone et al., 2019)*.

Prime editing (PE) surpasses the base editing limitation to only four substitution types, offering greater precision and flexibility. To mediate editing, the PE system nicks the target site on the genome and extends a new DNA sequence from this nick site using the pegRNA as a template (Figure 18). This edited DNA strand is then incorporated into the genome through endogenous cellular processes that can be promoted by also nicking the non-edited strand (Figure 18).

The prime editing mechanism involves the fusion of a Cas9 nickase and a reverse transcriptase (RT) with a prime editing guide RNA (pegRNA) to form a complex. PegRNA directs the complex to the complementary DNA target site. The Cas9 nickase creates a single-strand break (nick), three bp upstream of the PAM site in one of the DNA strands, exposing a 3'-hydroxyl group without making a double-stranded break. The primer binding site (PBS) of pegRNA anneals to the exposed 3' end, and the reverse transcriptase enzyme, also part of the prime editor complex, extends the nicked DNA strand using the RNA template sequence (edit site) located within the pegRNA.

This synthesizes a new DNA flap containing the desired edit. Flap resolution and repair occur when the newly synthesized 3' flap, containing the edited sequence, competes with and displaces the original 5' flap. Cellular nucleases remove the unwanted 5' flap, and DNA ligases seal the nick, incorporating the

edited DNA into the genome. Complementary strand correction occurs when a second gRNA is used to nick the opposite strand, allowing the edit to be copied to both strands and completing the duplex. Cellular nucleases then remove unedited DNA, fully integrating the new sequence.

