Czech University of Life Sciences Prague

Faculty of Forestry and Wood Sciences

INTRODUCTION TO FOREST GENETICS

For master students of forestry engineering

2026

doc. Ing. Jan Stejskal, Ph.D.

Czech University of Life Sciences Prague

Faculty of Forestry
and Wood Sciences

INTRODUCTION TO FOREST GENETICS

For master students of forestry engineering

doc. Ing. Jan Stejskal, Ph.D.

2026

TABLE OF CONTENTS

PREFACE

Dear Students,

Welcome to **"Introduction to Forest Genetics."** This book is designed specifically for master's students in forestry engineering to provide you with the essential knowledge of genetics as it applies to forests and trees.

The field of forest genetics is critical for understanding how trees grow, adapt, and respond to their environment. It is also key to improving tree breeding and conservation efforts. This book will guide you through the basic principles of genetics, explore their application to forests, and introduce you to modern tools and techniques used in genetic research.

The chapters are organized to take you from foundational concepts, like the structure and function of DNA, to advanced topics like tree improvement and genetic conservation. The focus is on clear explanations and practical applications, making it easier for you to connect science with real-world forestry challenges.

Whether your interest lies in understanding tree biology, improving forest productivity, or preserving biodiversity, this book is a useful resource. I hope it will support your studies and help you gain the knowledge you need to contribute to sustainable forestry.

I wish you success in your learning and your future work!

Jan Stejskal

Author

ACKNOWLEDGMENTS

I gratefully acknowledge my colleagues Jiří Korecký, Jaroslav Čepl, and Milan Lstibůrek for their invaluable support and collaboration during the writing of this textbook.

1. THE FOUNDATION OF HEREDITARY MATERIAL

Molecular genetics is a biology field that studies the structure and function of genes and genetic principles at the molecular level. It involves preserving and transmitting genetic information, typically in connection with recombination processes, gene expression, gene regulation, and repair mechanisms or mutations. It examines how genetic information is encoded in DNA, transferred to RNA, and ultimately translated into proteins that perform various functions within the cell.

An alternatively used term, **molecular biology**, is a broader field of biology that studies biological processes at the molecular level. It includes genetics, biochemistry, biophysics, and other disciplines that deal with the molecular mechanisms of life. However, when the term molecular biology is mentioned in genetics, it often specifically refers to the processes related to the transmission and preservation of genetic information, making the terms molecular genetics and molecular biology equivalent.

The life of a tree begins with a single cell containing all the genetic information necessary for its entire life cycle. In this respect, trees are no different from any other living organisms. This essential information is inherited from both parent trees. Before delving into the principles of population genetics, we need to understand the nature of hereditary material.

1.1 HISTORY OF MOLECULAR GENETICS

The historical development of molecular genetics is characterized by a series of groundbreaking discoveries that gradually revealed the fundamental principles of heredity and gene function. The first significant step was made by **Gregor Mendel**, who, in the 1860s, experimented with pea plants and formulated the basic laws of inheritance based on numerous observations. His work is considered pioneering, and he is regarded as the father of genetics. However, in the context of molecular genetics, his contributions were only theoretical foundations and considerations, as the level of available tools at the time did not allow for research at the cellular level.

At the beginning of the 20th century, **Thomas Hunt Morgan** contributed to the development of molecular genetics by researching fruit flies (*Drosophila melanogaster*). His work led to the formulation of the chromosomal theory of inheritance, which demonstrated that genes are arranged linearly on chromosomes. This discovery was crucial for further development in genetics.

In 1944, **Oswald Avery, Colin MacLeod, and Maclyn McCarty** demonstrated that DNA carries genetic information. This discovery paved the way for understanding the molecular basis of heredity. In 1953, with the help of Rosalind Franklin and Maurice Wilkins, James Watson, and Francis Crick revealed the structure of the double helix of DNA, explaining how DNA stores and transmits genetic information.

In the 1960s, **Marshall Nirenberg and Har Gobind Khorana** deciphered the genetic code, showing how the sequence of nucleotides in DNA determines the sequence of amino acids in proteins. This discovery was key to understanding how genes control biological processes.

The 1970s brought another breakthrough, the development of recombinant DNA technology by **Herbert Boyer and Stanley Cohen.** This technology enabled gene cloning and led to genetic engineering and biotechnology development.

In 1983, **Kary Mullis** developed the polymerase chain reaction (PCR), a revolutionary technique for amplifying DNA. PCR became a fundamental tool in molecular biology and genetics, allowing for rapid and precise analysis of genetic material.

Completing the **Human Genome Project** in 2003 provided a detailed map of all human genes and their arrangement. This milestone opened new possibilities for medicine and biology, enabling a better understanding of genetic diseases and the development of personalized medicine.

The latest groundbreaking discovery in molecular genetics is the CRISPR-Cas9 technology developed by **Jennifer Doudna and Emmanuelle Charpentier** in 2012. This technology allows for precise and efficient gene editing, which has far-reaching applications in genetic research and disease treatment.

1.2 THE FUNDAMENTS OF MOLECULAR GENETICS

Informational macromolecules can be regarded as the fundamental building blocks of molecular genetics. Large biological molecules carry genetic information or ensure functional activity within cells. The main types of informational macromolecules include nucleic acids (DNA, RNA) and proteins.

Nucleic Acids

DNA (Deoxyribonucleic Acid): Carries the genetic information of an organism. DNA mostly comprises two strands of nucleotides that form a double helix. Each nucleotide comprises deoxyribose (a sugar), a phosphate group, and one of four nitrogenous bases (adenine, thymine, cytosine, guanine).

RNA (Ribonucleic Acid): Plays various roles in transmitting and realizing genetic information. RNA is mostly composed of a single strand of nucleotides: ribose (a sugar), a phosphate group, and one of four bases (adenine, uracil, cytosine, guanine). There are different types of RNA, such as mRNA (messenger RNA), tRNA (transfer RNA), and rRNA (ribosomal RNA), as well as other types of RNA, mainly with regulatory functions (e.g., miRNA, siRNA).

Proteins

Proteins are composed of long chains of amino acids. They are responsible for various functions within the cell, including catalyzing biochemical reactions (enzymes), transporting molecules, providing structure to cells and tissues, and signaling.

These macromolecules closely interact to ensure cellular processes' proper function and regulation. For example, DNA is transcribed into mRNA and then translated into proteins that perform specific tasks within the cell.

The Central Dogma of Molecular Biology

The central dogma of molecular biology is a concept that describes the flow of genetic information within cells from DNA through RNA to proteins. This principle was first formulated by Francis Crick in 1958. In general, it captures the storage, transmission, and expression processes of genetic information. This process is the fundamental mechanism for the function and regulation of all living organisms. Its schematic representation is crucial for understanding how genetic information is transcribed and ultimately manifested in functional proteins.

The central dogma can be described using three main points:

DNA Replication: Before cell division, DNA replicates to ensure that each new cell receives a complete set of genetic information. During this process, the double helix of DNA unwinds, and each strand serves as a template for synthesizing a new complementary strand.

Transcription: Information from DNA is transcribed into mRNA (messenger RNA). This process occurs in the cell nucleus and involves synthesizing an mRNA strand complementary to one of the DNA strands. RNA polymerase is the enzyme that directs this process. The resulting mRNA then carries the genetic information from the nucleus to the cytoplasm, undergoing post-transcriptional modifications.

Translation: In the cytoplasm, mRNA binds to ribosomes, serving as a protein synthesis template. tRNA (transfer RNA) brings amino acids linked together in the order specified by the mRNA. Each triplet of bases on the mRNA (codon) specifies one amino acid. This process forms a specific protein, which can perform various functions within the cell.

The following schematic can express the central dogma of molecular biology: DNA → Replication → DNA → Transcription → RNA → Translation → Protein.

1.3 NUCLEIC ACIDS

Nucleic acids are biopolymers composed of nucleotides, the fundamental building blocks of DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). They are crucial for storing and transmitting genetic information in living organisms.

The most common structure of DNA is the double helix, in which two strands of nucleotides are connected by hydrogen bonds between complementary, i.e., opposing nitrogenous bases (A-T, C-G), with the strands oriented antiparallel. The function of the DNA molecule is to store genetic information and ensure its transmission to daughter cells during cell division.

In contrast, RNA is usually single-stranded but can form secondary structures. Its functions include transmitting genetic information from DNA to proteins (via mRNA), being a component of ribosomes (rRNA), participating in the transport of amino acids (tRNA), and regulating gene expression (miRNA, siRNA).

Nucleotides

The basic building units of nucleic acids are nucleotides. Each nucleotide is composed of three components:

Sugar component (pentose): In DNA, this is deoxyribose, and in RNA, it is ribose. These sugars differ by one hydroxyl group, which significantly impacts their function. The absence of a hydroxyl group on the 2' carbon of deoxyribose in the DNA molecule results in greater stability than the RNA molecule. The stability of DNA is essential for the long-term storage of genetic information. DNA must withstand damage and changes to transmit genetic information accurately from generation to generation. On the other hand, the reactivity of RNA is advantageous because RNA molecules often function as transient copies of genetic information, which are needed only for a short period. The instability of RNA allows for its rapid degradation, which is important for regulating gene expression and other cellular processes.

Phosphate group (PO_4^{3-}): Composed of one phosphorus atom surrounded by four oxygen atoms. Nucleotides can contain one, two, or three phosphate groups. Based on the number of phosphate groups, nucleotides are classified as mono-, di-, or triphosphates. Phosphate groups play a crucial role in the structure of nucleic acids and the cell's energy metabolism and regulatory processes.

Nitrogenous bases: adenine (A), guanine (G), cytosine (C), and thymine (T) in DNA, which is replaced by uracil (U) in RNA. Nitrogenous bases are organic molecules containing nitrogen and are key components of nucleotides in nucleic acids (DNA and RNA). These bases are essential for encoding genetic information through specific pairing. There are two main groups of nitrogenous bases: purines and pyrimidines.

Purines are double-ring structures consisting of a six-membered and a five-membered ring. They include Adenine (A) and Guanine (G). **Pyrimidines** are single-ring structures consisting of a six-membered ring. They include Cytosine (C), Thymine (T), and Uracil (U).

Base Pairing

In DNA and RNA, nitrogenous bases pair according to specific rules, which allow for accurate replication and transcription of genetic information. This pairing is known as base complementarity or Watson-Crick pairing (named after the discoverers of the DNA structure) or Chargaff's rule, named after the biochemist who elucidated the specific ratios of nitrogenous bases. In DNA, Adenine (A) pairs with thymine (T) through two hydrogen bonds, and guanine (G) pairs with cytosine (C) through three hydrogen bonds. Contrastingly, in RNA, Adenine (A) pairs with uracil (U) through two hydrogen bonds, and guanine (G) pairs with cytosine (C) through three hydrogen bonds. This specific base pairing is fundamental for storing and transmitting genetic information, ensuring that the sequence of bases in one strand of DNA or RNA precisely matches the sequence in the complementary strand (Fig. 1.1).

Nomenclature of Nucleotides

Each nucleotide comprises three basic components: a nitrogenous base, a pentose sugar (ribose in RNA, deoxyribose in DNA), and one or more phosphate groups. The nomenclature of nucleotides involves naming the nitrogenous base, the sugar, and the number of phosphate groups. A nucleoside is a nucleoside when a nitrogenous base is covalently bonded to a pentose sugar through a glycosidic bond. Nucleosides do not contain a phosphate group. A nucleotide is formed when one or more phosphate groups are added to a nucleoside.

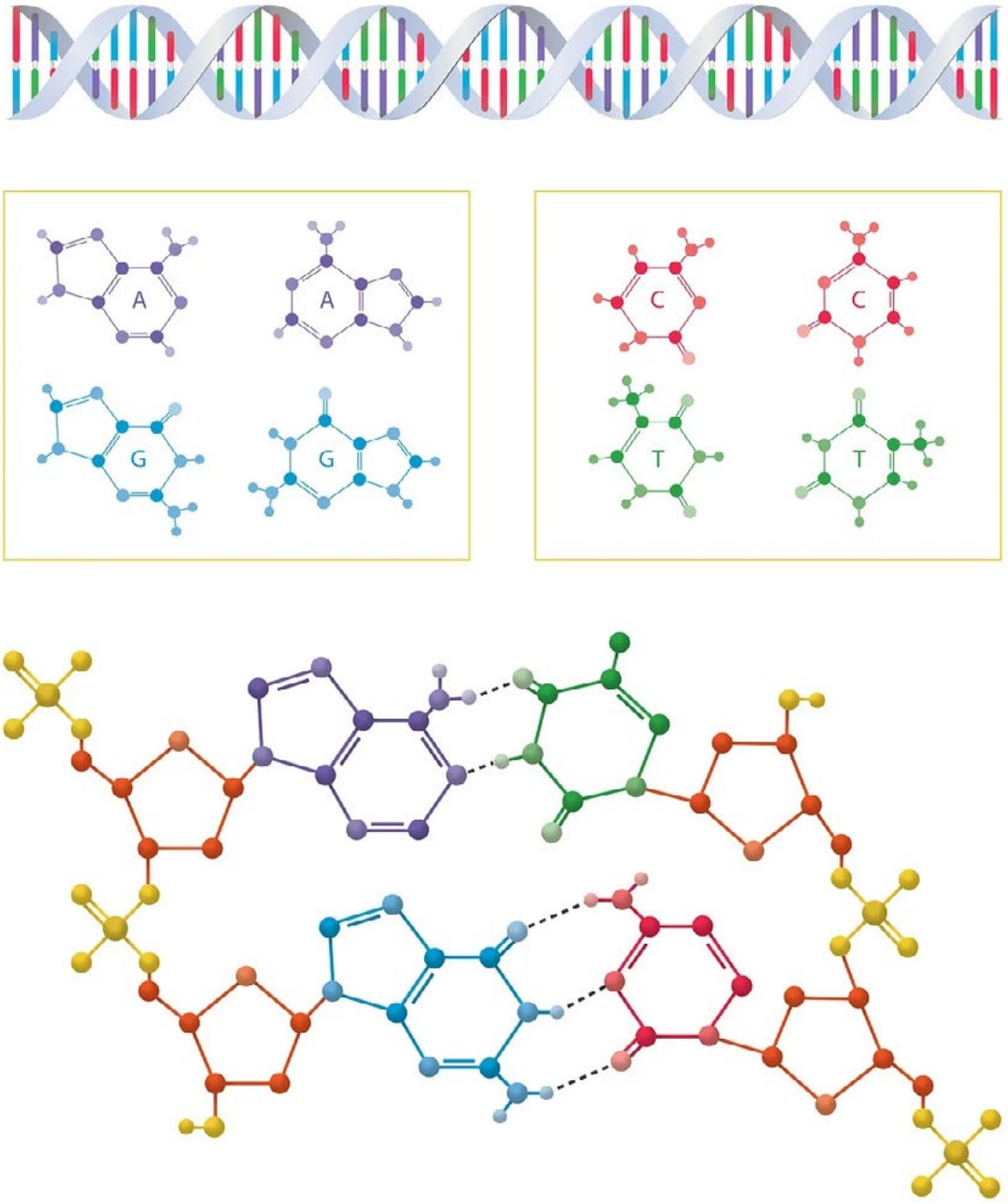


Fig. 1.1 Schematic drawing of specific base pairing within the DNA strand (source: https://pixabay.com/cs/).

Table 1.1 Comparison of Nucleosides and Nucleotides

Term	Description
Nucleoside	• Combination of a nitrogenous base and a pentose sugar without a phosphate group. • For ribose: *adenosine*, *guanosine*, *cytidine*, *uridine*. • For deoxyribose: *deoxyadenosine*, *deoxyguanosine*, *deoxycytidine*, *thymidine* (the prefix *deoxy-* is usually not used for thymidine).
Nucleotide	• Addition of a phosphate group to a nucleoside. • Nomenclature: *[nucleoside] + [number of phosphates] + phosphate*. • Examples: – For one phosphate: ∗ Ribose: *adenosine monophosphate (AMP)*, *guanosine monophosphate (GMP)*, *cytidine monophosphate (CMP)*, *uridine monophosphate (UMP)*. ∗ Deoxyribose: *deoxyadenosine monophosphate (dAMP)*, *deoxyguanosine monophosphate (dGMP)*, *deoxycytidine monophosphate (dCMP)*, *thymidine monophosphate (TMP)*. – For two phosphates: ∗ *adenosine diphosphate (ADP)*, *deoxyadenosine diphosphate (dADP)*.

Examples:

- ATP: Adenosine triphosphate – contains adenine, ribose, and three phosphate groups.
- dGTP: Deoxyguanosine triphosphate – contains guanine, deoxyribose, and three phosphate groups.
- CMP: Cytidine monophosphate – contains cytosine, ribose, and one phosphate group.
- dTTP: Deoxythymidine triphosphate – contains thymine, deoxyribose, and three phosphate groups.

This nomenclature principle ensures clear and consistent naming of various nucleotides, facilitating their identification and study in biochemistry and molecular biology. To simplify matters, we often refer to nucleotides only by their nitrogenous base component, such as “adenine nucleotide,” “guanine nucleotide,” and so on.

1.4 STRUCTURE AND FUNCTION OF DNA AND RNA

Molecular Structure of DNA: DNA (deoxyribonucleic acid) in its most common form is composed of two polynucleotide chains that coil into a double helix. These chains are antiparallel, meaning that one side of the DNA runs in the 5'→3' direction (from the five end to the three end) and the other side runs in the antiparallel 3'→5' direction. This is crucial for processes like replication and transcription. The primary function of DNA is to store genetic information. DNA carries the genetic information of an organism in the form of nucleotide sequences. These sequences are codes for the synthesis of proteins and the regulation of various cellular processes.

Molecular Structure of RNA: RNA (ribonucleic acid) is typically a single-stranded molecule, unlike the double helix of DNA. Uracil replaces thymine, which is present in DNA. Due to complementary base pairing within a single strand, RNA can form various structures, including hairpin loops and internal loops. The functions of RNA vary depending on the specific type of molecule (Table 1.2):

- **mRNA (messenger RNA):** mRNA carries genetic information from DNA to ribosomes, serving as a protein synthesis template.
- **tRNA (transfer RNA):** tRNA brings specific amino acids to the ribosome during translation, according to the codon sequence on the mRNA.
- **rRNA (ribosomal RNA):** rRNA is a major component of ribosomes, where it participates in protein synthesis. rRNA ensures the structure of the ribosome and catalyzes the peptide bond between amino acids.
- **Regulatory RNA:** RNA molecules, such as miRNA (microRNA) and siRNA (small interfering RNA), play key roles in regulating gene expression and protecting the genome from viruses and transposons.

Tab. 1.2 Comparison of DNA and RNA

Characteristic	DNA	RNA
Sugar	Deoxyribose	Ribose
Nitrogenous Bases	Adenine (A), Thymine (T), Cytosine (C), Guanine (G)	Adenine (A), Uracil (U), Cytosine (C), Guanine (G)
Structure	Double-stranded (double helix)	Single-stranded (various shapes and structures)
Stability	High stability (longer lifespan)	Lower stability (short-lived molecules)
Function	Stores genetic information, replication, transcription	Transfers genetic information, protein synthesis, gene expression regulation
Location in Cell	Nucleus (in eukaryotes), mitochondria, chloroplasts	Nucleus, cytoplasm, ribosomes

1.5 HISTORY OF THE DISCOVERY OF NUCLEIC ACIDS

19th Century:

1869: Swiss chemist Friedrich Miescher first isolated a substance from pus in surgical bandages (white blood cells) and named it "nuclein." This substance, later identified as DNA, was found to contain phosphorus and to be in the cell nucleus. **1880s-1890s:** Walther Flemming's work on chromosomes and cell division (mitosis) provided early evidence that nuclein (nucleic acids) was a component of chromosomes identified as carriers of hereditary material. However, their exact role was not yet clear.

20th Century:

In **1928**, Frederick Griffith discovered the phenomenon of transformation, demonstrating that genetic information could be transferred between bacteria. **1944**: Oswald Avery, Colin MacLeod, and Maclyn McCarty identified DNA as the molecule responsible for carrying genetic information, confirming Griffith's earlier findings. In **1950**, Erwin Chargaff found that the ratios of adenine (A) to thymine (T) and cytosine (C) to guanine (G) are always approximately 1:1, which was crucial for understanding base pairing in DNA.

Finally, in **1953**, James Watson and Francis Crick proposed the iconic double helix structure of DNA based on X-ray crystallography data produced by Rosalind Franklin and Maurice Wilkins. Later, only Watson, Crick, and Wilkins were jointly awarded the Nobel Prize for their original model of the molecular structure of nucleic acids and its significance for information transfer in living material. Unfortunately, Rosalind Franklin, whose X-ray crystallography work was crucial to the discovery, was not included. She passed away in 1958, and the Nobel Prize was not awarded posthumously. Her contributions are now widely recognized as foundational to understanding DNA's structure.

21st Century:

The development of advanced technologies, including DNA sequencing and CRISPR, has revolutionized our understanding of nucleic acids. These advancements have significantly enhanced their applications in medicine and biotechnology, deepening our knowledge of DNA's structure and functions.

1.6 DNA REPLICATION

DNA replication is the process of copying DNA before cell division, ensuring that each new cell receives a complete set of genetic information. This process is crucial for the growth and reproduction of organisms.

The first phase of replication is initiation, which begins at specific points called origins of replication. The enzyme helicase unwinds the double helix of DNA by breaking the hydrogen bonds between the bases, creating a replication fork. The unwound DNA is stabilized by single-strand binding proteins (SSBs).

The second phase is elongation, during which new DNA strands are synthesized using the original DNA as a template. The key enzyme here is DNA polymerase, which synthesizes a new DNA strand in the 5' → 3' direction. DNA polymerase can only add nucleotides to the free 3' end, causing differences in the synthesis of the two strands. **Leading strand**: Synthesized continuously in the direction of the replication fork. **Lagging strand**: Synthesized discontinuously as a series of short fragments called Okazaki fragments because it must be synthesized in the direction opposite to the replication fork.

The entire process is concluded by termination. Elongation primers are removed, and the Okazaki fragments are joined by DNA ligase, creating a continuous DNA strand. DNA polymerase can be proofread during DNA synthesis. If an incorrect nucleotide is added, the polymerase removes and replaces it with the correct one.

The result of replication is two identical double helices of DNA, each consisting of one original (parental) strand and one newly synthesized (daughter) strand. This semiconservative mechanism ensures that each new cell receives an accurate copy of genetic information.

This process is highly regulated and ensures the accuracy and integrity of genetic information preservation during cell division (damage to regulation leads to cancerous growth).

1.7 ORGANIZATIONAL STRUCTURE OF NUCLEIC ACIDS

DNA has several levels of structure that together ensure its stability and functionality. These levels include primary, secondary, tertiary, and quaternary structures. Each level of structure plays a crucial role in storing and expressing genetic information.

1.7.1 PRIMARY STRUCTURE OF DNA

The primary structure of DNA refers to the linear order of nucleotides in the DNA chain. Nucleotides are linked by phosphodiester bonds, creating a stable linear backbone of DNA. The nucleotide sequence is written from the 5' end to the 3' end, for example, 5'-ATCG-3'.

Microsatellites can be considered as a specific example of primary structure. They are characterized by a basic repeating unit (e.g., nucleotides CG), repeat length (e.g., two nucleotides), number of repeats (e.g., 6), and total length (e.g., 12 nucleotides).

Example sequence:

5'- AGCTAGGTAGATCGAA**CGCGCGCGCGCG**ATTGCATCGAGCCAT – 3'

Microsatellites are used as genetic markers due to their specific characteristics, such as high length variability across individual genomes. They are also called **SSR markers** (Simple Sequence Repeats). For detailed information, see the chapter on genetic markers.

The significance of Primary Structure is in (1) Genetic Information: The order of nucleotides (sequence) in DNA encodes genetic information, which is essential for protein synthesis and regulation of cellular processes. **(2) Transcription and Translation**: The primary structure of DNA is the foundation for the formation of mRNA during transcription and subsequent protein synthesis during translation. **(3) Mutations**: Changes in the nucleotide sequence can lead to mutations, affecting the structure and function of proteins and having significant biological consequences.

Primary structure is determined by sequencing, which involves reading the individual types of nucleotides in the DNA chain. Different sequencing methods include **Sanger Sequencing**, based on chain termination using dideoxynucleotides. It is used for sequencing short DNA segments. **Next-Generation Sequencing (NGS)**: Modern sequencing methods allow the sequencing of large genomes quickly and efficiently. **Nanopore Sequencing**: A technique that enables the reading of long chains of DNA or RNA by passing individual molecules through nanopores.

1.7.2 SECONDARY STRUCTURE OF DNA (CONFORMATION)

The secondary structure of DNA refers to the specific arrangement of two polynucleotide chains into a double helix. This structure results from specific base pairing and other stabilizing interactions.

There are three main forms of DNA secondary structure: B-DNA, A-DNA, and Z-DNA, which can vary depending on nucleotide sequence, physiological conditions, and interactions with proteins and other molecules. Each form has its specific properties and biological significance (Table 1.3).

A significant characteristic of the different forms of DNA secondary structure is the major and minor grooves, which arise from the geometry of base pairing and are crucial for interactions with proteins and other molecules. The different accessibility and structure of the grooves allow specific binding of regulatory proteins, enzymes, and other molecules involved in processes such as replication, transcription, and DNA repair. **Major Groove**: Wider and deeper, allowing specific interactions with regulatory proteins, enzymes, and other molecules. It is key for regulating gene expression, replication, and DNA repair. **Minor Groove**: Narrower and shallower, less accessible, but still important for binding specific molecules and supporting structural stability of DNA.

Table 1.3 Comparison of various secondary DNA structures

Feature	B-DNA	A-DNA	Z-DNA
Structure	Most common form of DNA in living cells; right-handed double helix.	Right-handed double helix, but differs slightly from B-DNA.	Left-handed double helix, significantly different from B-DNA and A-DNA.
Parameters	• Bases per turn: 10.5 • Rise per turn: 3.4 nm • Helix diameter: 2.0 nm	• Bases per turn: 11 • Rise per turn: 2.8 nm • Helix diameter: 2.3 nm	• Bases per turn: 12 • Rise per turn: 4.5 nm • Helix diameter: 1.8 nm
Grooves	• Major groove: Wide and deep, allows access for regulatory proteins. • Minor groove: Narrow and shallow.	• Major groove: Narrow and deep. • Minor groove: Wide and shallow.	Does not have distinct major and minor grooves like B-DNA and A-DNA.
Occurrence and Significance	• Occurrence: Dominant form under physiological conditions (high humidity, neutral pH). • Significance: Optimal for replication and transcription due to easy access for enzymes and regulatory proteins.	• Occurrence: Forms under low humidity or in a dehydrated environment; also found in DNA-RNA hybrids and double-stranded RNA. • Significance: Less accessible to regulatory proteins, which can affect gene expression; stabilizes DNA under specific conditions.	• Occurrence: Found in regions with high GC content, supercoiling, specific ions, or chemical modifications. • Significance: Plays a regulatory role in gene expression by changing DNA topology and affecting protein binding.

1.7.3 TERTIARY STRUCTURE OF DNA

The tertiary structure of DNA refers to a higher level of organization, where the double helix of DNA forms more complex structures to fit into the cell nucleus or other cellular organelles. This structure, formed by additional double helix winding, is called supercoiling.

Supercoiling can be positive or negative. In living cells, most DNA is in a negatively supercoiled state. Supercoiling is important for regulating access to DNA during transcription and replication.

At the tertiary structure level, DNA is organized into chromatin. This occurs when DNA wraps around specific proteins (histones) to form nucleosome structures. A nucleosome is the basic unit of chromatin and consists of DNA wrapped around eight histone proteins (a histone octamer). Each nucleosome includes 147 DNA base pairs (about 1.65 turns of the DNA molecule).

Nucleosomes link together to form chromatin fibers, allowing further condensation of DNA. Chromatin can be euchromatin (less condensed, active in transcription) or heterochromatin (highly condensed, inactive).

1.7.4 QUATERNARY STRUCTURE OF DNA

The quaternary structure of DNA refers to the highest level of organization, where tertiary structures of DNA further fold into even more complex structures. This includes the organization of chromatin fibers into chromosomes and their interactions with other molecules and structures in the cell. The quaternary structure is crucial for the precise division of genetic material during cell division, gene expression regulation, and DNA protection from damage.

Interactions with other molecules include protein-DNA interactions. Interacting proteins include enzymes (e.g., DNA polymerase, RNA polymerase), regulatory proteins (e.g., transcription factors), and structural proteins. The quaternary structure also includes RNA-DNA interactions (rRNA, tRNA, miRNA), which play key roles in protein synthesis and regulation of gene expression.

1.7.5 CHROMATIN

Chromatin is a complex of DNA and proteins found in the nuclei of eukaryotic cells. It is crucial for packaging long DNA molecules into a compact structure that fits within the cell nucleus while also enabling the regulation of gene expression and DNA replication. The basic unit of chromatin is the nucleosome, which consists of approximately 147 base pairs of DNA wrapped around a histone octamer. The octamer contains two molecules, each of histones H2A, H2B, H3, and H4. Histone H1 (which may or may not be present depending on the level of chromatin organization) links nucleosomes and helps stabilize their structure.

Functionally, chromatin is involved in DNA condensation, regulation of gene expression, replication, and repair. Chromatin exists in cells in two primary types: euchromatin and heterochromatin, which differ in structure and function. Euchromatin is a less condensed form of chromatin. This region of the genome is actively transcribed and contains often expressed genes. Due to its open structure, euchromatin is accessible to transcription factors and RNA polymerases, facilitating easy access to DNA and supporting gene expression.

Conversely, heterochromatin is a more condensed form of chromatin. This region of the genome is genetically inactive and contains typically silenced genes. Heterochromatin is further divided into two categories: constitutive heterochromatin and facultative heterochromatin. Constitutive heterochromatin is permanently condensed and includes regions such as centromeres and telomeres. Facultative heterochromatin, on the other hand, changes its condensation state depending on cell type and developmental stage, meaning its structure and function can be regulated according to cellular needs.

Post-translational modifications of histones (e.g., acetylation, methylation, phosphorylation, ubiquitination) affect chromatin structure and gene expression regulation. Histone acetylation is typically associated with gene activation. In contrast, histone methylation can lead to either activation or silencing of genes depending on the specific site and degree of methylation (an epigenetic modification).

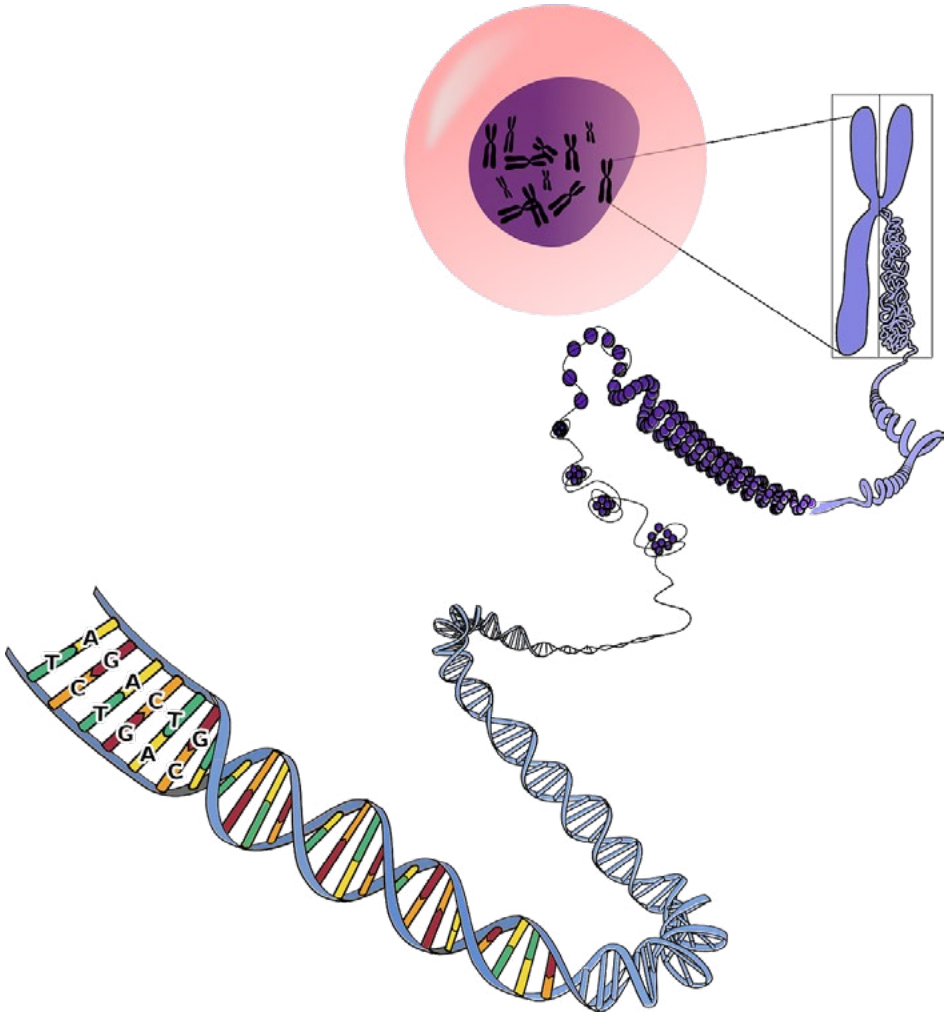


Fig. 1.2 Schematic depiction of structural levels in genetic information (source: https://pixabay.com/cs/)

1.7.6 ORGANIZATIONAL LEVELS OF CHROMATIN

Nucleosome Fiber (10 nm Fiber): This is a string of nucleosomes connected by DNA, resembling beads on a string. This structure is the least condensed form of chromatin and is often associated with actively transcribed DNA.

Solenoid (30 nm Fiber): Nucleosomes are coiled into a more condensed structure, forming the 30 nm fiber. The presence of histone H1 facilitates the solenoid structure, which links nucleosomes and helps maintain their arrangement. The solenoid represents a higher level of DNA condensation, aiding the organization of DNA into a more compact form within the nucleus.

Chromosome: For cell division, chromatin is further condensed into the structure of chromosomes, which are visible under a microscope (size 1-10 μm). This condensation level is essential for accurately distributing genetic material to daughter cells during cell division. Chromosomes comprise chromatin fibers made up of DNA and proteins, primarily histones. Chromatin fibers are of two types: **Euchromatin**: This part of the chromosome is less condensed and contains most of the genes active in transcription. **Heterochromatin** is more condensed and often includes less active or completely inactive genes.

Chromosomes have a characteristic shape visible during cell division (mitosis and meiosis). Each chromosome consists of two chromatids joined by a centromere. The ends of chromosomes are called telomeres, which play a key role in protecting DNA from degradation and fusion with other chromosomes.

Function of Chromosomes The main function of chromosomes is to transfer genetic information from generation to generation. Genes, segments of DNA encoding specific proteins, are arranged linearly on chromosomes. Each gene has a particular location (locus) on the chromosome. During cell division (mitosis), chromosomes replicate and are evenly distributed to daughter cells, ensuring that each new cell contains the same genetic information as the parent cell. In sexual reproduction (meiosis), recombination and reductional division occur, producing gametes (sperm and eggs) with half the number of chromosomes, which is essential for genetic variability.

Homologous Chromosomes Homologous chromosomes are pairs of chromosomes with the same structure and carry the same genes but may contain different alleles or variants of these genes. These chromosomes come from each parent—one from the mother and one from the father. During meiosis, the process of forming gametes, homologous chromosomes pair up and undergo genetic recombination (crossing over), contributing to offspring variability. Conversely, heterologous chromosomes are chromosomes that are not part of the same pair and have different gene sequences.

Number of Chromosomes in a Cell The number of chromosomes varies between species and can range from a few to hundreds of chromosomes. This number is characteristic of each species. The different number of chromosomes between species is one of the intrinsic barriers to interspecies crossing. In somatic cells, chromosomes occur in pairs. The chromosome number (2n) indicates the number of chromosomes in a diploid cell, which is the typical set of chromosomes in somatic cells of a given species (examples of various organisms listed in Table 1.4).

Gymnosperms typically have 22 or 24 chromosomes, with exceptions like *Pseudotsuga menziesii* (2n = 26) and *Pseudolarix amabilis* (2n = 44), the latter being a polyploid species. Polyploids have multiples of the haploid number, such as tetraploids (4n), hexaploids (6n), and higher. For instance, *Sequoia sempervirens* is a notable hexaploid with 66 chromosomes.

Chromosome numbers and ploidy levels vary more among angiosperms than gymnosperms, with most species having X values between 10 and 20. Genera like *Populus* and *Eucalyptus* are mostly diploids, whereas *Salix*, *Betula*, and *Acacia* include many polyploid species.

Pine and spruce chromosomes are typically metacentric and similar in length, often lacking distinct landmarks for easy identification in cytogenetic studies. Most northern hemisphere conifers have 24 chromosomes (12 chromosome pairs). These chromosomes are large and easily studied under a light microscope, unlike the smaller chromosomes of hardwoods. Genes are positioned linearly on chromosomes at specific loci. Genes with multiple forms at the same locus are **alleles**. Homozygosity refers to individuals with two identical alleles (**AA** or **aa**), while **heterozygosity** involves one dominant and one recessive allele (**Aa**). Multiple alleles can exist within a population, leading to polymorphism. Codominant alleles, like those in human ABO blood groups, display distinct phenotypes for each genotype.

All genes on the same chromosome belong to the same **linkage group**. In diploid organisms, homologous chromosomes appear in pairs, containing the same loci in the same order. During **mitosis**, somatic cell division ensures that daughter cells receive identical chromosomes. **Meiosis**, involved in gamete formation, halves the chromosome number and allows for gene recombination, producing four haploid cells. Fertilization restores the original chromosome number. In conifers like ***Picea abies*** and ***Pinus sylvestris***, with 12 chromosome pairs, numerous recombinant combinations are possible.

Tab. 1.4 Chromosome numbers for various organisms.

Organism	Chromosome Number (2n)
Corn (*Zea mays*)	20
Tomato (*Solanum lycopersicum*)	24
Apple (*Malus domestica*)	34
Scots Pine (*Pinus sylvestris*)	24
Norway Spruce (*Picea abies*)	24
Silver Fir (*Abies alba*)	24
English Oak (*Quercus robur*)	24
European Beech (*Fagus sylvatica*)	24
Ash (*Fraxinus excelsior*)	46
Sycamore Maple (*Acer pseudoplatanus*)	52
Small-leaved Lime (*Tilia cordata*)	82
Human (*Homo sapiens*)	46
Domestic Dog (*Canis lupus familiaris*)	78
Domestic Cat (*Felis catus*)	38
Domestic Cow (*Bos taurus*)	60
Fruit Fly (*Drosophila melanogaster*)	8
Red Deer (*Cervus elaphus*)	68
Roe Deer (*Capreolus capreolus*)	70
Wild Boar (*Sus scrofa*)	36
Common Pheasant (*Phasianus colchicus*)	82
Mallard Duck (*Anas platyrhynchos*)	80

Karyotype A karyotype is the complete set of chromosomes in a single cell of an organism. It is typically represented as a microphotograph of chromosomes arranged by size and shape. In humans, the karyotype normally comprises 22 pairs of autosomes and one pair of sex chromosomes (XX in women and XY in men).

Examples of karyotypes:

- Male: 46, XY
- Female: 46, XX
- Drosophila melanogaster (male): 8, XY
- Drosophila melanogaster (female): 8, XX

Regarding sex chromosomes, it is important to note that plants do not have specific sex chromosomes like animals. In plants, sex is usually not determined by specialized sex chromosomes (like X and Y chromosomes in mammals) but rather by genetic regulation and hormonal signals that influence the development of sex organs. For example, in Norway, spruce (Picea abies) is a monoecious plant, meaning that an individual produces both male (pollen) and female (seed) cones on the same plant. Therefore, in Norway spruce, there are no sex chromosomes as such, and all chromosomes in its karyotype are autosomes that carry the genetic information necessary for the growth and development of the plant.

1.7.7 DISCOVERY OF CHROMOSOMES

The history of chromosome discovery dates to the 19th century when scientists began using microscopes to study cellular structures. Chromosomes were first discovered and described by the German biologist Walther Flemming in 1879. Before Flemming, chromosomes had been observed under the microscope, but their significance and role were not properly understood or detailed.

Thomas Hunt Morgan demonstrated that genes are arranged linearly on chromosomes and that chromosomes are carriers of genetic information. His discovery was based on his experiments with fruit flies (Drosophila melanogaster) in the early 20th century. For his discoveries in genetics, Thomas Hunt Morgan was awarded the Nobel Prize in Physiology or Medicine in 1933.

1.8 LOCALIZATION OF DNA IN THE CELL

In eukaryotic cells, there are three main types of DNA: nuclear DNA, mitochondrial DNA (mtDNA), and chloroplast DNA (cpDNA).

Nuclear DNA: Most of the DNA in eukaryotic cells is stored in the nucleus, surrounded by the nuclear membrane. DNA in the nucleus is organized into chromatin, further divided into euchromatin (less condensed and transcriptionally active) and heterochromatin (more condensed and transcriptionally inactive). During cell division, DNA condenses into chromosomes, visible under a microscope.

In eukaryotic organisms, nuclear DNA is inherited **biparentally**, meaning that the offspring receive half of their genetic information from the mother (egg) and half from the father (sperm) during the formation of gametes (eggs and sperm), recombination of homologous chromosomes occurs, increasing genetic variability in the offspring.

Mitochondrial DNA (mtDNA): Mitochondria contain their DNA, which is called mitochondrial DNA (mtDNA). Mitochondrial DNA is circular and has its genome encoding some proteins necessary for mitochondrial function. Mitochondrial DNA is independent of nuclear DNA and replicates autonomously.

Mitochondrial DNA is almost exclusively inherited from the mother. This mode of inheritance is known as **maternal inheritance (uniparental)**. However, studies have shown that in some conifers (e.g., *Sequoia sempervirens*), paternal inheritance of mtDNA is likely (Neale et al. 1986).

Chloroplast DNA (cpDNA): Chloroplasts, found in plant cells, contain their DNA, called chloroplast DNA (cpDNA). Chloroplast DNA is circular and encodes genes essential for photosynthesis and other chloroplast functions. Chloroplast DNA replicates independently of nuclear DNA and is crucial for photosynthetic processes. Chloroplast DNA is often inherited **uniparentally**. In most gymnosperms, especially conifers (*Pinaceae*), chloroplast DNA (cpDNA) is paternally inherited, meaning cpDNA is transmitted through pollen grains (male gametes). Most angiosperms inherit cpDNA maternally, meaning cpDNA is derived from the egg cell (the male gamete does not contribute cpDNA).

Plasmid DNA: Plasmid DNA, or plasmids, are small circular DNA molecules found outside the chromosomal DNA in cells, most commonly in bacteria. These molecules are known for their ability to replicate autonomously, meaning they can copy themselves independently of the chromosomal DNA. In biotechnology, plasmids have significant importance. They are often used as vectors for gene cloning, where they serve to insert and clone foreign genes in genetic engineering. Plasmids are usually modified to enhance their function. They contain specific sequences where plasmid replication begins and selectable markers, such as antibiotic resistance genes, allowing the selection of cells containing the plasmid. Additionally, plasmids contain multiple cloning sites (MCS) and short DNA sequences with multiple restriction enzyme sites, facilitating the insertion of foreign DNA.

The endosymbiotic hypothesis posits that mitochondria and chloroplasts originated from proteobacteria and cyanobacteria. These bacteria formed a symbiotic relationship with a host cell, providing advantages such as improved energy production. Over time, extensive gene transfer from these organelles to the host's nuclear genome occurred, reducing their genetic autonomy but integrating their functions into the eukaryotic cell (Sato, 2019). Endosymbiotic theories of organelles revisited: retrospects and prospects. Springer Nature).