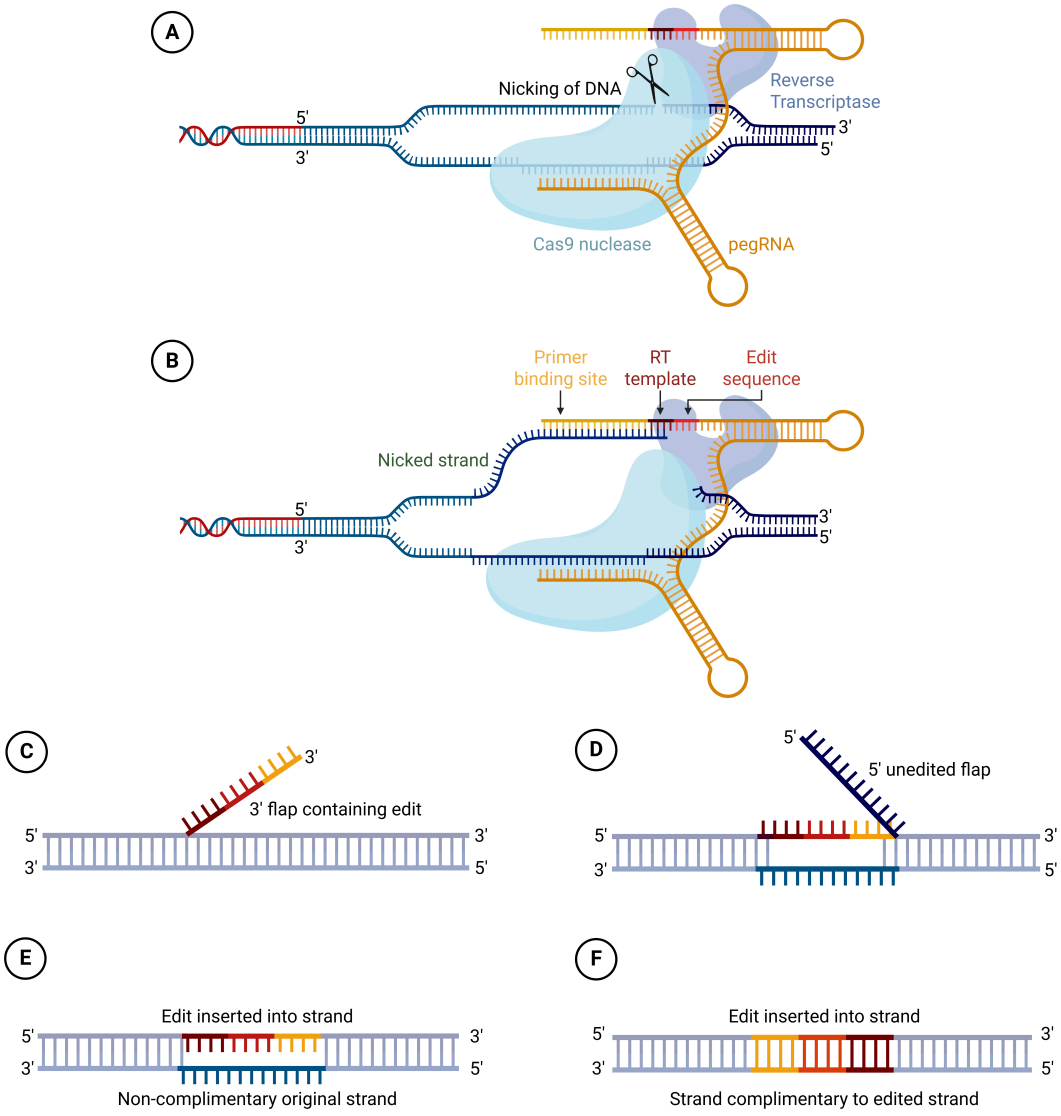


Figure 18 Prime editing system and its editing mechanism. **A)** Spacer sequence of pegRNA guides the complex to the target site, and the complex binds to the DNA. Cas9 identifies a PAM and nicks three nucleotides upstream of the PAM. **B)** a primer binding site (PBS) binds to nicked genomic DNA, and reverse transcriptase (RT uses reverse transcription template (RTT) as a template to copy the edit into a 3' DNA flap. The edited 3'-flap **C)** competes with the original 5'-flap **D)** for binding to the target DNA. If a 3'-flap successfully competes with the original 5'-flap, the desired edit will be incorporated **E)**; however, editing will be lost if the 5'-flap outcompetes the 3'-flap. **F)** *Image adapted from the BioRender template created by Joanne Rose (Anzalone et al., 2019).*

Using an improved prime editing system, Gupta et al. (2023) developed two strategies to confer resistance to bacterial blight (BB) caused by *Xanthomonas oryzae* pathovar *oryzae* (*Xoo*). The first approach

involved the Knock-in of TAL effector binding elements (EBE) derived from the BB-susceptible gene SWEET14 into the promoter of a dysfunctional executor R gene, *xa23*. This enables TALE-dependent activation of *xa23* and BB resistance, achieving a 47.2% editing rate in T_0 plants with 18% biallelic editing. The second approach involved editing the transcription factor TFIIA gene *TFIIAγ5*, which is required for TAL effector-dependent BB susceptibility, converting it to a resistant xa5 allele. This resulted in an editing efficiency of 88.5% and 30% biallelic edits among T_0 plants. Both strategies produced T_1 plants with broad resistance against multiple *Xoo* strains. These plants showed significantly shorter lesion lengths compared to the wild type. These approaches, which utilize prime editing, hold promise for reducing BB disease occurrence in rice and can be applied to other crops.