1.8.1 TYPES OF CHEMICAL BONDS IN NUCLEIC ACIDS

Several types of chemical bonds ensure their stability and functionality in nucleic acids such as DNA and RNA. The fundamental types of bonds are covalent phosphodiester bonds and weaker hydrogen bonds. Although hydrogen bonds are weaker than covalent bonds, their cumulative effect along the entire DNA or RNA molecule length provides significant stability. It is crucial for maintaining the double helix of DNA and the secondary structures of nucleic acids.

Phosphodiester Bond: Phosphodiester bond is a covalent bond that links adjacent nucleotides in a DNA or RNA strand. This bond forms between the 5'-phosphate group of one nucleotide and the 3'-hydroxyl group of the adjacent nucleotide. Phosphodiester bonds create the backbone of nucleic acids, consisting of alternating sugar (deoxyribose in DNA and ribose in RNA) and phosphate groups.

Hydrogen Bonds: Hydrogen bonds form between complementary bases of two DNA strands (or, in some cases, RNA) and are key for stabilizing the DNA double helix. In DNA, adenine (A) pairs with thymine (T) through two hydrogen bonds, and cytosine (C) pairs with guanine (G) through three hydrogen bonds. In RNA, adenine (A) pairs with uracil (U) instead of thymine.

Glycosidic Bond: A glycosidic bond is a covalent bond between the sugar (deoxyribose or ribose) and a nitrogenous base (adenine, thymine, cytosine, guanine, or uracil). This bond forms between the 1'-carbon of the sugar and a nitrogen atom on the base (N1 for pyrimidines and N9 for purines).

Van der Waals Forces and Hydrophobic Interactions: Van der Waals forces and hydrophobic interactions also contribute to stabilizing nucleic acids, particularly within the structure of the DNA double helix. These interactions are weaker than covalent and hydrogen bonds but play an important role in maintaining the correct three-dimensional structure.

Ionic Interactions: Ionic interactions are especially important in nucleic acids interacting with proteins and other molecules. For example, positively charged ions (such as Mg^{2+}) can stabilize nucleic acids' negatively charged phosphate backbone.

These bonds work together to create a stable and functional structure of nucleic acids, enabling replication, transcription, and other key biological processes. Although hydrogen bonds in nucleic acids are individually weak, they are collectively crucial for the stability and function of genetic material in living organisms.

1.9 RNA

RNA (ribonucleic acid) is a crucial molecule involved in various biological processes, particularly in the coding, decoding, regulating, and expressing genes. There are several types of RNA, each serving different functions. Here are the main types:

Messenger RNA (mRNA) carries genetic information from DNA to the ribosome, serving as a protein synthesis template. In this process, mRNA is transcribed from a DNA template in the nucleus. Once transcribed, it travels from the nucleus to the ribosome in the cytoplasm. The mRNA sequence is read and translated at the ribosome into a specific protein, following the genetic instructions encoded in the mRNA. This translation process is essential to produce proteins, which perform many functions within the cell. Before the RNA molecule is released from the nucleus, it undergoes processing mainly involving excising introns - short DNA sequences interrupting the coding regions (exons).

Transfer RNA (tRNA) brings amino acids to the ribosome to be added to a growing polypeptide chain during protein synthesis. Each tRNA molecule has an anticodon that pairs with a complementary codon on the mRNA, ensuring that the correct amino acid is incorporated into the protein. This precise matching of anticodon to codon is crucial for accurately translating the genetic code into functional proteins, as it guarantees that amino acids are added in the proper sequence dictated by the mRNA.

Ribosomal RNA (rRNA) forms the core of the ribosome's structure and catalyzes protein synthesis. rRNA molecules combine with proteins to form ribosomes, the sites of protein synthesis. Within the ribosome, rRNA ensures the proper alignment of the mRNA and tRNAs and catalyzes the formation of peptide bonds, which link amino

acids together to form a polypeptide chain. This catalytic activity is essential for translating genetic information into functional proteins, making rRNA a crucial component of the protein synthesis machinery.

Nuclear RNA (snRNA) is involved in pre-mRNA processing, particularly in splicing. snRNA molecules are components of the spliceosome, a complex responsible for removing introns from pre-mRNA and ligating exons to form mature mRNA. This splicing process is essential for gene expression, as it ensures that the mRNA contains the correct coding sequence to be translated into a functional protein. SnRNA plays a critical role in producing mature mRNA that accurately reflects the genetic information encoded in the DNA by removing non-coding regions and joining coding regions.

MicroRNA (miRNA) regulates gene expression by binding to complementary sequences on target mRNA, leading to its degradation or inhibition of translation. This regulation occurs post-transcriptionally, meaning it occurs after the mRNA has been synthesized from DNA. By binding to specific mRNA molecules, miRNAs can prevent the production of certain proteins, thus playing a critical role in fine-tuning gene expression. This post-transcriptional regulation by miRNAs affects various cellular processes, including development, differentiation, proliferation, and apoptosis, ensuring that genes are expressed at the right levels and times.

Small Interfering RNA (siRNA) is involved in the RNA interference (RNAi) pathway, where it interferes with the expression of specific genes. siRNAs can bind to complementary mRNA sequences and promote their degradation, effectively silencing particular genes. This process of gene silencing by siRNA is crucial for regulating gene expression and maintaining cellular homeostasis. By targeting and degrading specific mRNA molecules, siRNAs can prevent the production of proteins that might be harmful or unnecessary for the cell, thus playing a vital role in controlling various biological processes and defending against viral infections.

Long Non-Coding RNA (lncRNA) regulates gene expression at various levels, including chromatin modification, transcription, and post-transcriptional processing. lncRNAs can act as scaffolds, guides, decoys, or enhancers to influence gene expression, although their mechanisms are not fully understood. By serving these diverse roles, lncRNAs can modulate the structure of chromatin, interact with transcription factors, and influence the stability and translation of mRNA. This regulatory capacity allows lncRNAs to significantly impact various cellular processes, contributing to the complexity of gene expression regulation.

These different types of RNA work together to ensure the proper functioning of cellular processes, particularly the synthesis and regulation of proteins.

1.10 ENZYMES OF MAJOR IMPORTANCE IN MOLECULAR GENETICS

Molecular genetics relies on a wide range of enzymes that naturally occur in cells but, when used purposefully, allow controlled manipulation of DNA and RNA. Key enzymes in molecular genetics include polymerases, ligases, and nucleases, particularly restriction endonucleases. These enzymes are essential for polymerase chain reaction (PCR), DNA cloning, genome sequencing, and gene editing.

Polymerases

Polymerases are enzymes that catalyze the synthesis of DNA or RNA from nucleotides according to a template molecule. There are various polymerases, with DNA and RNA polymerases being the most significant in molecular genetics.

DNA Polymerases: These enzymes catalyze the synthesis of a new DNA strand using an existing DNA strand as a template. They play a crucial role in DNA replication, necessary for cell division and reproduction. DNA polymerases also repair damaged DNA.

RNA Polymerases: These enzymes synthesize RNA from a DNA template during transcription. RNA polymerases are essential for the formation of mRNA, which is subsequently translated into proteins, and for synthesizing other types of RNA, such as tRNA and rRNA.

Ligases

Ligases are enzymes that catalyze the formation of phosphodiester bonds between the 3'-hydroxyl group of one nucleotide and the 5'-phosphate group of another nucleotide. This process is key for joining DNA fragments. In molecular genetics, DNA ligases join DNA fragments during cloning, create recombinant DNA, and perform other genetic manipulations.

Nucleases:

Nucleases are enzymes that hydrolyze (cleave) phosphodiester bonds in DNA or RNA, leading to the degradation of these molecules. There are two main categories of nucleases: exonucleases and endonucleases.

Exonucleases: These enzymes remove nucleotides from the ends of DNA or RNA strands. They are useful for removing excess DNA or RNA and for various nucleic acid purification and preparation techniques. **Endonucleases**: These enzymes cleave phosphodiester bonds within DNA or RNA molecules. The most important endonucleases in molecular genetics are restriction endonucleases.

Restriction Endonucleases:

Restriction endonucleases, or restriction enzymes, are specialized proteins that recognize specific DNA sequences (restriction sites) and cleave the DNA at these sites. These enzymes were discovered in bacteria, serving as a defense mechanism against phage infections.

Principle of Restriction Endonucleases: These enzymes recognize specific, usually palindromic, DNA sequences and make precise cuts in the phosphodiester bonds at these sites. This specific cleavage allows scientists to cut DNA into precisely defined fragments, which is essential for techniques such as restriction mapping, DNA cloning, and restriction fragment length polymorphism (RFLP) analysis.

1.11 GENE EXPRESSION

Gene expression is the process by which a cell converts genetic information encoded in DNA into functional molecules such as proteins or various types of RNA. This process is crucial for all living organisms because it determines their phenotypic traits and ability to respond to environmental changes. Gene expression is a two-step process: the first phase is called transcription, and the second is translation.

Transcription

The first step in gene expression is transcription, where DNA is transcribed into messenger RNA (mRNA). This process occurs in eukaryotic cells' nucleus and prokaryotic cells' cytoplasm. The main enzyme involved in transcription is RNA polymerase.

Steps of Transcription:

Initiation: RNA polymerase binds to the promoter and unwinds a short section of the DNA double helix.

Elongation: RNA polymerase adds nucleotides complementary to the DNA template, synthesizing pre-mRNA.

Termination: Upon reaching a terminator sequence, RNA polymerase releases the newly formed RNA, and the DNA rewinds into a double helix.

The resulting pre-mRNA must undergo post-transcriptional modifications to become mature mRNA, which can then be transported to ribosomes for translation. These modifications include:

- **Addition of a 5' cap**: This modified nucleotide protects the mRNA from degradation and aids its transport to the cytoplasm.
- **Polyadenylation of the 3' end**: Adding a poly(A) tail increases mRNA stability and facilitates its export from the nucleus.
- **Splicing (removal of introns)**: Introns, which do not code for proteins, are removed, and exons are joined together, forming mature mRNA.

Some viruses can reverse-transcribe RNA back into DNA, a process called reverse transcription. Genetic laboratories also use this method to convert RNA into complementary DNA (cDNA) because DNA is more stable and easier to analyze.

Translation

Translation is the process of converting mRNA into a sequence of amino acids, leading to protein formation. This process occurs in ribosomes, free in the cytoplasm or attached to the endoplasmic reticulum.

Steps of Translation:

Initiation: The ribosome binds to the mRNA near the start codon (AUG).

Elongation: tRNA molecules bring amino acids to the ribosome, where peptide bonds link them into a growing polypeptide chain.

Termination: Upon reaching a stop codon (UAA, UAG, UGA), the polypeptide chain is released, and the ribosome disassembles.

Regulation of Gene Expression

The regulation of gene expression is crucial for cellular differentiation, organismal development, and adaptation to environmental changes. This process can occur at various levels. **Transcriptional Regulation** involves transcription factors that can increase or decrease the transcription of specific genes. **Post-transcriptional Regulation** affects mRNA stability and splicing, which can alter the amount and variants of produced mRNA. **Translational Regulation** controls the efficiency with which mRNA is translated into protein. **Post-translational Modifications** include protein modifications such as phosphorylation or glycosylation, which affect their function and lifespan. These mechanisms control gene expression, allowing cells to respond appropriately to internal and external stimuli.

Epigenetic Regulation

Epigenetic regulation involves changes in gene expression that are not caused by changes in the DNA sequence but can still be inherited. These changes mainly include DNA methylation and histone modifications.

DNA Methylation: Add methyl groups to cytosine bases in DNA, usually leading to gene silencing. Methylation can hinder the binding of transcription factors to DNA or attract proteins that silence gene expression. For Norway spruce, the study reveals that DNA methylation, particularly in gene bodies, plays a role in its ability to respond to environmental changes. Over 24 years, differentially methylated positions (DMPs) were identified in genes associated with acclimation and stress responses, such as those influencing phenology and cold stress. These changes suggest that methylation patterns may contribute to phenotypic plasticity (Chapter 5 on Geographic variation), aiding the spruce in adapting to contrasting climatic conditions. However, the overall stability of global methylation levels indicates resilience, with targeted changes in specific genes potentially facilitating long-term survival and ecological success (Heer et al., 2018).

Histone Modifications: Involve chemical modifications of histone proteins around which DNA is wrapped. These modifications, such as acetylation, methylation, or phosphorylation, affect the accessibility of DNA to transcription factors and thus regulate whether a gene is active or silent. Epigenetic changes allow for fine-tuned regulation of gene expression.

1.12 GENETIC CODE

The genetic code is a set of rules by which the information stored in DNA is translated into amino acid sequences of proteins. This process occurs through triplets of bases (codons) in mRNA. Each codon specifies one amino acid or a signal to stop protein synthesis.

Characteristics of the Genetic Code:

Redundancy (Degeneracy): Several codons can encode the same amino acid. For example, the amino acid leucine is encoded by six different codons (UUA, UUG, CUU, CUC, CUA, CUG). Degenerate codons usually differ in their third nucleotide position.

Universality: The genetic code is almost universal across all organisms, from bacteria to humans, with few exceptions.

Unambiguity: Each codon specifies only one amino acid.

Non-overlapping and Continuous: Codons are read in triplets without overlapping and without interruptions between them.

Start and Stop Codons

Start Codon (AUG): Signals the beginning of translation and determines where the ribosome synthesizes the protein. AUG encodes the amino acid methionine, the first amino acid in a newly synthesized polypeptide.

Stop Codons (UAA, UAG, UGA): Signal the end of translation. These codons do not correspond to any tRNA, causing the release of the polypeptide chain from the ribosome. Accurate recognition of stop codons is crucial for producing functional proteins. Mutations that change a stop codon to an amino acid (nonsense mutations) can lead to extended or nonfunctional proteins. The chemical characteristics of amino acids and their codons are listed in Table 1.5.

Table 1.5 Amino acids and their codons.

Chemical Characteristic	Amino Acid	Codons
Hydrophobic	Glycine	GGU, GGC, GGA, GGG
Hydrophobic	Alanine	GCU, GCC, GCA, GCG
Hydrophobic	Valine	GUU, GUC, GUA, GUG
Hydrophobic	Leucine	UUA, UUG, CUU, CUC, CUA, CUG
Hydrophobic	Isoleucine	AUU, AUC, AUA
Hydrophobic	Methionine	AUG
Hydrophobic	Proline	CCU, CCC, CCA, CCG
Polar Uncharged	Serine	UCU, UCC, UCA, UCG, AGU, AGC
Polar Uncharged	Threonine	ACU, ACC, ACA, ACG
Polar Uncharged	Cysteine	UGU, UGC
Polar Uncharged	Tyrosine	UAU, UAC
Polar Uncharged	Asparagine	AAU, AAC
Polar Uncharged	Glutamine	CAA, CAG
Polar Charged	Aspartate (Aspartic acid)	GAU, GAC
Polar Charged	Glutamate (Glutamic acid)	GAA, GAG
Polar Charged	Lysine	AAA, AAG
Polar Charged	Arginine	CGU, CGC, CGA, CGG, AGA, AGG
Polar Charged	Histidine	CAU, CAC
Aromatic	Phenylalanine	UUU, UUC
Aromatic	Tyrosine	UAU, UAC
Aromatic	Tryptophan	UGG
Special Function	Selenocysteine	UGA (in presence of SECIS)
Special Function	Pyrrolysine	UAG (in presence of PYLIS)
Special Function	Stop Codons	UAA, UAG, UGA

Amino Acids

Amino acids are the fundamental building blocks of proteins. Twenty standard amino acids combine into polypeptide chains during protein synthesis. Each amino acid has a distinctive structure, and chemical properties determine its function in proteins. Additionally, three special amino acids (selenocysteine, pyrrolysine, and N-formylmethionine) are found in certain organisms under specific conditions, bringing the total to 23.

Each amino acid has a central carbon atom, known as the alpha-carbon. Attached to this carbon are four groups: an amino group ($-NH_2$), a carboxyl group ($-COOH$), a hydrogen atom (H), and a variable side chain called the R-group. The R-group is specific to each amino acid and determines its chemical and physical properties.

The amino group ($-NH_2$) and carboxyl group ($-COOH$) are essential functional groups that enable amino acids to join into polypeptide chains through peptide bonds. When two amino acids join, the carboxyl group of one amino acid reacts with the amino group of another, forming a peptide bond and releasing a water molecule. This process, known as condensation, is fundamental for protein formation.

Each amino acid's variable side chain (R-group) determines its specific properties. For example, glycine, the simplest amino acid, has only a hydrogen atom as its R-group. Alanine has a methyl group ($-CH_3$) as its side chain. These differences in R-groups lead to diverse chemical properties of amino acids, affecting their behavior in proteins and their interactions with other molecules.

The general structure of amino acids with a central alpha-carbon, amino group, carboxyl group, and specific R-group is crucial for their role in biological systems. These basic building blocks form polypeptides, which further fold into complex protein structures essential for life.

Amino acids are classified into essential (which must be obtained from the diet) and non-essential (which can be synthesized by the body).

Bonds in Proteins (Peptides)

A peptide bond is a specific type of covalent bond that links amino acids into polypeptide chains, forming the primary structure of proteins. Peptide bonds are crucial for forming the primary structure of proteins.

A peptide bond forms through condensation between the carboxyl group (-COOH) of one amino acid and the amino group ($-NH_2$) of another. During this reaction, a water molecule (H_2O) is released.

Peptide bonds are synthesized during protein synthesis on ribosomes in a process called translation. The cleavage of peptide bonds occurs during proteolysis by enzymes called proteases. This process is important for regulating protein levels, removing damaged or misfolded proteins, and recycling amino acids.

1.13 CELL CYCLE

The cell cycle is a series of precisely regulated steps that lead to cell growth and division. This cycle is essential for organisms' growth, development, and maintenance. The cell cycle consists of four main phases: G1 phase, S phase, G2 phase, and M phase.

G1 Phase (First Growth Phase): The cell grows and synthesizes necessary proteins and organelles. This phase is critical for preparing the cell for DNA synthesis. A checkpoint at the end of the G1 phase determines whether the cell will enter the S phase or transition into a resting state (G0 phase).

S Phase (Synthesis Phase): During this phase, DNA replication occurs, ensuring that each daughter cell has a complete set of genetic information after division. DNA replication is tightly regulated to prevent mutations and errors.

G2 Phase (Second Growth Phase): The cell grows and prepares for mitosis. DNA repair and synthesis of proteins necessary for cell division occur during this phase. A checkpoint at the end of the G2 phase ensures that DNA is correctly replicated and the cell is ready for division.

M Phase (Mitotic Phase): The M phase includes the division of the nucleus (mitosis) and subsequent division of the cytoplasm (cytokinesis). Mitosis is further divided into prophase, metaphase, anaphase, and telophase. During these steps, chromosomes are precisely distributed between two daughter cells.

Regulation of the Cell Cycle: The cell cycle is controlled by a series of checkpoints that ensure each phase is completed correctly before the cell proceeds to the next phase. If errors or DNA damage are detected, checkpoints can halt the cycle to allow for repair or trigger cell death (apoptosis) if the damage is irreparable. The cell cycle is thus essential for the proper functioning and survival of organisms, enabling growth, tissue regeneration, and cell reproduction. Its precise regulation is crucial for maintaining genetic stability and preventing cancer-related diseases.

1.13.1 MITOSIS

Mitosis is the process of cell division that ensures the even distribution of chromosomes to two daughter cells and is critical for growth, regeneration, and asexual reproduction in organisms. This process consists of several phases: prophase, metaphase, anaphase, and telophase.

Prophase: Chromosomes condense and become visible as double-stranded structures called sister chromatids, joined by a centromere. The nuclear membrane begins to disintegrate. Centrosomes (organelles organizing microtubules) move to opposite poles of the cell, and the mitotic spindle, a structure of microtubules, begins to form, which will facilitate chromosome movement.

Metaphase: Chromosomes align at the cell's equatorial plane, the metaphase plate. Each chromosome is attached to spindle microtubules from both poles of the cell through kinetochores.

Anaphase: Sister chromatids separate and are pulled by the spindle microtubules to opposite cell poles. This movement ensures that each daughter cell receives an equal number of chromosomes.

Telophase: New nuclear membranes form around the separated sets of chromosomes at opposite cell poles. Chromosomes begin to decondense and return to their thread-like form. The mitotic spindle disassembles.

Cytokinesis: This process follows telophase and involves the division of the cell's cytoplasm into two daughter cells. A cleavage furrow forms in animal cells, pinching the cell in two. A cell plate forms in plant cells, eventually leading to a new cell wall between the daughter cells.

Mitosis ensures that each daughter cell receives an identical set of chromosomes as the parent cell, which is essential for genetic stability and proper organism function.

1.13.2 MEIOSIS

Meiosis is a specialized type of cell division that produces gametes (sex cells) with the chromosome number reduced by half (haploid cells). This process is crucial for sexual reproduction as it ensures genetic diversity and the correct chromosome number in offspring.

Meiosis consists of two consecutive divisions: meiosis I and II, each comprising several phases. In prophase I, a key process called crossing over occurs, where homologous chromosomes pair up and exchange genetic material, increasing genetic variability.

Following this, in metaphase I, bivalents (paired homologous chromosomes) align at the cell's equatorial plane; in anaphase I, homologous chromosomes separate; and in telophase I, two daughter cells form, each with half the original chromosome number. Meiosis II continues without another round of DNA replication, with prophase II, where chromosomes condense; metaphase II, where chromosomes align at the equatorial plane; anaphase II, where sister chromatids separate; and telophase II, which, along with cytokinesis, produces four genetically unique haploid cells.

This process ensures that each resulting gamete has a unique combination of genes. The outcome of meiosis is four haploid cells, each containing half the original number of chromosomes and a genetically unique set of genes due to crossing over and the independent assortment of homologous chromosomes.

Chromosome aberrations, such as deletions, duplications, inversions, and translocations, can cause meiotic irregularities, often leading to lethal gametes. Meiosis timing varies in some conifers; male meiosis occurs in autumn (e.g., *Juniperus*), autumn to spring (e.g., *Larix*), or spring (e.g., *Picea* and *Pinus*). In contrast, female meiosis typically happens in spring. In angiosperms like birch and oak, meiosis occurs in late summer or spring, respectively. Temperature extremes can cause injuries and irregularities during male meiosis.

Significance of meiosis to heredity

- Meiosis ensures that each gamete has a complete haploid set of chromosomes.
- Meiosis results in a random assortment of paternal and maternal chromosomes in gametes. Each pair of homologous chromosomes of the gametes receives the paternal chromosome of the pair, and half gets the maternal chromosome.
- Meiosis results in an independent assortment of chromosomes from different homologous pairs. This means that chromosomes originating from the paternal and maternal parents randomly assort gametes such that one gamete may receive a mix of paternal and maternal origin chromosomes. Therefore, genes on non-homologous chromosomes segregate independently (complies with the 3rd law of Mendel).
- Through crossing over, alleles at different loci on the same chromosome can recombine, further enhancing genetic diversity in offspring of sexually reproducing tree species.

1.13.3 BASIC INFORMATION ON TREES' CHROMOSOME CYTOLOGY

Chromosome cytology involves the microscopic examination of chromosome number, size, morphology, and behavior during nuclear division. Discovered in the mid-19th century, chromosomes became visible when stained with basic dyes. After the invention of light microscopy in the late 19th century, advances allowed for a detailed study of chromosomes, revealing key features of fertilization and nuclear division in somatic (**mitosis**) and germ cells (**meiosis**).

Cytogenetics also involves studying morphological differences among chromosomes through karyotype analysis. This includes examining metaphase chromosomes stained with special chemicals to identify variables like chromosome arm length and secondary constrictions. While karyotype analysis in forest trees has been limited, it has been conducted in species such as *Pinus*, *Picea*, *Pseudotsuga menziesii*, *Sequoiadendron giganteum*, and *Calocedrus decurrens*. Recent techniques like fluorescent in situ hybridization and confocal microscopy revitalize cytogenetics in forest trees.

Time of meiosis in trees

A summary of the timing of meiosis in various conifer genera and individual species with variability in this trait is provided in Table 1.6. Three main patterns are observed: (1) meiosis begins and completes in autumn, (2) meiosis starts in autumn and completes in spring, and (3) meiosis begins and completes in spring. The discovery that meiosis in pollen mother cells of the *Larix* genus starts in autumn led to a revision of the earlier belief that meiosis occurs exclusively in either autumn or spring. Further detailed studies on *Pseudotsuga menziesii*, *Thuja plicata*, and *Tsuga heterophylla* revealed that these species share the same timing pattern as *Larix*.

In most investigated species, both female and male meiosis occur in spring. However, in many *Larix* species, the timing of meiosis depends on weather conditions. Meiosis occurs in late summer in birch, hazel, and alder, while in elm, aspen, and oak, it happens in spring. Night length is likely the environmental trigger in species where meiosis begins in late summer or autumn. Conversely, in species where meiosis occurs primarily or entirely in spring, the timing is determined mainly by heat accumulation (heat sum).

Table 1.6 Time of meiosis in some conifer genera and species.

Meiosis starts and is completed during autumn	Meiosis starts during autumn and is completed during spring	Meiosis starts and is completed during spring
• *Cedrus* • *Cryptomeria* • *Juniperus chinensis* • *J. horizontalis* • *J. virginiana* • *Taxus*	• Meiosis of pollen mother cells in *Larix*, *Pseudotsuga*, *Thuja*, and *Tsuga* usually show this pattern	• *Abies* • *Athrotaxis* • *Cunninghamia* • *Juniperus communis* • *J. rigida* • *Keteleeria* • *Picea* • *Pinus* • *Pseudolarix* • Megaspore mother cells in *Larix* and *Taxus*

1.14 MUTATIONS AND TYPES OF MUTATIONS

Mutations are changes in the genetic material of an organism. These changes can affect individual genes, larger regions of DNA, or entire chromosomes. Mutations can result from errors during DNA replication, exposure to mutagens (such as chemicals or radiation), or they can occur spontaneously. In some cases, mutations can lead to new traits in an organism, while in others, they can be harmful or neutral.

Mutations play a crucial role in evolution because they provide genetic variability on which natural selection acts. Without mutations, the adaptation of organisms to changing environments would not be possible. On the other hand, mutations can also cause many genetic diseases and cancer.

Mutations can have various consequences depending on their type and location:

Neutral Mutations: Do not have a significant impact on the organism.

Harmful Mutations: Can cause genetic diseases, cancer, or reduce the survival and reproductive abilities of the organism.

Beneficial Mutations: Can provide evolutionary advantages and improve the organism's ability to survive and reproduce.

Types of Mutations

Mutations can be classified according to different criteria, such as their extent, cause, or impact on protein function. Table 1.7 shows the main types of mutations.

Tab 1.7 The Main types of mutations are sorted by their extent and cause.

Category	Details
By Extent: Gene Mutations (Point Mutations)	• **Substitution:** Replacement of one nucleotide with another. – **Silent Mutations:** A nucleotide change that does not alter the amino acid in the protein. – **Missense Mutations:** A nucleotide change that alters an amino acid in the protein. – **Nonsense Mutations:** A nucleotide change that creates a stop codon, leading to premature termination of protein translation. • **Indel Mutations:** Insertion or deletion of one or more nucleotides. – **Frameshift Mutations:** Insertion or deletion of a nucleotide that changes the reading frame of the gene, affecting all subsequent amino acids.
By Extent: Chromosomal Mutations	• **Deletion:** Loss of a chromosome segment. • **Duplication:** Duplication of a chromosome segment. • **Inversion:** Reversal of a chromosome segment by 180 degrees. • **Translocation:** Transfer of a chromosome segment to another chromosome.
By Extent: Genomic Mutations	• **Aneuploidy:** Change in the number of individual chromosomes (e.g., trisomy 21, which causes Down syndrome). • **Polyploidy:** Change in the total number of chromosome sets (common in plants).
By Cause: Spontaneous Mutations	Arise without external intervention, usually during DNA replication or repair.
By Cause: Induced Mutations	Arise because of exposure to mutagens such as chemicals, radiation, or viruses.

1.15 GENOME

The term genome refers to an organism's complete set of genetic material. It encompasses all the DNA, including all its genes (genic DNA) and non-coding sequences of the DNA (non-genic DNA). Eukaryotic DNA includes unique, single, and low-copy functional genes, found in only one or a few copies per haploid genome, and various types of repetitive or spacer DNA. In higher organisms, particularly conifers, non-coding DNA, often called "junk DNA," constitutes a significant portion of the genome. Over 99% of DNA in species like Norway spruce and Scots pine is likely of minor or no importance for adaptation. However, recent advances in genomics have increasingly challenged this idea. For example, advances in understanding transposable elements (TEs) in plant genomes have also shifted perceptions of non-coding DNA. These TEs, once considered "junk," are now known to contribute significantly to genetic variation and evolution. They play roles in gene regulation, genome structure, and stress responses, highlighting their importance beyond genetic filler (Ramakrishnan et al., 2022).

The retrotransposons found in maize have also been identified in pine and spruce species. In *Pinus elliottii* (slash pine), a type of retrotransposon known as Ty1-copia-like sequences has been studied extensively. Research has shown that these retrotransposons are distributed evenly across all 12 chromosomes of *Pinus elliottii*. These elements represent a significant part of the genome of this species, contributing to its genomic complexity and large size. This distribution pattern was detailed by Kamm et al. (1996) and further supported by subsequent research indicating a high copy number and dispersed presence across the chromosomes, though with reduced hybridization in centromeric regions (Voronova et al., 2014).

1.15.1 GENOME SIZE

The gene order appears highly conserved within genera (e.g., *Pinaceae*) and across wide taxonomic families, regardless of genome size. This conservation aids in gene mapping and enables the development of various molecular markers for breeding purposes. The total amount of nuclear DNA determines the genome size for a species, generally reported as the C-value. Genome sizes vary greatly among plant species, with angiosperm trees like *Populus* and *Eucalyptus* having significantly smaller genomes than gymnosperms like *Pinus*, which has at least 1×10^{10} nucleotide bases in its nuclear genome. This discrepancy leads to the C-value paradox, where genome size does not correlate with the organism's apparent complexity. Research explores why plants, particularly gymnosperms, possess extensive DNA, suggesting adaptation to environmental stresses as a possible reason. Specifically, pine species have some of the largest genomes among plants, often exceeding 20 Gb, one of the largest known plant genomes.

In contrast, Eucalyptus species have much smaller genomes, around 0.64 Gb. This significant difference highlights the unique evolutionary paths and genomic architectures between these two types of plants. Due to their large genomes, pine species like *Pinus taeda* (loblolly pine) have a high number of genes, estimated to be around 50,000 functional genes (Ritland, 2012). This high gene count is reflective of the large genome size and the presence of many gene duplications and repetitive elements. The genome of Eucalyptus grandis contains about 36,000 to 40,000 functional genes. This number is lower than that of pine species but still substantial, reflecting its complex genetic makeup.

DNA content estimation in conifers, based on Feulgen cytophotometry, shows they have some of the largest genomes among higher plants. DNA content can also vary among trees within the same species, with some studies indicating increased DNA content with increasing latitude. This variation supports the hypothesis that increased DNA content in conifers might be an adaptive response to stressful environments. Genome size plays a significant role in plant evolution, with differences among species and individuals being at least partly adaptive.

Number of functional genes in plants

Estimating the number of functional genes in the genome of a conifer is challenging due to their large and complex genomes. However, comprehensive DNA sequencing projects have provided insights. For example, the genome of *Pinus taeda* (loblolly pine) has been sequenced, revealing it contains a vast number of genes, but exact functional gene counts are still being refined.

Arabidopsis thaliana, *Oryza sativa* (rice), and *Populus trichocarpa* (poplar) have approximately 25,000, 50,000, and 45,000 genes, respectively. *Populus trichocarpa* is believed to be a recently formed tetraploid. Conifers, possibly ancient polyploids, are thought to have around 50,000 functional genes due to their large genome sizes and gene duplication events.

The genome of *Eucalyptus grandis*, one of the most studied species of *Eucalyptus*, has been extensively sequenced and analyzed. The Eucalyptus grandis genome is estimated to contain around 36,000 to 40,000 functional genes. This figure is based on comprehensive genome sequencing and annotation efforts that have identified and characterized thousands of genes across various gene families.

Comparative genomic analyses have revealed that *Eucalyptus* species have a significant number of gene families, with about 27,032 genes grouped into these families in the hybrid species *E. urophylla* × *E. grandis*, accounting

for a large portion of its predicted genes. These analyses also show the presence of many unique and conserved gene families, indicating a complex and diverse genetic structure.

These estimates are based on comprehensive sequencing efforts and comparisons with other plant genomes, illustrating conifer genomes' complexity and long evolutionary history.

The large genome size and high gene number in pines may contribute to their long life cycles, adaptation to diverse environments, and resilience to various stresses. Eucalyptus species, with their smaller but still complex genomes, show significant gene family diversification and specific adaptations that make them dominant in certain ecological niches, especially in terms of wood production and secondary metabolite synthesis.

1.16 GENETIC MARKERS

Genetic markers are crucial for many types of genomic studies in forest trees. Genetic markers are often used in natural populations of forest trees to determine the presence and patterns of genetic variation, to understand mating systems, and to account for inbreeding. Moreover, genetic markers allow studying taxonomic and phylogenetic relationships among species. In more practical applications, genetic markers are routinely used to monitor the efficiency of various tree improvement programs. They are indispensable for constructing genetic maps and for multiple types of marker-assisted breeding.

Specifically, genetic markers are heritable traits that can identify and track genetic differences among individuals, populations, or species. Genetic markers should be heritable and stable across generations, often meeting the principles of Mendelian inheritance. However, specific situations exist where this is unmet (e.g., mitochondrial and chloroplast DNA inheritance, inherited from only one parent, or epigenetic markers influenced by environmental factors).

Genetic markers can be morphological, biochemical, or molecular. Nowadays, molecular markers are almost exclusively used; therefore, some sources (incorrectly) equate the term molecular marker with genetic marker.

1.16.1 TYPES OF GENETIC MARKERS

Morphological Markers

Definition: Visible physical characteristics of genetically determined organisms can be observed without molecular techniques.

Examples: Flower color, leaf shape, morphological characteristics of bark. In clonal seed orchards (see chapter on tree improvement), morphological markers such as cone size, cone shape, and seed weight/shape are easily observed among clones.

Advantages: Easy identification without laboratory equipment, quick and inexpensive method.

Disadvantages: Phenotypes can be influenced by the environment, often polygenic inheritance, making it unclear how genotypes relate to the "value" of the trait. Only a limited number of available morphological markers exist in forest trees, often originating from point mutations (e.g., needle albinism, dwarfism, etc.). These mutations generally lack production significance and frequently have negative or even lethal effects on individuals. The number of morphological markers is limited and usually associated with a deleterious phenotype. Nevertheless, seedling mutant characters have been used to study mating systems in several conifer species (e.g., *Pinus ponderosa*).

Biochemical Markers

Definition: Proteins or enzymes whose variations can be detected by biochemical techniques such as electrophoresis or chromatography.

Several biochemical markers exist and have been extremely valuable for studying tree genetics. **Monoterpenes** (substances in plant resins and oils) were the first biochemical markers in trees and have been mostly used for taxonomic studies. The small number of monoterpene markers and their dominant expression have limited their other applications.

Studies based on **allozymes**, the most widely used genetic markers in the 1980s and 1990s, have contributed greatly to our knowledge of the population genetics of forest trees. Allozymes are accessible and are highly polymorphic and codominant. They have been used to describe patterns of genetic variation both within and among forest tree populations and for estimating mating systems and gene flow. Allozymes are also very useful for genetic fingerprinting and paternity analysis.

Examples: Allozymes (different forms of an enzyme encoded by different alleles of the same gene), isoenzymes (various forms of an enzyme encoded by distinct genes), monoterpenes

Advantages: Direct insight into gene expression, although complex detection with frequent genotype noise (see features of the genetic code).

Disadvantages: Limited number of available biochemical markers and environmental influence.

Molecular Markers

Definition: Molecular markers are specific DNA sequences that can be analyzed to detect genetic differences among individuals, populations, or species. These markers are tools used in genetic studies for genome mapping, studying genetic diversity, identifying genes associated with certain traits, and selecting individuals in breeding programs.

Advantages: Molecular or DNA-based markers are the most recent to be developed and have many advantages over morphological and biochemical markers. The primary benefits are: (1) There is potentially an unlimited number of molecular markers available, and (2) DNA markers are generally not affected by developmental differences or environmental influences, leading to their high accuracy and specificity; ability to detect variations at the DNA level, allowing detailed genetic analyses. Lastly, they are also independent from phenotypic traits, thus better identifying environmental influences.

Disadvantages: Higher costs and technical complexity. Requirement for specific laboratory equipment and expertise.

1.16.2 TYPES OF MOLECULAR MARKERS

There are two general classes of molecular markers: those based on DNA-DNA hybridization and those based on the polymerase chain reaction (PCR).

Restriction fragment length polymorphism (RFLP) markers rely on DNA-DNA hybridization and have been used for organelle genetic analysis, genetic linkage mapping, and diversity analysis in forest trees. In principle, there are variations in the length of DNA fragments after enzymatic cleavage by specific restriction enzymes.

The three PCR-based molecular marker types used widely in forest trees are:

(1) Random amplified polymorphic DNA (RAPD)

(2) Amplified fragment length polymorphisms (AFLP)

(3) Simple sequence repeats (SSR)

All three marker types generally reveal polymorphism in non-coding regions of genomes. The RAPDs and AFLPs are diallelic, dominant markers, whereas SSRs are codominant and multiallelic.

Random Amplified Polymorphic DNA (RAPD)

Principle: Random amplification of genome fragments using short oligonucleotide primers.

Usage: Quick and inexpensive method for detecting genetic diversity.

Amplified Fragment Length Polymorphism (AFLP) combines PCR amplification and restriction enzymes to detect polymorphisms. The method has previously been used to analyze the genetic structure of populations and breeding programs.

Simple Sequence Repeats (SSR), also widely known as **Microsatellites**

Principle: Short tandem repeats of DNA sequences (1-6 bp) with a high degree of polymorphism.

Usage: Genetic diversity analysis, population genetics, pedigree reconstruction for breeding (see BWB in the chapter on tree improvement).

1.16.3 SSRS PROMINENT ROLE IN FOREST GENETICS OF PAST DECADES

SSR markers were first developed for human genetic mapping (Litt & Luty, 1989); short, tandemly repeated sequences of two, three, or four nucleotides are frequently found throughout the entire genome. For example, the dinucleotide repeat AC is commonly found in *Pinus*. Since the number of tandem repeats at a locus can vary, SSR markers tend to be highly polymorphic genetic marker types. For example, one allele might have 10 copies of the AC tandem repeat (AC)i0, whereas another allele would have 11 copies (AC)n, another 12 copies (AC)i2, and so forth. SSRs are widespread in the genome and show a high variation among individuals in a population, making them very useful for genetic studies.

Structure: Repeating DNA sequences, such as (CA)n and (AT)n, where "n" denotes the number of repeats.

Variability: Highly polymorphic due to different numbers of repeats in different individuals.

Locus-specific: Each SSR locus can be specific to a particular genome location.

Developing SSR markers requires significant investment because it involves several complex steps. First, genomic DNA libraries enriched with microsatellite sequences must be created and screened to identify clones containing SSR sequences (Ostrander et al., 1992). The DNA sequences of these clones must then be determined to identify the unique regions flanking the SSRs, which are essential for designing PCR primers to amplify these sequences from individual samples. Once suitable primers are designed, researchers must assess whether the SSRs are polymorphic and whether the resulting gel band patterns are genetically interpretable.

Some of the earliest SSR markers developed in trees were derived from the chloroplast genome already in the 1990s. This process was facilitated by the availability of the complete DNA sequence of the chloroplast genome in *Pinus thunbergii* (Wakasugi et al., 1994). Chloroplast SSRs are particularly useful due to their high polymorphism compared to other cpDNA markers, making them valuable for various studies. For example, in conifers, cpDNA is inherited paternally, making cpDNA markers ideal for determining the male parentage of offspring (paternity analysis) and tracking pollen dispersal in populations where the SSR genotypes of all male trees are known (Stoehr et al., 1998).

SSR markers based on nuclear DNA have also been developed for several forest tree species, including *Pinus* species, *Picea*, *Quercus*, and *Populus*. These studies describe isolating and cloning a limited number of SSRs, analyzing their inheritance patterns, and evaluating their utility across related species.

Detection and Analysis Procedure

DNA Isolation: Extraction of DNA from plant material using standard methods (column isolation method, CTAB method).

PCR Amplification: Use of specific primers for SSR locus amplification. Primers are designed to delimit repetitive sequences.

Amplified fragments are separated using **gel electrophoresis**, often on polyacrylamide or agarose gel. Fragment sizes are estimated by comparison with a DNA standard.

Detection is done by gel staining (e.g., ethidium bromide) or using fluorescently labeled primers for fragment detection. Length separation is often performed in a capillary sequencer. The software is used for further analysis for fragment size determination and subsequent polymorphism identification.

Advantages of SSR Markers:

High polymorphism rate: SSR markers are highly polymorphic, revealing many genetic variations among individuals.

Codominance: Allows detection of heterozygous states, providing more information than dominant markers.

Reproducibility (compared to, e.g., RFLP, AFLP, RAPD): High reproducibility of results due to specific primers and standardized protocols.

Wide distribution: Spread throughout the genome and found in both coding and non-coding regions (more common in the non-coding areas, considered selectively neutral markers).

Disadvantages of SSR Markers:

Technical complexity: Requires precise primer design and PCR condition optimization.

Primer sequence mutations: Variability in sequences surrounding SSR can affect amplification efficiency.

SSR markers are highly effective and useful tools in genetic studies, providing robust information about genetic diversity, enabling genetic differentiation, supporting breeding programs, and having broad applications in various fields of genetics and forestry. Their high polymorphism, codominance, and wide distribution in the genome make them ideal for multiple applications in forest genetics.

1.16.4 THE ADVENT OF SNP MARKERS IN FOREST TREES

A Single Nucleotide Polymorphism (SNP) is a genetic variation where a single nucleotide—A, T, C, or G—differs at a specific position in the genome among individuals. SNPs are essential for studying genetic diversity and adaptation in forest trees because they are abundant and widely distributed throughout the genome, especially in non-coding regions. As the most common type of gene variation, SNP markers are extensively used in genomic research and various applications, such as association genetics, which requires detecting polymorphisms over short physical distances along chromosomes. This high-resolution mapping demands a very dense marker coverage, which can be challenging for species with large genomes, such as those in the *Pinaceae* family.

Markers derived from SNPs are particularly valuable in association genetics because they can detect fine-scale genetic variation, which is crucial for identifying regions of the genome associated with specific traits. In forest trees like loblolly pine (*Pinus taeda*), SNPs are frequent, occurring approximately once every 60 base pairs (Brown et al., 2004), making them much more abundant than in humans, where SNPs occur about once every 1,000 base pairs. This abundance and straightforward discovery process make SNPs suitable for tree genetic studies. Typically, SNPs are bi-allelic, meaning they have two possible variants.

SNPs can be discovered by aligning multiple DNA sequences or by de novo sequencing of individuals from a population. In conifers, haplotypes, or unique combinations of SNPs on a single DNA fragment, can be easily identified using haploid megagametophyte tissue. Once SNPs are identified in a small sample, all individuals in an association population need to be genotyped to determine SNP frequencies. However, due to the high nucleotide diversity in forest trees, genotyping every discovered SNP is often not feasible due to cost.