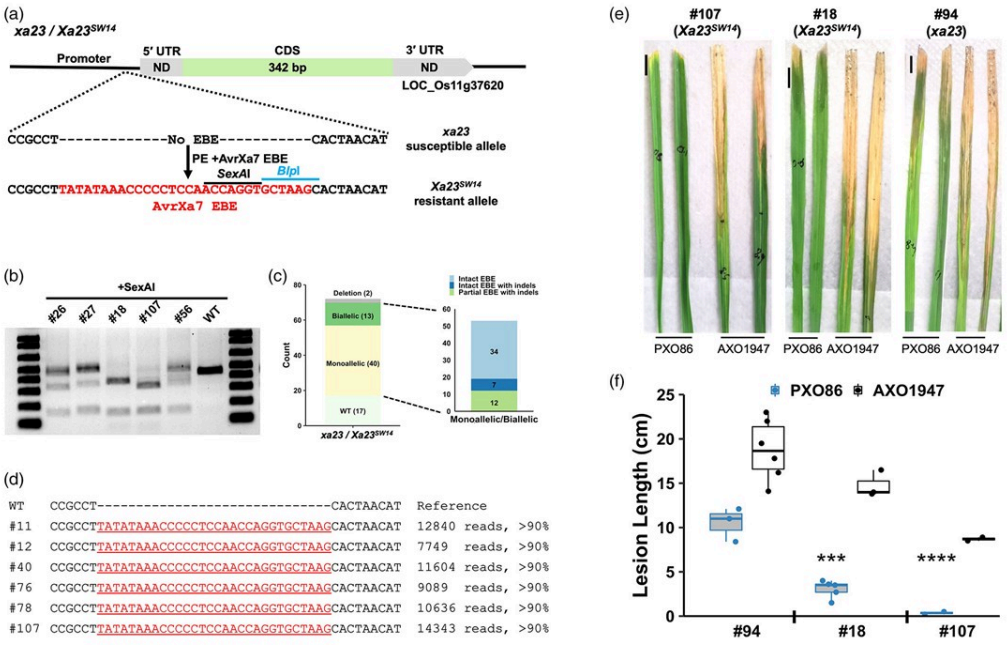


Figure 19 PE-based knock-in of EBE in *xa23*. **a)** Gene structures of xa23 and Xa23SW14. Intronless CDS (coding sequence) and untranslated sequences (ND, not determined) are shown. Intended knock-in sequence of EBEs is in red, with restriction sites shown. **b)** Gel image showing SexAI cleavage of relevant PCR products of some representative T0 and the wild-type (WT) lines. **c)** Summary of editing events based on gel and deep sequencing of PCR amplicons. Complete, partial digestions, and undigested lines are diallelic, monoallelic edits, and wild type, respectively. Number in parentheses shows the count out of the total T0 lines. The second bar illustrates the types of edits that occurred in monoallelic and diallelic lines based on the deep amplicon sequencing of xa23. **d)** Deep sequencing of xa23/Xa23SW14 amplicons. Independent biallelic lines with intact EBE (in red) inserted without indels are shown. Total number and percentage of reads of individual edited lines are presented on the right. Dashes in reference separate the two bases where EBE is inserted. **e,f)** Lesion lengths of two biallelic and a wild-type line inoculated with PXO86 and AXO1947. Lesion lengths were measured 12 days post-inoculation on three to five leaves of individual plants (n = 3–5). Scale bar, 1 cm. ***P-value <0.001 and ****P-value <0.0001 with t-test adjusted using Bonferroni correction for multiple comparisons. *Adapted from Gupta et al. (2023)*.