Various methods exist for SNP genotyping, each with specific chemistries and technologies designed for high throughput, enabling the genotyping of thousands of trees across numerous SNPs. These methods rapidly evolve as commercial labs strive to reduce costs and enhance throughput. Genomic studies using SNPs also involve statistical tests to find associations between SNP genotypes and quantitative traits, like QTL detection, which helps identify genetic markers linked to specific phenotypic traits.

Characteristics of SNP markers:

Single Nucleotide Polymorphisms (SNPs) represent a change of one nucleotide in a DNA sequence, such as the sequence AAGCCTA having the SNP variant AAGCTTA. SNPs are binary, meaning only two alleles (variants) exist at a single locus, and each SNP locus is specific to a particular position in the genome, making them easily identifiable and analyzable. Their detection involves isolating DNA from plant material using methods like column isolation or the CTAB method, followed by amplifying specific genome regions containing SNPs through PCR with specific primers. DNA sequencing techniques, including Sanger sequencing for accurate detection in small fragments, next-generation sequencing (NGS), and SNP chip technology, are then employed to identify and analyze SNPs.

Next-Generation Sequencing (NGS): High-throughput method allowing large-scale genome sequencing and SNP identification.

DNA Chips: Technology enabling simultaneous analysis of thousands of specific SNPs on a small glass or silicon chip.

Bioinformatics Analysis: Analysis of sequence data using bioinformatics tools and software for SNP identification and interpretation. Includes mapping sequences to the reference genome and variant detection.

Advantages of SNPs:

High Density: SNPs are very common in the genome, allowing high-density markers for detailed genetic mapping.

Reproducibility: SNP analysis is highly reproducible due to precise detection methods and standardized procedures.

Codominance: SNP markers detect heterozygous states, providing more information than dominant markers.

Automation and High Throughput: Technologies like DNA chips and NGS enable rapid and automated analysis of many SNPs.

Disadvantages of SNPs:

Technical complexity: Requires specialized laboratory equipment and analysis and data interpretation expertise.

Costs: More for detection and analysis than other markers, especially when using NGS.

Applications of SNP Markers in Forest Genetics:

Single-nucleotide polymorphism (SNP) markers have become invaluable tools in forest genetics, offering diverse applications that enhance our understanding and management of forest resources. They play a critical role in analyzing genetic diversity and population structure, shedding light on evolutionary processes and migration patterns within and between species. SNP markers are instrumental in genome mapping and gene identification, facilitating the discovery of genes linked to key traits like disease resistance and growth characteristics. They underpin genome-wide association studies (GWAS), enabling the identification of genes associated with complex traits and adaptive responses. SNP markers also accelerate breeding programs through Marker-Assisted Selection (MAS) and genomic selection, streamlining the identification and propagation of desired genotypes. Beyond breeding, SNPs aid in species identification and tracking wood origin, proving vital in combating illegal logging and revising taxonomic classifications. Together, these applications underscore the transformative potential of SNP markers in advancing forest conservation and management strategies.

1.17 REFERENCES

Brown, G. R., Gill, G. P., Kuntz, R. J., Langley, C. H., & Neale, D. B. (2004). Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences*, *101*(42), 15255–15260. https://doi.org/10.1073/pnas.0404231101

Heer, K., Ullrich, K. K., Hiss, M., Liepelt, S., Schulze Brüning, R., Zhou, J., Opgenoorth, L., & Rensing, S. A. (2018). Detection of somatic epigenetic variation in Norway spruce via targeted bisulfite sequencing. *Ecology and Evolution*, *8*(19), 9672–9682. https://doi.org/10.1002/ece3.4374

Kamm, A., Doudrick, R. L., Heslop-Harrison, J. S., & Schmidt, T. (1996). The genomic and physical organization of Ty1-copia-like sequences as a component of large genomes in Pinus elliottii var. Elliottii and other gymnosperms. *Proceedings of the National Academy of Sciences*, *93*(7), 2708–2713. https://doi.org/10.1073/pnas.93.7.2708

Litt, M., & Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, *44*(3), 397–401.

Ostrander, E. A., Jong, P. M., Rine, J., & Duyk, G. (1992). Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences*, *89*(8), 3419–3423. https://doi.org/10.1073/pnas.89.8.3419

Ramakrishnan, M., Satish, L., Sharma, A., Kurungara Vinod, K., Emamverdian, A., Zhou, M., & Wei, Q. (2022). Transposable elements in plants: Recent advancements, tools and prospects. *Plant Molecular Biology Reporter*, *40*(4), 628–645. https://doi.org/10.1007/s11105-022-01342-w

Ritland, K. (2012). Genomics of a phylum distant from flowering plants: Conifers. *Tree Genetics & Genomes*, *8*(3), 573–582. https://doi.org/10.1007/s11295-012-0497-4

Sato, N. (2019). *Endosymbiotic Theories of Organelles Revisited: Retrospects and Prospects*. Springer Singapore. https://doi.org/10.1007/978-981-15-1161-5

Stoehr, M. U., Orvar, B. L., Vo, T. M., Gawley, J. R., Webber, J. E., & Newton, C. H. (1998). Application of a chloroplast DNA marker in seed orchard management evaluations of Douglas-fir. *Canadian Journal of Forest Research*, *28*(2), 187–195. https://doi.org/10.1139/x97-201

Voronova, A., Belevich, V., Jansons, A., & Rungis, D. (2014). Stress-induced transcriptional activation of retrotransposon-like sequences in the Scots pine (Pinus sylvestris L.) genome. *Tree Genetics & Genomes*, *10*(4), 937–951. https://doi.org/10.1007/s11295-014-0733-1

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., & Sugiura, M. (1994). Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine Pinus thunbergii. *Proceedings of the National Academy of Sciences*, *91*(21), 9794–9798. https://doi.org/10.1073/pnas.91.21.9794

2. MENDELIAN GENETICS

Genetics is grounded in several foundational principles. Genes are considered the individual constant units of heredity, transferred via sex cells. Each parent contributes half of their genetic information (a subset of chromosomes) to their offspring. Functional forms of genes, known as alleles, can be either dominant or recessive. In progeny, various phenotypic traits exhibit specific numeric segregation patterns. This seemingly simple yet profound logic dates back to Johan Gregor Mendel, who first observed and summarized these phenomena. Mendelian inheritance is the monogenic inheritance of qualitative traits, named in his honor.

2.1 MENDELIAN TRAITS IN FOREST TREES

This book primarily discusses genetics in forest trees, where Mendelian inheritance is relatively rare. Most traits related to growth and production, which are important for tree improvement, exhibit quantitative inheritance, which will be further explored in the next chapters. Exceptions are summarized in White et al. (2007), highlighting traits such as:

- Cone color in *Pinus monticola* (Steinhoff, 1974)
- Seedling foliage color in *Picea abies* (Langner, 1953)
- Chlorophyll deficiencies and other morphological variants in *Pinus taeda* (Franklin, 1969)
- Diameter growth in *Pinus patula* (Barnes et al., 1987)
- Blister rust resistance in *Pinus lambertiana* (Kinloch et al., 1970)

Many other morphological traits under simple genetic control could likely be identified if geneticists observed such traits in segregating populations. Moreover, the above-mentioned studies have been recently challenged by modern genomic methods (**see Chapter 7 on genomics of forest trees**).

2.2 MENDEL IN THE HISTORICAL CONTEXT

Irrespective of the genetic architecture of any trait, it is subjected to natural selection. The concept of natural selection was first described by another important figure of biology and related sciences, Sir Charles Darwin. In the next section, we put the contributions of Mendel and Darwin into a historical context, summarizing their views on the mechanism of inheritance and their impact. Both scientists' lives, milestones, and achievements are outlined in Table 2.1.

Table 2.1 Comparative Timeline: Mendel and Darwin

Year	Mendel	Darwin
1809	—	Born in Shrewsbury, England.
1822	Born in Heinzendorf, Moravia.	—
1831-1836	—	Voyage on HMS Beagle, leading to his observations on natural selection.
1843	Joined the Augustinian monastery in Brno.	—
1856-1863	Conducted experiments on pea plants.	—
1859	—	Published *On the Origin of Species.*
1865	Presented his findings on inheritance.	—
1866	Published *Experiments on Plant Hybridization.*	—
1871	—	Published *The Descent of Man.*
1882	—	Died as one of the most prominent scientists of his time.
1884	Died without significant recognition for his work.	—

Regarding the inheritance mechanism, Darwin proposed the theory of natural selection, explaining how species evolve based on differential survival and reproduction. However, Darwin lacked a clear mechanism for how traits were inherited.

On the other hand, Mandel provided the inheritance mechanism through his segregation and independent assortment laws. His experiments demonstrated how traits are passed from parents to offspring predictably.

Their outlook on natural variation differed as Darwin recognized that variation within species was crucial for natural selection. He observed that individuals with advantageous traits were more likely to survive and reproduce. In contrast, Mendel explained that the source of this variation is genetic recombination and the independent assortment of alleles during reproduction.

Although Mendel's work was not known to Darwin, Mendel's principles of heredity provided the genetic foundation that Darwin's theory of natural selection needed. When Mendel's work was rediscovered in 1900, it filled in the gaps in Darwin's theory, showing how traits are inherited and how variation arises within populations. The eventual synthesis of Mendelian genetics with Darwinian evolution led to the modern evolutionary synthesis in the early 20th century. This integrated Darwin's theory of evolution by natural selection with Mendel's principles of genetic inheritance, providing a comprehensive framework for understanding how evolutionary processes operate at both the genetic and population levels.

What sets both forefathers apart is the historical impact at the time of their findings. Mendel´s work initially went unrecognized, only gaining attention decades later when it was rediscovered. Mendel's work initially garnered only mild interest among his peers in Brno and Vienna. It took almost 50 years for his findings to gain worldwide recognition. Once recognized, his principles became the cornerstone of genetics.

In contrast, Darwin´s work was immediately impactful and controversial, significantly influencing the scientific community and public thought. Darwin's ideas spurred further research into the mechanisms of evolution, eventually leading to the incorporation of Mendelian genetics.

Mendel's work was rediscovered independently by three scientists around the turn of the 20th century:

Hugo de Vries: A Dutch botanist, de Vries conducted his experiments on plant hybrids and came to conclusions similar to Mendel's. In 1900, while searching the literature, he came across Mendel's 1866 paper and recognized the significance of Mendel's findings to his work.

Carl Correns: A German botanist, Correns also conducted hybridization experiments and developed principles like Mendel's. In 1900, Correns discovered Mendel's work and realized that Mendel had preceded him in formulating inheritance laws. Correns was instrumental in ensuring Mendel received proper recognition.

Erich von Tschermak: An Austrian agronomist and botanist, von Tschermak conducted breeding experiments and derived results aligned with Mendel's laws. In 1900, he found Mendel's paper and acknowledged the priority of Mendel's discoveries.

These three scientists not only rediscovered Mendel's laws but also helped to validate and popularize them, leading to widespread acceptance and recognition of Mendel's foundational contributions to the field of genetics. Their independent confirmation of Mendel's findings marked the beginning of modern genetics.

Mendel's discoveries on genetic inheritance provided the missing piece to Darwin's theory of natural selection. While Darwin explained how evolution occurs, Mendel explained how traits are inherited. Together, their findings form the basis of modern evolutionary biology, demonstrating how genetic variation leads to evolutionary change. The rediscovery of Mendel's work at the turn of the 20th century allowed scientists to merge these ideas, leading to a deeper understanding of the mechanisms driving evolution and heredity.

2.3 MENDEL´S LAWS IN THE LIGHT OF HIS HYBRIDIZATION EXPERIMENTS

One effective way to introduce Mendelian genetics is by recapitulating Mendel's experiments. In his most famous experiment, a homozygote for the spherical seed allele was crossed with a homozygote for the wrinkled seed allele. This sets up the experimental context where Mendel's **Law of Segregation** can be later observed. These parents, known as founders or true inbred lines, were established through test crosses ensuring dominant homozygosity. Each parent contributed gametes of only one kind, S or s, resulting in offspring with the genotype Ss and the spherical seed phenotype. This describes the creation of the F_1 generation and demonstrates the concept of dominance (**Law of Dominance**).

When the F_1 plants (the first filial generation) self-pollinate, they produce two gametes, S and s. Mendel deliberately self-pollinated some F_1 offspring to create an F_2 generation, whose plants segregated for both traits found in the inbred parents. This illustrates Mendel's **Law of Segregation** during the formation of gametes in the F_1 generation. The S and s gametes combine randomly in four different ways to form F_2 plants, resulting in three combinations producing the "spherical seed" phenotype and one combination producing the "wrinkled seed" phenotype. This results in the observed 3:1 ratio. The wrinkled seed phenotype corresponds to the "ss" genotype.

In contrast, the spherical seed phenotype is observed with the genotypes "SS" or "Ss," indicating that the spherical seed trait is dominant, and the wrinkled seed trait is recessive. This reinforces the **Law of Dominance**, where dominant traits mask recessive ones in heterozygous individuals. However, the **complete dominance** scenario we just introduced is just one of the options, and the concept of dominance as the basic interaction between alleles within a single locus will be further elaborated.

These experiments are known as **monohybrid** crosses because the peas were inbred lines that differed in a **single trait** (Fig. 1.1). Neither mixed types nor intermediate offspring were observed. This highlights Mendel's use of monohybrid crosses to simplify observing inheritance patterns for a single trait. Mendel's significant contribution to science was observing the trait types in the offspring, counting them, and calculating the ratios of each type. In the F_2 generation, he observed approximately three times as many plants of one trait type as the other for all seven traits he studied. Additionally, the trait observed in the F_1 generation was more prevalent in the F_2 generation.

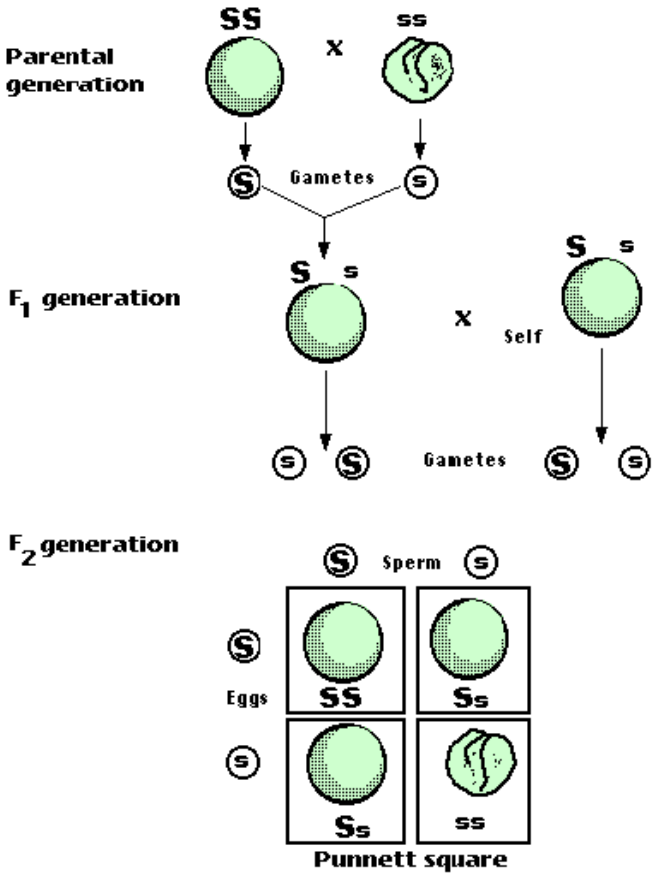


Fig. 1.1. Example of one of Mendel's monohybrid crosses with peas. The experiments involved three generations of plants. First, the true-breeding round and wrinkled-seeded parental plants (Pj) were crossed to form the F_1 generation, which comprised uniformly round-seeded plants. A single F_1 generation plant was then self-pollinated to create the F_2 generation. The Punnett square shows the F_2 offspring resulting from selfing the F_1. There are three possible genotypes (SS, Ss, and ss), but only two phenotypes (round and wrinkled) segregate in a 3:1 ratio.

The **Law of Independent Assortment** states that the alleles of two (or more) different genes get sorted into gametes independently. In other words, the allele a gamete receives for one gene does not influence the allele received for another gene. This law is a key principle of Mendelian genetics and can be illustrated using a dihybrid cross, which examines the inheritance of two different traits.

Example of the Law of Independent Assortment with Two Traits:

Let's consider Mendel's classic experiment with pea plants, focusing on two traits: (1) Seed shape: round (R) is dominant to wrinkled (r). (2) Seed color: yellow (Y) is dominant to green (y).

Parental Generation (P):

- One parent is homozygous dominant for both traits: RRYY (round, yellow seeds).
- The other parent is homozygous recessive for both traits: rryy (wrinkled, green seeds).

F_1 Generation:

- All offspring (F_1) are heterozygous for both traits: RrYy (round, yellow seeds).

F_2 Generation:

To see the Law of Independent Assortment in action, we need to examine the F_2 generation produced by self--pollinating the F_1 hybrids (RrYy x RrYy).

Formation of Gametes:

According to the Law of Independent Assortment, each allele pair segregates independently. Therefore, the F_1 plant can produce four types of gametes with equal probability:

RY, Ry, rY, ry

Punnett Square:

A 4x4 Punnett square can be used to show the combinations of these gametes:

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

From the Punnett square, we can see the following phenotypic combinations and their frequencies:

1. **Round, Yellow (R_Y_)**: RRYY, RRYy, RrYY, RrYy (9 combinations)
2. **Round, Green (R_yy)**: RRyy, Rryy (3 combinations)
3. **Wrinkled, Yellow (rrY_)**: rrYY, rrYy (3 combinations)
4. **Wrinkled, Green (rryy)**: rryy (1 combination)

This results in a phenotypic ratio of **9:3:3:1** with the following phenotypic ratios:

- 9/16 Round, Yellow
- 3/16 Round, Green
- 3/16 Wrinkled, Yellow
- 1/16 Wrinkled, Green

The independent assortment of alleles results in this 9:3:3:1 ratio. Each trait is inherited independently, meaning the allele for seed shape (R or r) is sorted into gametes independently of the allele for seed color (Y or y). This independence creates the diversity of phenotypic combinations seen in the F_2 generation.

We can observe the combinability of any number of traits, provided each trait is determined by a single gene on different chromosomes. Mendel's laws will still apply in these cases. As possible combinations increase, the Punnett square will also grow larger. To determine the number of genotypic and phenotypic categories, we can use the following general formulas, where n is the number of traits being observed:

3^n: number of genotypic classes

2^n: number of phenotypic classes

Mendel's Law of Independent Assortment explains how different traits are passed from parents to offspring independently. This principle is fundamental to understanding the genetic variation observed in sexually reproducing organisms. Gregor Mendel was fortunate in his experiments that all the traits he observed were located on different chromosomes. This allowed him to observe the independent assortment of traits. However, genes are located next to each other on chromosomes, forming what is known as genetic linkage. This linkage is not entirely fixed, as during **meiosis**, parts of chromosomes can be exchanged, leading to the separation of linked genes through a process called **crossing-over**.

Crossing over during meiosis is a crucial process that impacts genetic structure by breaking genetic linkage, the tendency of genes close together on the same chromosome to be inherited. This exchange of chromosome segments between homologous chromosomes occurs during prophase I at chiasmata and creates new combinations of alleles, increasing genetic diversity. By separating linked genes, crossing over allows for the independent assortment of traits, producing recombinant phenotypes and altering inheritance patterns. This genetic recombination enhances evolutionary adaptability by providing a greater range of traits for populations to respond to environmental changes, contributing to biodiversity and the success of species over time.

Mendel's meticulous experiments and the clear patterns he observed in pea plants provided a framework still used today to study and manipulate genetics. His work bridges classical genetics with modern molecular biology, demonstrating the lasting impact of his discoveries on genetics.

2.4 BEYOND THE LAWS OF MENDEL

Exceptions to complete dominance include several patterns of inheritance where the heterozygous phenotype differs from both homozygous phenotypes. These exceptions include incomplete dominance, codominance, multiple alleles, pleiotropy, epistasis, and polygenic inheritance.

Incomplete Dominance: In incomplete dominance, the heterozygous phenotype is an intermediate blend between the two homozygous phenotypes. For Example, in snapdragons, a cross between red (RR) and white (rr) flowers results in offspring with pink (Rr) flowers.

Codominance: In codominance, both alleles in the heterozygote are fully expressed, resulting in a phenotype that displays both traits simultaneously. An example is the ABO blood group system in humans, where individuals with genotype IAIB exhibit both A and B antigens on their red blood cells.

Multiple Alleles: Some genes have more than two alleles within a population. Although an individual can only possess two alleles, the multiple allele system allows for a greater variety of genetic combinations and phenotypes.

The ABO blood group system is an example of multiple alleles involving three alleles: IA, IB, and i. In a population, there can be multiple alleles at a single gene locus, though each individual still only possesses two alleles (one from each parent). This results in various genotypes across the population, such as A_1 A_1, A_1 A_3, A_2 A_n, and so on. This genetic diversity is crucial for the adaptability and evolution of species.

Pleiotropy: Pleiotropy occurs when a single gene influences multiple seemingly unrelated phenotypic traits. An example is Marfan syndrome, a genetic disorder caused by mutations in the FBN1 gene, which affects connective tissue and leads to a wide range of symptoms, including tall stature, long limbs, and cardiovascular issues.

Epistasis: Epistasis happens when one gene affects the expression of another gene. For Example, in Labrador retrievers, one gene determines coat color (B for black, b for brown), but another gene (E) determines whether pigment will be deposited in the fur. A dog with two recessive alleles (ee) will be yellow regardless of the B/b genotype.

Polygenic Inheritance: Polygenic inheritance involves multiple genes contributing to a single trait, resulting in continuous variation. Human height, skin color, and eye color are polygenic traits, where several genes contribute to the overall phenotype.

These exceptions to complete dominance illustrate the **complexity of genetic inheritance** and how different mechanisms can influence the expression of traits.

In **forest genetics**, exceptions to complete dominance can be observed through several examples:

Incomplete Dominance: An example of incomplete dominance can be found in the coloration of needles in certain pine species. For instance, in some pine hybrids, the needle color of the heterozygous individuals (Rr) may appear as an intermediate shade between the two homozygous parent colors (RR and rr).