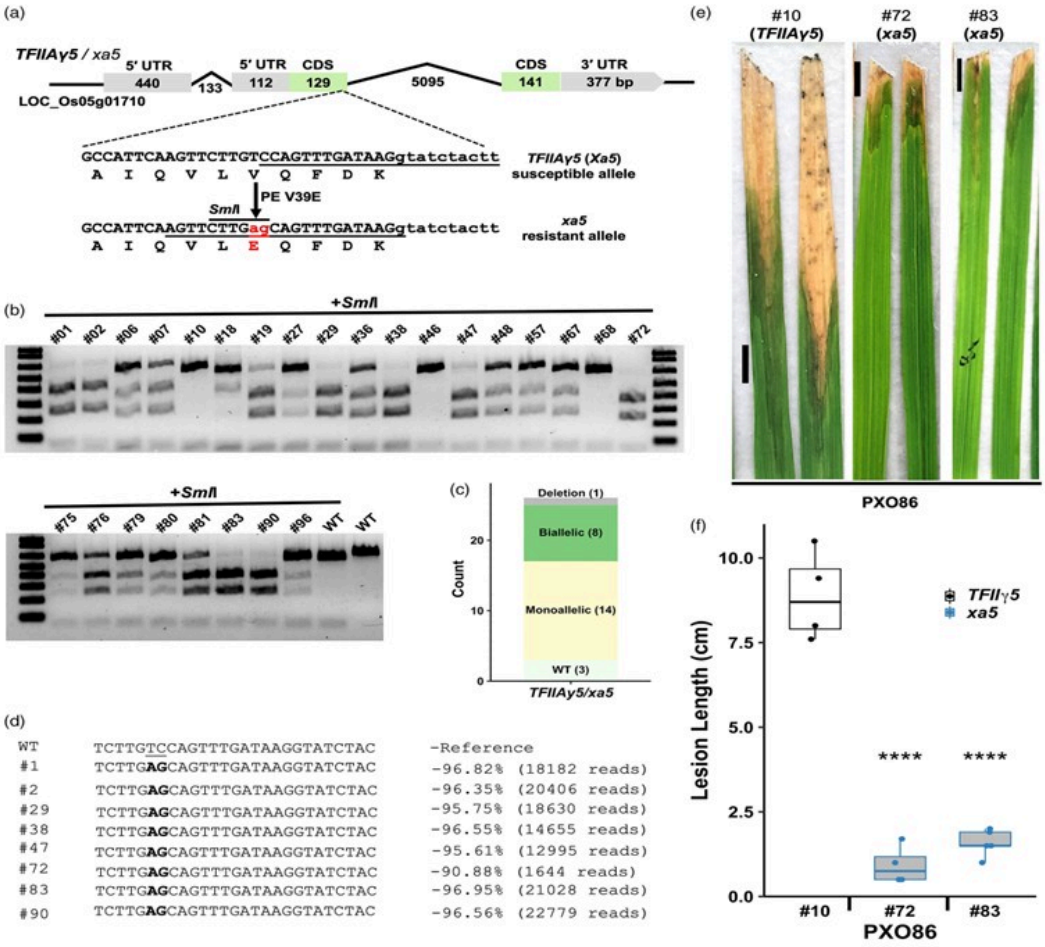


Figure 20 Prime editing of TFIIAγ5 to generate the xa5 allele. **a)** Gene structures of TFIIAγ5 and xa5. PE target site (underlined) in TFIIAγ5 allele and nick site (underlined) in the edited strand of the xa5 allele are shown. **b)** Genotyping of edited lines as indicated above, lanes with *SmlI* digestion of relevant PCR amplicons. **c)** Bar plot summarizing the SmlI digestion. Complete, partial digestions, and undigested lines are diallelic, monoallelic edits, and wild type, respectively. **d)** Deep sequencing of biallelic lines with desired edits. The wild-type bases are underlined in the reference sequence, and edited bases are bold-highlighted in sequenced lines. The numbers on the left indicate the percentage and number of reads mapped to the xa5 allele. **e-f)** Disease phenotypes of edited biallelic lines. Lesion lengths were measured 12 days post-inoculation with PXO86 on three to five leaves of individual plants (n = 3–5): scale bar, 1 cm. **** represents P-value <0.0001 with t-test adjusted using Bonferroni correction for multiple comparisons. *Adapted from* Gupta et al. (2023).

Chapter Summary

Genetic engineering techniques are crucial for agricultural biotechnology, enabling the application of functional genomics to study complex traits and accelerate crop improvement. This technology is not new and has progressed through various stages, from early transgenesis using methods like meganucleases and zinc finger nucleases to the current use of precise, programmable tools like CRISPR/Cas9 systems. Among these technologies, the CRISPR/Cas system is the most adapted technology

due to its simplicity and effectiveness. CRISPR/Cas reagents can be introduced into plant cells using either *Agrobacterium*– or biolistic-mediated delivery methods. Edited plants that carry either the CRISPR reagents or selectable marker genes can be backcrossed to non-transgenic plants to segregate the transgene from the edited plant. CRISPR/Cas9 reagents can also be delivered as an in vitro-assembled RNP complex for DNA-free editing. Improving methods to recover edited events from this method is a goal for academics and industry alike. Other toolkits, such as base and prime editing, enable single-nucleotide changes and small indels, respectively, without introducing double-stranded breaks, thereby reducing off-target effects and unintended mutations. Collectively, these tools facilitate the precise manipulation of plant genomes to improve crop traits, including disease and pest resistance, as well as other attributes, for which encoding genes are identified. Base and prime editing can facilitate allele swapping experiments without requiring backcross breeding of desirable alleles that may co-inherit many linked undesirable genes, resulting in a yield penalty. Thus, genome editing holds considerable promise for the precision breeding of crop plants in sustainable agriculture.