Codominance: Codominance can be seen in the expression of disease resistance in some tree species. For Example, in American chestnut (*Castanea dentata*), codominance can occur with resistance genes to chestnut blight, where both alleles are expressed, leading to a tree that shows partial resistance characteristics from both alleles.

Multiple Alleles: In the case of *Pinus sylvestris* (Scots pine), multiple alleles may determine the shape and size of cones. Different alleles contribute to variations in cone morphology, which can affect seed dispersal mechanisms and the overall fitness of the trees in various environments.

Pleiotropy: An example of pleiotropy in forest genetics is seen in the *Quercus robur* (oak) tree, where a single gene can affect leaf morphology, wood density, and disease resistance. Mutations in such genes can lead to a broad range of phenotypic changes impacting the tree's overall growth and survival.

Epistasis: Epistasis can be observed in the height growth of hybrid poplars (*Populus* spp.), where one gene controls the growth rate. In contrast, another gene can modify the effect, leading to variations in overall tree height. The interaction between these genes determines the final phenotype.

Polygenic Inheritance: Polygenic inheritance is evident in the growth traits of Douglas-fir (*Pseudotsuga menziesii*), where multiple genes contribute to traits like height, diameter, and wood density. The combined effect of these genes results in continuous variation in growth performance among individual trees.

These examples illustrate how complex genetic mechanisms influence various traits in forest tree species, reflecting forest ecosystems' rich diversity and adaptability.

2.5 CHI-SQUARE TEST IN MENDELIAN GENETICS

In both real and model cases, we verify the validity of Mendel's third law of independent assortment—whether genes are located on the same chromosome—using the chi-square (χ^2) test, also known as the goodness-of-fit test. Based on Mendelian inheritance, this statistical test compares the expected phenotypic ratios with the observed phenotypic ratios. We can determine whether the genes assort independently or exhibit genetic linkage by analyzing the discrepancies between the expected and observed values. To calculate test statistics that can be compared with statistical tables, we use the following formula:

$$\chi^2 = \sum \frac{(o - e)^2}{e} \quad (2.1)$$

Where *o* represents the observed frequencies and *e* represents the expected frequencies. This formula allows us to determine if there is a significant difference between the observed and expected values, helping us to assess the validity of Mendel's third law in our experiments.

The calculated value is compared with the tabulated value for N degrees of freedom at a significance level of α=0.05. If the calculated χ^2 value is less than the tabulated value, the difference between the observed and expected data is not statistically significant. This indicates that the observed trait is segregated in the expected ratio, confirming that Mendel's third law applies to the studied traits. If Mendel's law is not confirmed, it can be assumed that the genes of interest do not assort independently because they are located on the same chromosome. Such genes are said to be in genetic linkage. The closer these genes are to each other on the chromosome, the stronger their genetic linkage.

A similar application of this statistical test is documented in the study by Steinhoff, (1974) and further detailed in White et al. (2007). Steinhoff's study investigated the strobili color in *Pinus monticola* to determine its compliance with Mendel's law of segregation. The research specifically examined whether the inheritance of strobili color followed the expected Mendelian ratios. Steinhoff found that the observed phenotypic ratios closely matched the expected ratios, indicating that the strobili color trait in *Pinus monticola* adheres to Mendelian inheritance patterns. This conclusion was supported by a chi-square (χ^2) test, which showed no statistically significant difference between the observed and expected values. These findings reinforce the applicability of Mendel's laws to certain traits in coniferous species, contributing valuable insights to the field of genetic inheritance in plants.

2.6 SUMMARY OF MENDELIAN GENETICS

Gregor Mendel's pioneering work laid the foundational principles of modern genetics, profoundly shaping the field and influencing scientific thought in several critical ways. His experiments with pea plants revealed the mechanisms by which traits are inherited, establishing key concepts that continue to underpin genetic research and applications today.

Mendel formulated the fundamental laws of inheritance: the Law of Segregation and the Law of Independent Assortment. The Law of Segregation states that each individual carries two alleles for each trait, which separate during gamete formation. This ensures that offspring inherit one allele from each parent. This principle is deeply intertwined with the process of meiosis, discussed in detail in earlier studies. The Law of Independent Assortment describes how alleles of different genes are inherited independently, promoting genetic diversity. Together, these principles provided a framework for understanding the transmission of genetic traits.

Mendel introduced the concept of dominant and recessive alleles, explaining why some traits appear in offspring while others remain hidden despite the presence of both alleles. This insight was pivotal in deciphering patterns of inheritance and is still central to understanding how genetic traits are expressed.

Another critical contribution was Mendel's distinction between an organism's genotype (genetic makeup) and phenotype (observable traits). This differentiation allows for a clearer understanding of how genetic information translates into physical characteristics, a concept that remains fundamental to genetics and biology. Mendel's use of mathematics in studying inheritance added a level of precision to genetics that was unprecedented at the time. By applying probability to segregation ratios, Mendel could predict the distribution of traits in offspring. This predictive approach became a cornerstone of genetic studies and has broad applications in research, breeding, and medicine.

Mendel's discoveries provided the groundwork for understanding genes as units of heredity, paving the way for later breakthroughs such as identifying DNA as the hereditary material. The structure of DNA, uncovered by Watson and Crick, built on Mendelian principles and gave rise to the field of molecular genetics, which explores the molecular mechanisms underlying inheritance.

Mendel's principles have had a transformative impact on agriculture and animal breeding. By applying his laws, breeders can select desirable traits to improve crop yields, enhance livestock quality, and develop resilient plant varieties. His work remains central to modern breeding programs, including tree improvement and other agricultural advancements. Mendelian genetics has also been instrumental in understanding human genetic disorders. Patterns of inheritance, as described by Mendel, apply to many diseases, enabling genetic counseling and testing. This has profound implications for diagnosing, managing, and predicting genetic conditions.

Finally, Mendel's principles are the foundation for modern genetic research and biotechnology. Techniques such as gene mapping, genome sequencing, and genetic modification rely on understanding the inheritance patterns he discovered. These methods allow scientists to manipulate genetic material for research purposes and develop therapeutic interventions for various genetic conditions.

In summary, Mendel's contributions to genetics have had far-reaching effects, from elucidating the basic principles of heredity to enabling cutting-edge advancements in biotechnology, medicine, and agriculture. His work continues to inspire and guide genetic research, making him a foundational figure in the field.

2.7 REFERENCES:

Barnes, R. D., Matheson, A. C., Mullin, L. J., & Birks, J. (1987). Dominance in a metric trait of *Pinus patula* Schiede and Deppe. *Forest science*, *33*(3), 809–815.

Franklin, E. C. (1969). *Mutant forms found by self-pollination in loblolly pine.* https://www.cabidigitallibrary.org/doi/full/10.5555/19701608609

Kinloch, B. B., Parks, G. K., & Fowler, C. W. (1970). White Pine Blister Rust: Simply Inherited Resistance in Sugar Pine. *Science*, *167*(3915), 193–195. https://doi.org/10.1126/science.167.3915.193

Langner, W. (1953). Eine Mendelspaltung bei Aurea Formen von Picea abies (L.) Karst. Als Mittel zur Klärung der Befruchtungsverhältnisse im Walde. *Z Forstgenet Forstpflanzenzücht*, *2*, 49–51.

Steinhoff, R. J. (1974). *Inheritance of cone color in Pinus monticola.* https://www.cabidigitallibrary.org/doi/full/10.5555/19741619745

White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI.

3. BASICS OF POPULATION GENETICS

Population genetics extends our understanding of heredity and variability from individual organisms to the population level. While we will broaden our knowledge of inheritance principles to encompass entire populations, we will often use simple single-locus models as an introductory framework to grasp the fundamental concepts.

We shall first define "A population as a group of organisms of the same species that live in a specific geographic area and have the potential to interbreed. The population members share a common gene pool, and genetic exchange occurs freely among them." according to (Hartl & Clark, 2006).

In population genetics, we assess the impact of various factors that influence changes in the genetic structure of a population over time. These changes can be simplified as **evolution**. Beyond these evolutionary factors' direct (additive) effects, their more complex interactions (multiplicative effects) also become a subject of scientific interest. Population genetics builds on the basic principles of heredity and variability (Mendelian laws) and develops them further in the broader context of species evolution. Differences in the phenotypic expression of specific traits (e.g., the stem straightness of the Scots pine) allow for a deeper study of heredity and genetic variability.

Understanding a population's genetic structure and evolutionary dynamics requires analyzing how evolutionary factors interact. The additive effects of these factors can be straightforward, but their multiplicative interactions can lead to complex patterns of genetic change. For example, the interaction between natural selection and genetic drift can significantly impact allele frequencies, especially in small populations.

Population genetics has practical applications in conservation biology, agriculture, and medicine. For instance, it helps manage genetic diversity in endangered species, breed crops with desirable traits, and understand the genetic basis of diseases.

By studying populations' genetic structure and variability, researchers can gain insights into the evolutionary processes shaping biodiversity and develop strategies for effectively preserving and utilizing genetic resources. We begin by defining a simple, purely theoretical model of a population. The purpose is to establish a reference for quantifying the influence of evolutionary factors on the genetic structure of the population, which will be addressed subsequently.

3.1 GENETIC STRUCTURE OF A POPULATION

For simplicity, let us consider any species with a diploid set of chromosomes. Assume a larger number of individuals in the population, in order of thousands, to ensure statistical representativeness of the allele sample. These individuals segregate freely according to Mendel's laws. Generally, for any autosomal locus, an allele represents a specific haplotype (the functional form of a gene in haploid cells—gametes), and a pair of alleles represents a genotype (the functional forms of a gene in diploid cells—zygotes). The description of the genetic structure of a population is based on quantifying **allele frequencies** within the population and their distribution among individual organisms, i.e., **genotype frequencies**.

Allele Frequency: The proportion of a particular allele among all allele copies in the population.

Genotype Frequency: The proportion of a specific genotype among all individuals in the population.

3.2 HARDY-WEINBERG EQUILIBRIUM

After the initial rediscovery of Gregor Mendel's work on inheritance, there was a prevalent misunderstanding among scientists regarding recessive alleles. Many believed these alleles were inferior and would diminish in frequency over generations. This misconception was rooted in the observation that while recessive alleles persisted in populations, their apparent invisibility in phenotypes led researchers to think they were gradually being eliminated.

This incorrect assumption was challenged in 1907 by G. H. Hardy, a British mathematician, and Wilhelm Weinberg, a German physiologist. Hardy and Weinberg independently published works that refuted this idea, demonstrating through their now-famous Hardy-Weinberg principle that allele frequencies in a population would remain constant from generation to generation without evolutionary influences. This principle formed the foundation of population genetics and provided a mathematical baseline for studying genetic variation.

Hardy's publication specifically addressed a debate with British geneticist Undy Yule, who argued that the frequencies of alleles would naturally converge to the Mendelian 3:1 ratio over time (BOX 3.1).

BOX 3.1

"To the Editor of Science: I am reluctant to intrude in a discussion concerning matters of which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists. However, some remarks of Mr. Udny Yule, to which Mr. R. C. Punnett has called my attention, suggest that it may still be worth making…

Suppose that Aa is a pair of Mendelian characters, A being dominant, and that in any given generation the number of pure dominants (AA), heterozygotes (Aa), and pure recessives (aa) are as $p{:}2q{:}r$. Finally, suppose that the numbers are fairly large, so that mating may be regarded as random, that the sexes are evenly distributed among the three varieties, and that all are equally fertile. A little mathematics of the multiplication-table type is enough to show that in the next generation the numbers will be as $(p+q)^2 : 2(p+q)(q+r) : (q+r)^2$, or as $p_1 : 2q_1 : r_1$, say. The interesting question is: in what circumstances will this distribution be the same as that in the generation before? It is easy to see that the condition for this is $q^2 = pr$.

And since $q_1^2 = p_1 r_1$, whatever the values of p, q, and r may be, the distribution will in any case continue unchanged after the second generation"

Hardy's concise argument mathematically proved that allele frequencies remain stable under certain ideal conditions (no mutation, migration, selection, or genetic drift). This insight helped clarify the true nature of inheritance and the persistence of genetic variation in populations.

A population can be considered ideal if all the following assumptions are met:

- **Diploid Set of Chromosomes**: The species in question has a diploid set of chromosomes.
- **Sexual Reproduction**: The species reproduces sexually.
- **Non-overlapping Generations**: Each generation is distinct and does not overlap with others.
- **Equal Allele Frequencies**: Male and female individuals have the same allele frequencies.
- **Random Mating**: Mating occurs randomly within the population.
- **Infinite Population Size**: The population size is infinitely large.
- **Negligible Migration, Mutation, and Selection**: The effects of migration, mutation, and natural selection are negligible.

The genetic structure of real populations differs from that of an ideal population due to violating some (or even all) of the assumptions mentioned above. It is important to remember that the parameters of an ideal population are specific to the level of individual genes. While one locus may exhibit Hardy-Weinberg equilibrium, this may not be true for other loci.

We can use the chi-square test to verify whether a population is in Hardy-Weinberg equilibrium. Note the similarity to verifying the 3:1 segregation ratio demonstrated in the chapter on Mendelian inheritance. Since we are only sampling from the population, the proportions of different genotypes will be distorted by random sampling. The p-value of the chi-square test tells us the probability that the observed deviation from Hardy-Weinberg equilibrium is due to chance alone. If the probability of such a deviation occurring by chance is less than 5% (p = 0.05, meaning the deviation is pronounced), we consider it significant and can investigate its causes.

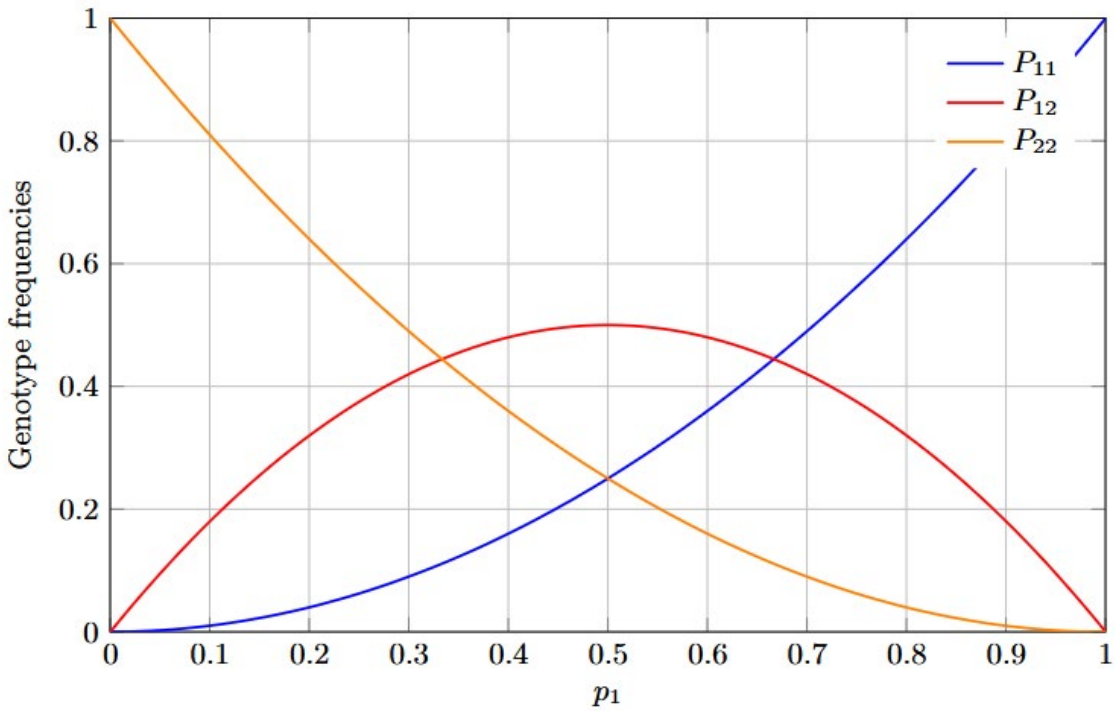


Fig. 3.1 The representation of heterozygotes (genotypes plotted on the Y axis) is maximal when the frequencies of alleles p_1 and p_2 are equal to 0.5.

3.2.1 APPLICATIONS AND SIGNIFICANCE

The Hardy-Weinberg model is crucial for several reasons: (1) It forms a baseline for evolutionary studies to detect when and how populations deviate from equilibrium due to evolutionary forces like selection, genetic drift, mutation, and migration; (1) It allows for predictions of genotype frequencies from known allele frequencies, aiding in understanding the genetic structure of populations; (3) Conservation Biology: It helps manage genetic diversity in conservation programs by identifying deviations from expected genetic distributions; (4) In medical genetics, the model can predict the distribution of genetic disorders within a population, if the population meets the ideal conditions.

The mentioned dependencies are utilized in the study of traits with simple inheritance. An example of such a trait is the autosomal recessive disorder phenylketonuria (a hereditary metabolic disorder) encoded by a single locus. Affected individuals are homozygous. Their frequency in the population is approximately 1 in 9,000 (in the Czech Republic). According to the Hardy-Weinberg model, the expected frequency of recessive homozygotes (symbolically denoted as A_2A_2) should match the observed frequency of the disorder in the population, i.e.,

$$p_2^2 = \frac{1}{9000} \tag{3.1}$$

$$p_2 = \sqrt{\frac{1}{9000}} \approx 0.01 \tag{3.2}$$

The frequency of the A_2 allele in the population is therefore 1.1%. The frequency of heterozygotes, i.e., the proportion of carriers in the general population, is approximately 2.2% (if recessive homozygotes do not participate in reproduction).

Also relevant is that heterozygotes are relatively common even when the A_2 allele is rare. A real example in human populations (Hartl, 1999) is cystic fibrosis, a disease expressed only by homozygous recessive individuals. The disease affects 1 in 2500 newborn Caucasians, and assuming Hardy-Weinberg equilibrium, we can calculate $q^2 = 1/2500$, $q = 0.02$, $p = 0.98$ and $2pq = 0.0392$. The latter value means that 1 in 25 Caucasians (1/25 = 0.04) are heterozygote carriers of the disease, but only 1 in 2500 are born with the disease.

These relationships can be adjusted for specific loci located on sex chromosomes or for a larger number of loci (presence of gene linkage on chromosomes).

Albino seedlings are commonly found in forest tree nurseries, such as this one of *Pinus ponderosa*. The lack of chlorophyll is presumably due to a mutation in a gene determinant for chlorophyll biosynthesis. The mutation is lethal if the seedlings do not photosynthesize and will not survive. Mutant phenotypes (e.g., albino needles, dwarfs, fused cotyledons) are often observed among the thousands of germinants in commercial nursery beds. Assume that in a *P. ponderosa* nursery, 20 albino mutants occur in a seedbed with 120,000 germinants and that this mutant is caused by a recessive allele at a single locus. What is the mutant gene's allele frequency (q) in this seedbed? The frequency of albino mutants is $X_{22} = q^2 = 20/120{,}000 = 0.0001667$, and $q = (0.0001667)172 = 0.013$. All genotype frequencies can be estimated from equation (BOX 3.1). Still, the accuracy of these estimates rests on the prior assumption that the population is in Hardy-Weinberg equilibrium.

3.3 EVOLUTIONARY FORCES

In an ideal population, the frequency of alleles remains constant across generations, a concept known as Hardy-Weinberg equilibrium. However, natural populations rarely conform to this ideal due to the influence of various evolutionary forces. These forces—genetic drift, mutation, migration, and selection—play crucial roles in shaping the genetic structure of populations by altering allele frequencies over time.

Genetic drift refers to random fluctuations in allele frequencies, which can have a significant impact, especially in small populations. Mutations introduce genetic variations by altering DNA sequences, providing raw material for

evolution. Migration, or gene flow, occurs when individuals move between populations, introducing new alleles and increasing genetic diversity. Natural and artificial selection acts on phenotypic variations, favoring some alleles over others based on their impact on survival and reproduction.

These evolutionary forces drive the dynamic nature of genetic structures in real populations, leading to the diversity of life we observe today. Understanding these mechanisms is fundamental to studying evolution, as they explain how populations adapt, evolve, and sometimes even speciate. In this chapter, we will explore these forces in detail, examining how they interact and contribute to the continuous process of evolution.

3.3.1 GENETIC DRIFT

The first evolutionary force we will examine is genetic drift, which refers to random fluctuations in allele frequencies. It was previously assumed that the population size was infinite, and the allele frequencies in gamete haplotypes corresponded to their frequencies in the genotypes of the parental population. However, when the population size is reduced, the selection of alleles will not be representative of the genetic structure of the parental population, leading to random intergenerational changes in allele frequencies. This phenomenon is known as genetic drift and arises because natural populations consist of a finite number of individuals. Thus, the segregation of alleles into gametes occurs randomly, much like tossing a coin. While numerous coin tosses approach a theoretical 50:50 ratio, a small number of tosses often yield significantly skewed results.

Similarly, in large populations, allele frequencies tend to shift only slightly. In contrast, in small populations, these frequencies can change dramatically from generation to generation, potentially leading to the loss of one allele and the fixation of another. Consequently, genetic drift acts towards reducing genetic diversity over time.

The main parameters affecting the intensity of genetic drift are:

Population Size: The most significant factor influencing genetic drift is population size. Genetic drift has a more pronounced effect in smaller populations because random fluctuations in allele frequencies are more important when fewer individuals are present. In larger populations, these random changes are diluted across many individuals, reducing the impact of drift.

Allele Frequency: The initial frequency of an allele in the population can affect how genetic drift influences its future frequency. Rare alleles are more likely to be lost purely by chance in small populations, whereas alleles with intermediate frequencies have a different dynamic under drift.

Bottlenecks and **Founder Effects**: Events drastically reducing the population size, such as natural disasters or other catastrophic events (bottlenecks), can intensify genetic drift. Similarly, when a new population is founded by a small number of individuals (founder effect), the initial allele frequencies can differ significantly from those in the source population, leading to strong drift effects.

Reproductive Strategy: How a population reproduces can influence the intensity of genetic drift. For example, populations that reproduce sexually with random mating will experience different drift dynamics compared to those with asexual reproduction or non-random mating patterns.

Generation Time: The number of generations over which genetic drift can act also plays a role. Populations with short generation times will experience genetic drift effects more rapidly than those with longer generations, as more opportunities for allele frequency changes arise in each period.

Understanding these parameters helps predict how genetic drift will influence a population's genetic structure and contributes to broader insights into evolutionary processes.

The random selection of gametes can be expressed statistically and easily simulated on a computer. We start the simulation by defining an initial population that follows the Hardy-Weinberg equilibrium, with an initial allele frequency of *p1*. We will select only ten individuals from the previous generation in each subsequent generation. This process results in the allele selection in the gamete haplotypes not representing the allele frequencies of the last generation. Consequently, this leads to random intergenerational changes. It is not difficult to demonstrate that the extent of this randomness (the degree of differentiation in allele frequencies) corresponds to:

$$\sigma^2_{\Delta p_1} = \frac{p_1 p_2}{2N} \tag{3.3}$$

Genetic drift thus depends on the initial allele frequencies and the population size.

To better understand genetic drift, we can observe its effects over several generations in an observed population. By tracking changes in allele frequencies, we can see how genetic drift can lead to significant shifts and even the eventual disappearance of certain alleles, illustrating its impact on the genetic makeup of populations. This examination will clarify how random processes contribute to the evolutionary landscape and influence the genetic diversity within natural populations. Currently, there are many available simulation platforms. However, Radford University launched the most user-friendly: https://sites.radford.edu/~rsheehy/Gen_flash/popgen/.

The changes mentioned above in the genetic structure of a population can be generalized to the genome level, even though they are presented here at the level of a single locus.

Since genetic drift reduces a population's heterozygosity, it can negatively impact the population's adaptation (if not offset by the effects of other evolutionary processes). We will discuss adaptation later. In real populations, it is possible to calculate the so-called effective population size (*Ne*). Simplified, *Ne* represents the genetic variability of a real population. The closer its value is to the actual population size (*N*), the more the population resembles the parameters of an ideal population. *Ne* thus allows for the comparison of several populations and the assessment of the impact of genetic drift.

In practical applications, there are several ways to calculate *Ne*. One method considers variable population sizes across different generations. Genetic drift will naturally affect the heterozygosity of the population in subsequent generations.

Let us consider that the population size in generation *i* is *Ni*, with a total of *t* generations. When there is a rapid decline in population size in a particular generation, intense genetic drift occurs (fluctuations in allele frequencies), according to the principles previously mentioned. This phenomenon is known as the bottleneck effect. *Ne* is then calculated using the harmonic mean:

$$\frac{1}{N_e} = \frac{1}{t}\left[\frac{1}{N_1} + \frac{1}{N_2} + \cdots + \frac{1}{N_t}\right] \tag{3.4}$$

Similarly, drift can occur when the number of individuals of one gender exceeds the number of individuals of the other. This imbalance indirectly increases the intensity of genetic drift by violating the expectation of random mating. Let N_m and N_f represent the number of male and female individuals in the population, respectively. Effective population size (N_e) is then calculated using the equation that incorporates the doubled harmonic mean:

$$N_e = \frac{4N_m N_f}{N_m + N_f} \tag{3.5}$$

The practical implications for forest tree populations will be explained in subsequent chapters.

The primary implication for silviculture and forest management is using solitary trees for natural regeneration. This approach inherently limits genetic diversity, as demonstrated in Fig. 3.2.

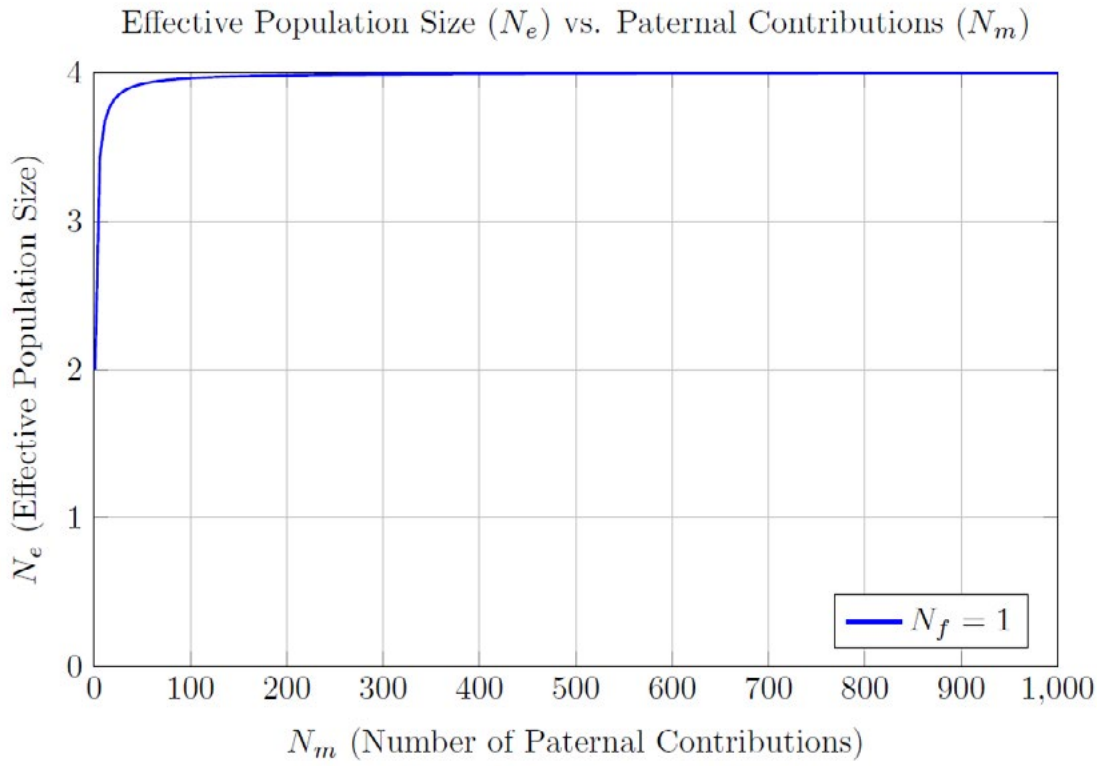


Fig. 3.2 Effective population size (N_e) as a function of the number of paternal contributions (N_m) for a single maternal tree ($N_f = 1$). The plot illustrates how *Ne* increases asymptotically with N_m, approaching a theoretical maximum of 4 as $N_m \to \infty$.

3.3.2 MUTATIONS

Mutations are integral to the process of evolution and the maintenance of genetic diversity within populations. While many mutations can be harmful, they also provide the raw material for natural selection to act upon, driving the adaptation and speciation of organisms. Understanding the mechanisms and consequences of genomic mutations is essential for evolutionary biology, genetics, and medicine.

In the long term, mutations are the most important source of genetic variability in populations. In populations of forest trees, significant cytological changes, such as anomalies in the number of chromosomes in karyotypes, can occur. The doubling of chromosome sets (polyploidy) has been evolutionarily significant in the emergence of new species of angiosperms.

In some species, polyploidy plays a major role, and it is common to encounter autopolyploidy (genome duplication within the same species) and, sometimes, allopolyploidy (genome duplication through the hybridization of different species). In forest trees, polyploidy is evolutionarily significant in the genera *Acacia*, *Alnus*, *Betula*, *Populus*, *Prunus*, and *Salix*.

At the DNA sequence level, we encounter transposons, sequences that change their position within the genome. Depending on their location, insertions can affect gene expression, and recombination between transposons can lead to chromosome rearrangements. The detailed study of these phenomena falls within the field of cytogenetics.

From the perspective of fundamental principles in population genetics, we are particularly interested in simple mutations that occur at very low frequencies at individual loci. The probability of a single allelic substitution at one locus per generation is generally in the range of 10^{-4} to 10^{-6}. For forest trees, it is typically in the range of 10^{-5} to 10^{-6}. However, the absolute frequency of new mutations is significant from an evolutionary standpoint. For example,

human DNA consists of approximately 3×10^9 nucleotide pairs. This results in an average of six new mutations per individual. Multiplying this by the total population size of 7.8 billion, we get approximately 47 billion new mutations per generation. This substantial increase in genetic diversity compensates for the loss of diversity due to genetic drift or selection. Similarly, the visibility and importance of simple mutations in forest trees are often amplified by the large population sizes covering vast species ranges.

In the following demonstration, we will return to the concept of an ideal population and relax one assumption by allowing allelic substitution at a frequency μ in one generation (see the previously mentioned frequency range of 10^{-5} to 10^{-6}). For simplicity, we will consider two alleles, A_1 and A_2, with frequencies p_1 and p_2. The expected intergenerational change in the frequency of allele A_1 (from generation t–1 to generation t) is then given by:

$$p_{1(t)} = p_{1(t-1)}(1 - \mu) \tag{3.6}$$

As evident, changes in allele frequencies due to mutations will be negligible throughout a few generations. However, long-term changes across many generations (from the initial generation 0 to generation t) can be expressed by the following formula:

$$p_{1(t)} = p_{1(0)}(1 - \mu)^t \tag{3.7}$$

The time frame to capture recurrent allelic substitution will be on the order of thousands of generations (Fig. 3.3).

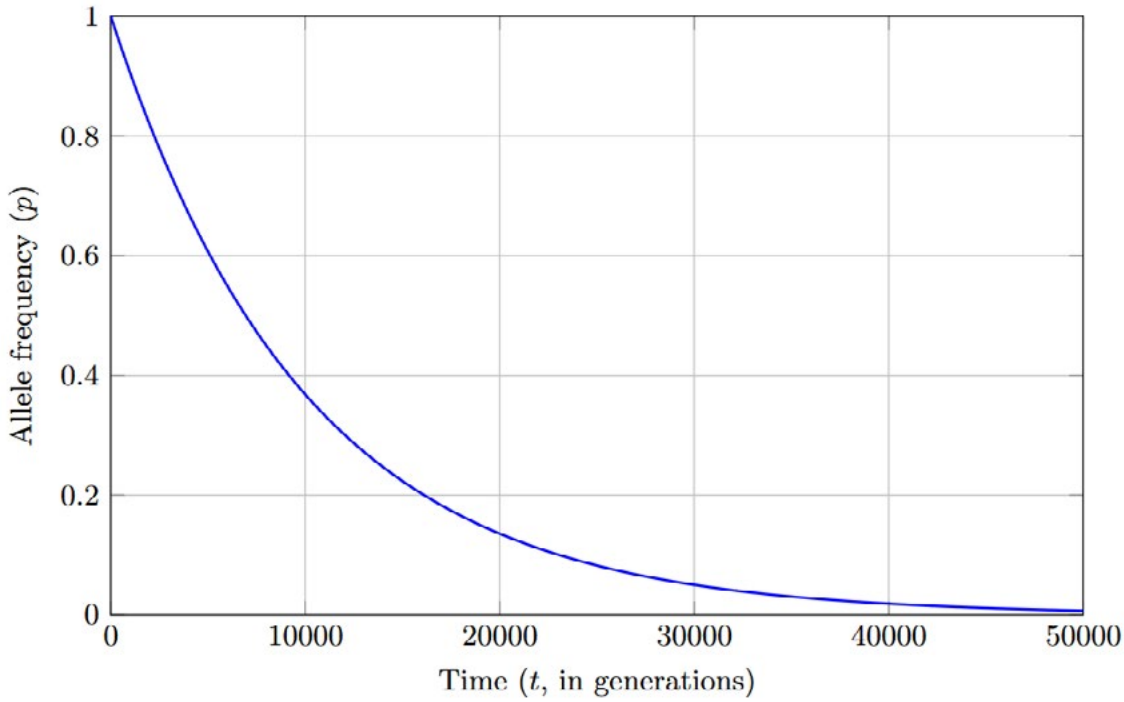


Fig. 3.3 Change in allele frequency under mutation pressure. In this example, allele *A* mutates to *a* at a rate of $\mu=1\times10^{-4}$ per generation; p_t is the allele frequency of *A* in generation *t*. We assume that $p_0=1$. With the given value of μ, the allele frequency decreases by half every 6931 generations. Modified from (Hartl & Clark, 2006).

Most mutations are selectively neutral because they either occur in non-coding regions of the genome, where they do not affect protein function, or they are synonymous mutations in coding regions that do not change the amino acid sequence due to the redundancy of the genetic code. Additionally, even when mutations alter proteins, many of these changes are minor, so they do not significantly impact the organism's fitness. This phenomenon is supported by Kimura's Neutral Theory of Molecular Evolution, which posits that most molecular changes are driven

by genetic drift rather than selection, corroborated by genomic studies and comparative genomics showing a high prevalence of neutral mutations in various species (Eyre-Walker & Keightley, 2007; Kimura, 1968).

Mutation-drift equilibrium is the balance between introducing new mutations and removing genetic variation due to genetic drift. In a population, mutations continuously generate new alleles, while genetic drift, the random fluctuation of allele frequencies, can lead to the loss of alleles, especially in small populations. At equilibrium, the rate at which new mutations arise is balanced by the rate at which genetic drift eliminates them, maintaining a steady level of genetic diversity. This concept is crucial in population genetics, as it helps explain the genetic variability observed in natural populations and the interplay between evolutionary forces.

In contrast to the Hardy-Weinberg model, mutation-drift equilibrium explicitly considers the impact of mutation and genetic drift on allele frequencies. While Hardy-Weinberg provides a baseline for understanding genetic variation in idealized conditions, mutation-drift equilibrium reflects a more dynamic and realistic scenario, where mutations introduce new genetic variations and genetic drift removes them, particularly in small populations.

In natural populations of forest trees, several factors influence genetic diversity and structure:

Large Population Size: Forest trees often exist in large populations, which can mitigate the effects of genetic drift. However, local populations or fragmented populations may experience significant drift.

Long Lifespan and Generation Time: Forest trees have long lifespans and extended generation times, which can slow down the impact of mutation and genetic drift compared to organisms with shorter lifespans.

High Mutation Rates: Trees, like many plants, can exhibit relatively high mutation rates, contributing significantly to genetic variability, which is crucial for adaptation to changing environmental conditions.

Polyploidy: Many forest tree species exhibit polyploidy, which increases their genetic makeup's complexity and can buffer harmful mutations' effects, maintaining higher genetic diversity.

Gene Flow: Trees often have mechanisms for long-distance pollen and seed dispersal, promoting gene flow between populations and counteracting the effects of genetic drift.

Equilibrium in Forest Trees

In forest tree populations, mutation drift equilibrium is crucial in maintaining genetic diversity. The continuous introduction of new mutations ensures a reservoir of genetic variation, while genetic drift, particularly in small or isolated populations, can lead to the loss of alleles. The balance between these forces determines the genetic structure of tree populations over time. In large, connected populations, the effect of genetic drift is minimized, and the genetic variation introduced by mutations is more likely to be retained, leading to higher overall genetic diversity. In contrast, in small or fragmented populations, genetic drift can significantly reduce genetic diversity, making the populations more vulnerable to environmental changes and diseases.

3.3.3 MIGRATION

Another evolutionary force acting on populations is migration. Migration is also called gene flow because it involves the movement of individuals from one population to another. Some populations may experience gene flow, while for others, it is rare. The extremes are defined such that if there is no gene flow between populations, they can be considered different species. Conversely, if gene flow is too frequent, populations can merge into one population. Here, we will demonstrate the consequences of migration between two populations. Migration is an evolutionary force that increases genetic diversity in a population. Migrants often bring completely new variants, making migration a more frequent source of advantageous novelties at the population level than mutation.

Random genetic drift causes genetic divergence (differentiation) of populations. This process is further supported by local adaptation (selection). Migration counteracts these processes by homogenizing populations, thus preventing the reduction of heterozygosity (genetic diversity) within populations. The impact of migration on the genetic structure of a population can be illustrated at the level of a single locus. As with mutations, we will assume that the population is in Hardy-Weinberg equilibrium and that the only deviation from the assumptions of an ideal population is the potential for gene flow (migration).

Let's again consider two alleles, A_1 and A_2, with frequencies p_1 and p_2, respectively, in a population we will denote as Population A. Additionally, we define the frequencies of the same alleles P_1^* P_2^* in Population B. Let's assume a one-way migration of gametes from Population B to Population A. This results in a gradual change in allele frequency (homogenization favoring the frequency in Population A). The intensity of this change is influenced by the migration coefficient $m \in \langle 0;\ 1 \rangle$, which quantifies the proportion of gametes in Population A originating from Population B. The expected intergenerational change in the frequency of allele A_1 in Population A (from generation t-1 to generation t) is then given by:

$$p_{1(t)} = p_{1(t-1)}(1 - m) + \dot{p_1}m \tag{3.8}$$

The allele frequencies in Population B remain unchanged (one-way migration).

The accumulated change in allele frequency in Population B from generation 0 to Generation T can then be expressed by the following formula:

$$p_{1(t)} = \dot{p_1} + \left(p_{1(0)} - \dot{p_1}\right)(1 - m)^t \tag{3.9}$$

The intensity of migration is often very high in forest trees. Estimates for forest trees range in the tens of percent of gametes (e.g., pollen grains) originating from more distant populations.

Migration in trees primarily occurs through the movement of pollen and seeds, the two main forms of gamete dispersal. Here are the actual forms of migration in trees:

Pollen Dispersal impacts on population structure

Wind Pollination (**Anemophily**): Many tree species rely on wind to carry pollen grains over long distances. This can result in gene flow between distant populations. Animal Pollination (zoophily): Some tree species are pollinated by animals, such as bees, birds, or bats. These pollinators can transfer pollen over varying distances, depending on their range of movement.

Seed Dispersal:

Wind Dispersal (**Anemochory**): Seeds of some trees are adapted to be carried by the wind. Examples include seeds with wings or other structures that allow them to be blown away from the parent tree. Animal Dispersal (**Zoochory**): Animals can transport seeds either externally (epizoochory) by carrying them on their bodies or internally (endozoochory) by consuming the seeds and later excreting them at different locations. This can lead to seeds being dispersed over long distances. Water Dispersal (**Hydrochory**): Trees near water bodies may have seeds that can float and be carried away by water currents. This type of dispersal can transport seeds to new areas downstream, and it often carries whole plant parts that can take root in new locations.

Water dispersal of seeds and whole plant parts is particularly relevant in several tree genera commonly found near water bodies. Notable examples include Salix (willows) and Populus (poplars), which thrive along rivers and streams, as well as Taxodium (bald cypress) and Nyssa (tupelo), which are prevalent in swamps and wetlands. Mangrove species like Rhizophora and Avicennia in coastal and estuarine environments also rely on water dispersal. Additionally, Platanus (sycamores) often grow along riverbanks and benefit from this form of propagation. This dispersal method aids these trees in colonizing new areas and maintaining genetic diversity within their populations.

Human-Mediated Dispersal:

Planting and Reforestation: Humans often transport tree seeds or young plants for planting in new areas, contributing to the migration and establishment of tree species in different regions.

Accidental Transport: Seeds can be inadvertently transported by humans through various means, such as clothing, vehicles, or agricultural equipment.

These forms of migration contribute to genetic diversity and the adaptation of tree populations to changing environments, playing a crucial role in the survival and evolution of forest ecosystems.

Wahlund Principle

The **Wahlund Principle** explains how the genetic structure of a population can be affected by its sub-division into smaller, genetically distinct sub-populations. It describes a situation where the overall population shows a deficit of heterozygosity compared to what would be expected under Hardy-Weinberg equilibrium due to the presence of a sub-population structure. This principle has significant implications in migration and population genetics, as demonstrated in BOX 3.2.

Key Points of the Wahlund Principle:

Sub-Populations: The overall population is divided into several sub-populations. Each may have different allele frequencies due to various evolutionary forces such as genetic drift, selection, and mutation.

Deficit of Heterozygosity: When these sub-populations are considered together as a single population, the overall heterozygosity (the proportion of heterozygous individuals) is lower than expected if the population were randomly mating. This is because the sub-populations are not in Hardy-Weinberg equilibrium with each other.

Genetic Differentiation: Genetic differentiation among sub-populations leads to excess homozygosity when pooling the entire population.

Migration involves the movement of individuals between sub-populations. This movement can have several impacts on the genetic structure of the population: (1) **Gene Flow**: Migration introduces new alleles into a sub-population, increasing genetic diversity and potentially altering allele frequencies; (2) **Reduction in Genetic Differentiation**: As individuals migrate and breed with members of other sub-populations, the genetic differences between sub--populations decrease. Over time, this can lead to homogenization of allele frequencies across the population; (3) **Impact on Heterozygosity**: Initially, if there is limited migration, sub-populations may remain genetically distinct, leading to a Wahlund effect (a deficit of heterozygosity). However, as migration increases and sub-populations interbreed more frequently, the overall population may move closer to the Hardy-Weinberg equilibrium, reducing the Wahlund effect.

BOX 3.2

Imagine a scenario with two sub-populations of a tree species in separate forest patches. Each sub-population has different allele frequencies for a specific gene:

Sub-population 1: Allele A_1 frequency = 0.7, Allele A_2 frequency = 0.3

Sub-population 2: Allele A_1 frequency = 0.3, Allele A_2 frequency = 0.7

If there is no migration, each sub-population might be in Hardy-Weinberg equilibrium on its own. Still, when combined, the overall population will show a heterozygosity deficit because the two sub-populations have different allele frequencies.

However, if individuals start migrating between the two patches and interbreeding, the allele frequencies will begin to equalize. Over many generations, the genetic structure of the combined population will stabilize, and the heterozygosity will approach the levels expected under Hardy-Weinberg equilibrium, diminishing the Wahlund effect.