References

Addgene. (2017). Chapter 3: Using CRISPR in Your Experiments. CRISPR 101: A Desktop Resource

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576* (7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Asmamaw, M., & Zawdie, B. (2021). Mechanism and applications of CRISPR/Cas-9-mediated genome editing. *Biologics: Targets and Therapy*, 353–361.

Bogdanove, A. J., & Voytas, D. F. (2011). TAL effectors: Customizable proteins for DNA targeting. *Science*, *333* (6051), 1843–1846. https://doi.org/10.1126/science.1204094

Char, S. N., Unger-Wallace, E., Frame, B., Briggs, S. A., Main, M., Spalding, M. H., Vollbrecht, E., Wang, K., & Yang, B. (2015). Heritable site-specific mutagenesis using TALENs in maize. *Plant Biotechnology Journal*, *13* (7), 1002–1010. https://doi.org/10.1111/pbi.12344

Dong, O. X., Yu, S., Jain, R., Zhang, N., Duong, P. Q., Butler, C., Li, Y., Lipzen, A., Martin, J. A., Barry, K. W., Schmutz, J., Tian, L., & Ronald, P. C. (2020). Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR-Cas9. *Nature Communications*, *11* (1), 1178. https://doi.org/10.1038/s41467-020-14981-y

Gupta, A., Liu, B., Chen, Q.-J., & Yang, B. (2023). High-efficiency prime editing enables new strategies for broad-spectrum resistance to bacterial blight of rice. *Plant Biotechnology Journal*, *21* (7), 1454–1464. https://doi.org/10.1111/pbi.14049

Hussain, A., Ding, X., Alariqi, M., Manghwar, H., Hui, F., Li, Y., Cheng, J., Wu, C., Cao, J., & Jin, S. (2021). Herbicide resistance: Another hot agronomic trait for plant genome editing. *Plants*, *10* (4), 621. https://doi.org/10.3390/plants10040621

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169* (12), 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Jacquier, A., & Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, *41* (2), 383–394. https://doi.org/10.1016/S0092-8674(85)80011-8

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337* (6096), 816–821. https://doi.org/10.1126/science.1225829

Kim, Y.-G., Cha, J., & Chandrasegaran, S. (1996a). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, *93* (3), 1156–1160. https://doi.org/10.1073/pnas.93.3.1156

Kim, Y.-G., Cha, J., & Chandrasegaran, S. (1996b). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, *93* (3), 1156–1160. https://doi.org/10.1073/pnas.93.3.1156

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533* (7603), 420–424. https://doi.org/10.1038/nature17946

Le Rhun, A., Escalera-Maurer, A., Bratovič, M., & Charpentier, E. (2019). CRISPR-Cas in Streptococcus pyogenes. *RNA Biology*, *16* (4), 380–389. https://doi.org/10.1080/15476286.2019.1582974

Lee, K., Zhang, Y., Kleinstiver, B. P., Guo, J. A., Aryee, M. J., Miller, J., Malzahn, A., Zarecor, S., Lawrence-Dill, C. J., Joung, J. K., & others. (2019). Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnology Journal*, *17* (2), 362–372. https://doi.org/10.1111/pbi.12982

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., & Gao, C. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications*, *8* (1), 14261. https://doi.org/10.1038/ncomms14261

Mansour, S. L., Thomas, K. R., & Capecchi, M. R. (1988). Disruption of the proto-oncogene int-2 in mouse

embryo-derived stem cells: A general strategy for targeting mutations to non-selectable genes. *Nature*, *336* (6197), 348–352. https://doi.org/10.1038/336348a0

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D., & Rebar, E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*, *29* (2), 143–148. https://doi.org/10.1038/nbt.1755

Miller, J., McLachlan, A., & Klug, Ajte. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO Journal*, *4* (6), 1609–1614. https://doi.org/10.1002/j.1460-2075.1985.tb03825.x

Osakabe, K., Osakabe, Y., & Toki, S. (2010). Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. *Proceedings of the National Academy of Sciences*, *107* (26), 12034–12039.

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., Wright, S. Y., Hinchliffe, E., Adams, J. L., Silverstone, A. L., & Drake, R. (2005). Improving the nutritional value of Golden Rice through increased provitamin A content. *Nature Biotechnology*, *23* (4), 482–487. https://doi.org/10.1038/nbt1082

Porto, E. M., Komor, A. C., Slaymaker, I. M., & Yeo, G. W. (2020). Base editing: Advances and therapeutic opportunities. *Nature Reviews Drug Discovery*, *19* (12), 839–859. https://doi.org/10.1038/s41573-020-0084-6

Puchta, H. (2005). The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. *Journal of Experimental Botany*, *56* (409), 1–14. https://doi.org/10.1093/jxb/eri025

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., & Mark Cigan, A. (2016). Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. *Nature Communications*, *7* (1), 13274. https://doi.org/10.1038/ncomms13274

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, *11* (9), 636–646. https://doi.org/10.1038/nrg2842

Xin, C., Yin, J., Yuan, S., Ou, L., Liu, M., Zhang, W., & Hu, J. (2022). Comprehensive assessment of miniature CRISPR-Cas12f nucleases for gene disruption. *Nature Communications*, *13* (1), 5623. https://doi.org/10.1038/s41467-022-33346-1

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P., & Potrykus, I. (2000). Engineering the provitamin

A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*, *287* (5451), 303–305. https://doi.org/10.1126/science.287.5451.303

Zhang, S., Shen, J., Li, D., & Cheng, Y. (2021). Strategies in the delivery of Cas9 ribonucleoprotein for CRISPR/Cas9 genome editing. *Theranostics*, *11* (2), 614. https://doi.org/10.7150/thno.47007

Zobrist, J. D., Lee, K., & Wang, K. (2023). Application of CRISPR/Cas9 for targeted mutagenesis in teosinte Zea mays ssp. Parviglumis. *Plant Biotechnology Journal*, *22* (4), 796. https://doi.org/10.1111/pbi.14269

About the Authors

Walter P. Suza is the George Washington Carver Endowed Chair and an Adjunct Associate Professor at Iowa State University. He holds a Ph.D. in plant sciences, specializing in molecular physiology, from the University of Nebraska-Lincoln, where his doctoral research identified the role of the *JASMONATE RESISTANT 1* (*JAR1*) gene in the production of jasmonoyl-isoleucine (JA-Ile) in wound response signaling in Arabidopsis and tomato. His postdoctoral work includes elucidating the spatial and temporal regulation of sterol biosynthesis in *Nicotiana benthamiana*. His current research applies genetic, physiological, molecular, and biochemical approaches to understand how phytosterols modulate cellular metabolism in maize and soybean. He teaches courses in Genetics and Biotechnology, Crop Physiology, and World Food Issues in the Department of Agronomy. In addition to co-developing courses for Iowa State's Distance Master's Program in Plant Breeding, he served for eight years as Director of Plant Breeding E-Learning in Africa (PBEA), expanding access to open educational resources in crop genetic improvement. His international experience includes work in central and southern Africa with the World Food Programme in Angola and UNICEF in Zimbabwe, where he focused on food security assessment, mapping vulnerable households, drought evaluation, and coordination of food aid projects. In 2024, he received the World Food Prize Top Agri-Food Pioneer Award.

Philip W. Becraft, Professor Emeritus, received his B.A. in Botany from the University of Montana, Missoula, his M.S. in Agronomy from Montana State University, Bozeman, and his Ph.D. in Genetics from the University of California, Berkeley. Upon graduation, Dr. Becraft completed an NSF postdoctoral Fellow/Courtesy Assistant Professor position in the Horticulture Science Department at the University of Florida, Gainesville. In 1996, Dr. Becraft joined the faculty of Iowa State University as Assistant Professor, was promoted to Associate Professor in 2002, and full Professor in 2009. His research was on regulation of plant growth and development. Using molecular genetics and genomics, Dr. Becraft's team established that a genetic hierarchy regulates aleurone development and isolated several of the key genes involved in endosperm development.

Faizo Kasule is a Graduate Research and Teaching Assistant in the Interdepartmental Genetics and Genomics Program at Iowa State University, working in the in the Department of Agronomy. His research focuses on advancing plant genetic transformation and genome editing in maize and sorghum, with a particular emphasis on utilizing morphogenes, peptides, and effectors to enhance transformation efficiency and regeneration. He contributes to teaching within the program while pursuing his graduate studies. Before joining Iowa State University, Faizo worked as a plant breeder with the National Agricultural

Research Organization (NARO) in Uganda, supporting efforts to improve millets and sorghum crops and strengthen food security. He also served as a Graduate Research and Teaching FAPA Assistant with the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) in Kampala, Uganda, and worked as a biotechnologist with the Alliance of Bioversity International and CIAT, Kampala, Uganda. Faizo holds a B.Sc. in Biotechnology and an M.Sc. in Plant Breeding and Seed Systems from Makerere University in Uganda.

Thomas Lübberstedt grew up on a horticultural farm in Hamburg, Germany, studied Ag.- and Horticulture (Technical University Munich, University of Hannover), and received his Dr. rer. nat. in Biology from the Ludwigs-Maximilians University in Munich in 1993. During this time he developed his interest in integrating plant breeding and biotechnologies including molecular genetics. He worked in three countries, Germany (1989-2001: LMU Munich: PhD, University of Hohenheim: Habilitation, Technical University of Munich, Heisenberg fellow), Denmark (Senior Scientist at Danish Institute of Agricultural Sciences, 2001-2007), and the U.S. (since 2007 at Iowa State University: ISU). At Iowa State University, Lübberstedt is K.J. Frey Chair, Director of M.S. Distance Plant Breeding, Faculty Scholar of Plant Sciences Institute, and founder and professor-in-charge of the Doubled Haploid (DH) Facility at ISU. Over the past ca. 12 years, Lübberstedt focused his research on DH technology and its application, which resulted in USDA-, NSF-, and FFAR-funded projects, patents and novel germplasm including haploid inducers. Lübberstedt served as chair of the Plant Breeding Coordinating Committee (PBCC) in 2015/16, president of the National Association for Plant Breeders (NAPB) 2023/24, and Director of R.F. Baker Center for Plant Breeding at ISU (2007-2023).

Madan K. Bhattacharyya received his B. Sc. (Agriculture: Genetics & Plant Breeding) from Assam Agricultural University and then M.S. (Olericulture: Vegetable Breeding) from Punjab Agricultural University. He received his Ph.D. in Plant Sciences from Western University, London, Canada as a Canadian Commonwealth Scholar. He then joined then John Innes Institute, now John Innes Center, where he cloned the peas "r" locus studied by Gregor Mendel. He started his lab in Noble Foundation, USA, and then joined Iowa State University in 2000, where he is currently a professor. Prof. Bhattacharyya leads a research program primarily focused on the genetic and molecular mechanisms of plant disease resistance and abiotic stress tolerance. His notable contributions include the first cloning and characterization of the plant signaling enzyme phospholipase C, understanding the molecular basis of soybean host resistance against Phytophthora sojae, Arabidopsis nonhost resistance against soybean pathogens, pathogenicity mechanisms of Fusarium virguliforme involved in soybean sudden death syndrome (SDS) development, isolation and characterization of an active endogenous transposable element in soybean, and genetic mechanisms underlying plant adaptation to adverse climatic conditions including cold, heat and drought stress. The overarching goal of his laboratory is to translate fundamental discoveries into strategies for enhancing soybean resilience to both biotic and abiotic stresses. Currently, his lab is developing drought and flood tolerance soybean cultivars and modifying a rice blast resistance gene to confer complete SDS resistance in soybean. Bhattacharyya also contributes to graduate education. He taught Plant Genetics

for a decade. He currently teaches Applied Plant Molecular Genetics & Biotechnology to graduate students. He has authored 99 peer reviewed articles with over 10,000 citations. In 2021, Bhattacharyya was elected as a fellow of the American Association for the Advancement of Science (AAAS) for his distinguished contributions to the field of plant-microbe interactions, particularly for understanding the interactions between soybean and its fungal and oomycete pathogens.