3.3.4 SELECTION

Natural selection is the central evolutionary force enabling organisms to adapt to their environments. Darwin (1859) proposed **natural selection** based on three premises: (1) More offspring are produced than can survive and reproduce; (2) Individuals vary in their ability to survive and reproduce, with some of these differences being genetically controlled; and (3) Genotypes that enhance survival or reproduction contribute disproportionately more offspring to the next generation, increasing the frequency of their alleles at the expense of less adapted genotypes. This process progressively improves a population's adaptation over generations.

The evolutionary biologist Ernst Mayr expressed the essence of natural selection in 1988: *Selection is not a forward-looking process but simply a name for the survival of those few individuals that have successfully outlasted the "struggle for existence."* (Mayr, 1988).

We might wonder why certain natural individuals have better prospects than others for passing their genes to the next generation. In plants, this could involve producing more flowers to attract more pollinators, having greater resistance to diseases, or tolerating lower temperatures. Among animals, it might be about attracting mates of the opposite sex or having a larger litter size. These traits are certainly not the only ones contributing to an individual's fitness.

One of the most prominent quantitative geneticists, Ronald Fisher, introduced a concept of great significance for natural selection in 1930: **the fundamental theorem of natural selection** (Fisher, 1958). While the detailed derivation of this concept is beyond the scope of this book, its essence states that *the rate of increase in mean fitness at any time is equal to the additive genetic variance of fitness at that time.*

In addition to natural selection, **artificial selection** is a key tool breeders exploit to enhance desirable traits in tree improvement programs (Chapter 6 on Tree Improvement). Both natural and artificial selection operate on similar principles, which are discussed here.

Although selection acts on phenotypes resulting from many gene loci, its consequences are best understood by examining how it changes allele frequencies at a single locus. Without other evolutionary forces to counteract it, **directional selection** eventually leads to the fixation of the favored allele in a population. Directional selection drives progressive adaptation to new environments and is widely applied by breeders in improvement programs.

Conversely, **stabilizing selection** or synonymously balancing selection maintains polymorphism by selecting two different alleles at the same locus. This type of selection is crucial for preserving genetic variation within populations, which is essential for continued adaptive evolution in changing environments. These three main types of natural selection are visualized in Fig. 3.4. The individuals with the highest fitness in a normal distribution depend on the environmental conditions.

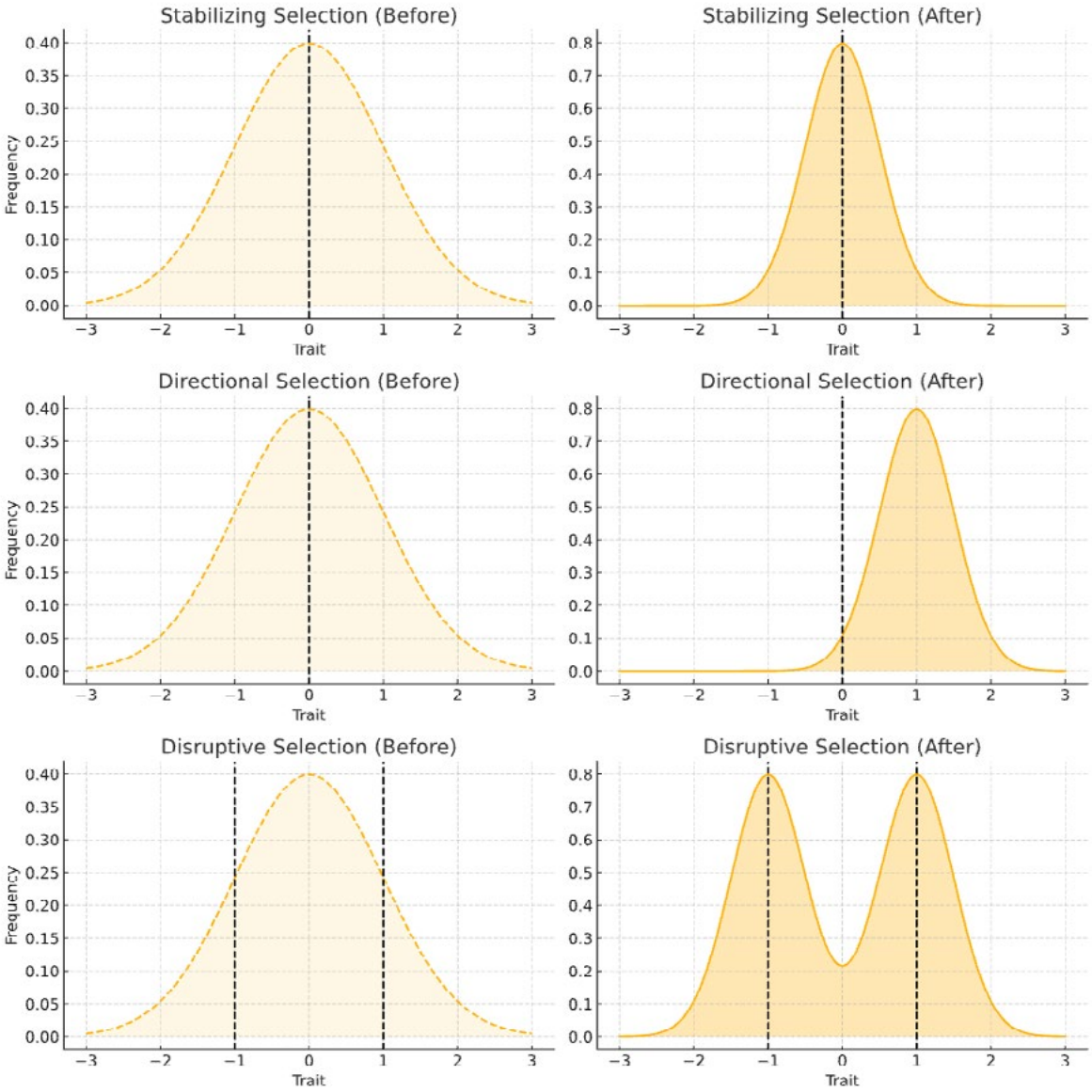


Fig. 3.4 The three main types of natural selection and the result of these types of selection.

Directional selection

Under the harsh conditions after the last glaciation, it is easy to see that plants in the left tail of the normal distribution had the highest fitness. This scenario favors directional selection. Directional selection is particularly significant in populations migrating along an ecological gradient or cline. While the hardiest individuals in the left tail were favored after the glaciation, the opposite tail may now have the highest fitness as temperatures rise due to increasing greenhouse gases in the atmosphere.

We can illustrate directional selection using alleles labeled 1 and 2. Assume that individuals in one tail predominantly have index 1 alleles, while those in the opposite tail mainly possess index 2 alleles. Depending on the direction of selection, the frequency of one allele will increase in the progeny. If index 1 alleles contribute to fitness under harsh conditions, an increase in the frequency of index 2 alleles is expected under global warming.

It is worth emphasizing that stabilizing and directional natural selection reduce genetic variation within a population while increasing variation among different populations. When considering phenotypic values, we observe that stabilizing selection within populations can manifest as disruptive selection among populations.

To understand the rate of change driven by directional selection, we must introduce an expression that quantifies the impact of selection on a particular genotype. For this purpose, we assume that AA, Aa, and aa genotypes have fitness values of 1, 1, and 1-**s**, respectively. Here, **s** is known as the **selection coefficient** and should not be confused with the selection differential (discussed in the next chapter on quantitative genetics). The selection coefficient is often expressed as a percentage.

In this context, the allele frequency of A is designated by p, and the a allele frequency is defined by q. By applying some algebra, which we will not detail here, we can derive the changes in allele frequency of a before and after natural selection. This derivation results in the following expression for the change in q, denoted as Δq:

$$\Delta q = -\frac{spq^2}{1 - sq^2} \tag{3.10}$$

Formula, valid under complete dominance, gives interesting information on the prerequisites for changes in allele frequencies via natural selection. The selection coefficient (s) quantifies the less-fit genotype's reduced reproductive success relative to the more-fit genotype. For example, if s=0.1, the fitness of aa individuals is 90% that of *AA* and *Aa* individuals.

The maximum speed of change occurs at an allele frequency of one-third of a allele. This can be understood intuitively because when q is one-third and p is two-thirds, both alleles have a significant presence in the population, allowing for noticeable shifts in allele frequency under selection pressure. The rate of change is highest when both alleles are common in the population. This is because the product pq (which appears in the numerator of the equation) is maximized when both p and q are around 0.5. In this scenario, there is a lot of genetic variation for selection to act upon. The equation also shows that the speed of change is largest when both alleles are common, while the change is minor when one allele is common and the other rare. The equation also tells us that the larger the value of s, the larger the speed of change.

Table 3.1 General model for allele frequency change under directional selection

Genotypes	A_1A_1	A_1A_2	A_2A_2	Total
Initial frequency	p^2	$2pq$	q^2	1.0
Relative fitness	1	$1-hs$	$1-s$	
Proportion after selection	p^2	$2pq(1-hs)$	$q^2(1-s)$	$T = 1 - 2pqhs - sq^2$
Frequency after selection	$\frac{p^2}{T}$	$\frac{2pq(1-hs)}{T}$	$\frac{q^2(1-s)}{T}$	1.0

Table 3.1 illustrates a general model for allele frequency change under directional selection. Here's a breakdown of the columns and the calculations:

Following the same procedures outlined in the numerical example, a rather complicated formula for Δq *is derives* after one generation of selection (Table 3.1). This expression simplifies to:

$$\Delta q = -\frac{sq^2(1-q)}{1-sq^2} \text{ (when } A_2 \text{ is recessive)} \tag{3.11}$$

$$\Delta q = -\frac{sq(1-q)^2}{1-sq(2-q)} \text{ (when } A_2 \text{ is dominant)} \tag{3.12}$$

$$\Delta q = -\frac{1}{2}\frac{sq(1-q)}{1-sq} \text{ (when there is no dominance)} \tag{3.13}$$

Regardless of the degree of dominance, the effect of directional selection against A_2 is to decrease its frequency in the population. However, the magnitude of change in allele frequency depends on the selection intensity (s) and the initial allele frequency (q).

The impact of allele frequency and dominance on the effectiveness of directional selection is illustrated in Fig. 3.5 for the three cases of dominance above when s = 0.2. For example, assume that the deleterious allele is recessive but nearly fixed in the population (i.e., q is near 1). This figure shows that it would take about 22 generations of directional selection to reduce the frequency of this allele to 0.50 and another 15 generations to reduce it to 0.25. However, selection for or against a recessive allele is very ineffective (i.e., little change in allele frequency) when the allele is rare (i.e., going from q = 0.05 to 0.025 would take approximately 100 generations). As shown earlier in this chapter, when a recessive allele is rare, it occurs mostly in heterozygotes and, therefore, is not exposed to selection. On the other hand, selection against a dominant allele is least effective when it is common (i.e. when the favored recessive allele is rare).

How effective is selection against an unfavorable allele if s = 1, which is the case of a lethal allele in a natural population, or the selective removal of all individuals expressing the allele in a breeding program? Only a single generation is required to remove the allele when it is dominant, but when recessive, it is increasingly difficult to remove as it becomes rare. Substituting s = 1 in Equation 3.11, $\Delta q = -q^2/(1+q)$, so that when q = 0.01, the change in q expected from removing all homozygous recessive individuals would be only 0.0001. Therefore, increasing the selection intensity, even to the maximum, has little additional impact on removing rare, recessive alleles when mating is at random.

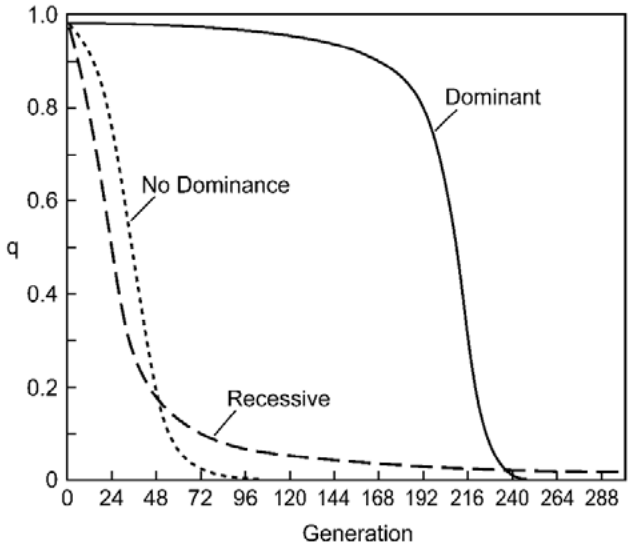


Fig. 3.5 Frequency of a deleterious allele (q) undergoing directional selection in a large, random mating population when its effect is dominant, recessive, or without any dominance. All plots assume that the deleterious allele is nearly fixed in the population at generation 0 and that the selection coefficient against this allele is 0.2. Values were originally calculated using the *Populus* computer program (University of Minnesota, https://cbs.umn.edu/populus/download), and the figure was modified from White et al. (2007)

Implications for breeding

Removing deleterious alleles in breeding populations can be greatly accelerated through progeny testing (Falconer & Mackay, 2009). By crossing parents displaying the recessive phenotype with all trees in the breeding population, heterozygous trees for the deleterious allele can be identified by the presence of recessive homozygotes among their progenies. Using progeny test data to eliminate heterozygous trees and parental phenotypes to eliminate recessive homozygotes allows for removing the recessive allele in a single generation, barring any misclassification of genotypes. This process is particularly effective in easily detectable traits, such as fungal resistance in trees.

Stabilizing selection

A prerequisite for stabilizing selection is that the mean value of the population remains unchanged; thus, stabilizing selection is the type of selection most common in stationary populations. This means that the individuals in the two tails of the normal distribution have the lowest fitness values. In situations without any heterozygotic advantage, all subsequent selections within a population lead to an increase in homozygosity. Since the mean value remains unchanged at stabilizing selection, the selection cannot lead to homozygosity for all positively acting alleles.

Stabilizing selection is a key evolutionary force that maintains phenotypic stability by favoring intermediate traits and selecting against extremes, ensuring the population mean remains unchanged. This process reduces genetic variation and increases homozygosity, though complete homozygosity is unlikely as intermediate phenotypes are preserved. While stabilizing selection promotes evolutionary stasis and adaptation to stable environments, it limits genetic diversity and adaptive potential, making populations vulnerable to rapid environmental changes. It can also constrain evolutionary innovation by suppressing novel traits and may slow speciation by maintaining phenotypic uniformity. Despite these limitations, stabilizing selection buffers populations against minor environmental fluctuations, highlighting its dual role in fostering stability and restricting adaptability.

Disruptive selection

If individuals in both tails of a trait's normal distribution have the highest fitness, we will observe disruptive selection. In high elevations or far northern latitudes with heavy snowfall during long winters, trees may benefit either from having narrow crowns that allow snow to slide off easily or from having extremely strong branches that can support the weight of the snow.

A narrow crown is likely the result of strong natural selection, as it helps prevent branches from breaking under the weight of snow. When branches break, the tree's crown is reduced, decreasing photosynthesis and growth. This, in turn, means fewer flowers and a lower chance of passing on the tree's genes to the next generation. Conversely, a tree with strong branches can support large amounts of snow without damage.

Thus, trees at both ends of the distribution—those with narrow crowns and strong branches—exhibit high fitness. This creates a scenario that favors disruptive selection, where extreme traits on both ends of the spectrum are advantageous, while intermediate traits may be less beneficial.

Over time, disruptive selection can promote divergence within populations, potentially leading to speciation as extreme phenotypes become more distinct. However, it may also create trade-offs as populations become polarized and less adaptable to changes that favor intermediate traits. Thus, disruptive selection drives diversity by rewarding specialized adaptations to distinct environmental challenges.

3.4 SUMMARY

Population genetics provides a foundational understanding of how genetic variation is maintained and shaped within populations over time. Through concepts like the Hardy-Weinberg equilibrium, we establish baselines for studying the genetic structure and the forces that disrupt it—mutation, genetic drift, migration, and selection. These evolutionary forces operate independently and interactively, influencing allele frequencies and driving the diversity essential for species' survival and adaptability.

For tree populations, the significance of population genetics lies in its ability to explain the genetic mechanisms that drive their resilience and adaptability. Tree populations often experience unique challenges, such as long lifespans, extended generation times, and exposure to diverse environmental pressures. Understanding genetic diversity in trees is critical for maintaining their capacity to adapt to climate change, disease, and habitat fragmentation. Concepts like effective population size, mutation-drift equilibrium, and gene flow offer insights into maintaining genetic variability, especially in fragmented or isolated populations. These principles are essential for sustainable forestry practices, conservation efforts, and the development of tree improvement programs, ensuring the long-term viability and ecological contributions of forest ecosystems.

3.4 REFERENCE

Darwin, C. (1859). On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life [Digital Book]. John Murray. https://library.si.edu/digital-library/book/onoriginofspec00darw

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, *8*(8), 610–618. https://doi.org/10.1038/nrg2146

Falconer, D. S., & Mackay, T. (2009). *Introduction to quantitative genetics* (4. ed., [16. print.]). Pearson, Prentice Hall.

Fisher, R. A. (1958). *The genetical theory of natural selection.* https://www.cabidigitallibrary.org/doi/full/10.5555/19601600934

Hartl, D. L. (1999). *A Primer of Population Genetics* (3rd edition). Sinauer Associates is an imprint of Oxford University Press.

Hartl, D. L., & Clark, A. G. (2006). *Principles of Population Genetics* (4th edition). Sinauer Associates is an imprint of Oxford University Press.

Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, *217*(5129), 624–626. https://doi.org/10.1038/217624a0

Mayr, E. (1988). *Toward a New Philosophy of Biology: Observations of an Evolutionist*. Harvard University Press.

White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI.

4. INTRODUCTION TO QUANTITATIVE GENETICS

Qualitative traits are typically categorical and discrete, often following simple Mendelian inheritance patterns influenced by one or a few genes with large effects (see Chapter Two for more details). They are measured descriptively or categorically, such as color, shape, or the presence or absence of a characteristic. Examples include flower color in pea plants (purple or white), blood type in humans (A, B, AB, or O), and seed shape (round or wrinkled). These traits are usually visible and easily distinguishable, controlled by a few major genes, and minimally influenced by environmental factors.

On the other hand, **quantitative traits** are continuous and vary along a gradient, influenced by multiple genes (polygenic), each contributing a small effect. These traits are measured numerically and often require statistical analysis to quantify. Examples include height, weight, yield, and other measurable characteristics. Quantitative traits exhibit a range of phenotypes that are not easily categorized, such as the continuous variation in human height. They are controlled by many genes and significantly influenced by environmental factors, making their expression more variable and complex.

Quantitative inheritance arose in the early 20th century as scientists sought to understand traits that did not follow simple Mendelian inheritance patterns. While Gregor Mendel's work laid the foundation for genetics, his laws primarily explained inheritance for traits controlled by single genes with clear dominant and recessive alleles. However, many characteristics in plants and animals, such as height, weight, and yield, did not fit this pattern and showed continuous variation.

Quantitative traits are those that occur on a continuous scale within a population and typically follow a normal distribution. A common example of a quantitative trait is an individual's height. It has long been known that offspring tend to resemble their parents in these traits.

Robert Bakewell, a pioneering breeder in the 18th century, systematically tracked pedigrees and selected the best parental individuals for breeding. This systematic approach laid the groundwork for modern breeding techniques. Sir Francis Galton, a cousin of Charles Darwin, later applied statistical methods to study these traits, directly building upon Darwin's ideas.

4.1 CONCEPT OF QUANTITATIVE INHERITANCE:

The concept of quantitative inheritance arose from key developments in genetics and statistical analysis in the early 20th century. Around 1900, Mendel's laws were independently rediscovered by Hugo de Vries, Carl Correns, and Erich von Tschermak, revitalizing the field of genetics and leading researchers to apply Mendelian principles to a wide range of traits. Even before this rediscovery, Francis Galton studied continuous variation and heredity, introducing concepts such as regression and correlation. Galton's work on human height demonstrated that such traits followed a **normal distribution** and suggested a hereditary component influenced by multiple factors. William Bateson, a strong advocate of Mendelian genetics, recognized that many traits did not follow simple Mendelian ratios and proposed that multiple genes might influence these traits.

The formal concept of quantitative inheritance was solidified by Ronald A. Fisher in 1918 with his seminal paper "The Correlation Between Relatives on the Supposition of Mendelian Inheritance." Fisher demonstrated that continuous variation could be explained by the combined effect of many genes, each contributing a small amount to the overall phenotype. This groundbreaking work integrated Mendelian genetics with the biometrical school's statistical methods, which studied continuous variation and heredity. Fisher's model showed how multiple genes contribute to quantitative traits, providing a mathematical and statistical framework that bridged the gap between Mendelian genetics and the observed continuous variation in traits like height, weight, and yield. This integration laid the foundation for population genetics, furthering our understanding of how genetic variation and environmental factors influence quantitative traits (Fisher, 1919).

The development of the concept of quantitative inheritance relied heavily on mathematical models and statistical methods, such as variance analysis, correlation, and regression, to analyze traits with continuous variation. Experimental breeding studies with plants and animals showed that traits like height, weight, and yield followed

patterns consistent with the additive effects of multiple genes. The modern synthesis of the 1930s and 1940s, which unified Mendelian genetics with Darwinian evolution, further reinforced the understanding of quantitative inheritance by highlighting how genetic variation in populations can lead to evolutionary changes over time. This integration of Mendelian principles with statistical methods and experimental breeding studies, driven by key contributions from scientists like Francis Galton, William Bateson, and especially R.A. Fisher, established the framework explaining how multiple genes contribute to the continuous variation observed in many traits.

4.1.1 THE SIGNIFICANCE OF NORMAL DISTRIBUTION IN QUANTITATIVE TRAITS

The normal distribution plays a crucial role in quantitative inheritance because it provides a statistical framework for understanding the continuous variation observed in many traits influenced by multiple genes.

Representation of Continuous Variation:

Quantitative traits, such as height, weight, and yield, exhibit continuous variation rather than discrete categories. When these traits are measured across a population, the values typically form a bell-shaped curve or normal distribution. This pattern arises because the combined effect of many small, independent genetic and environmental factors tends to produce a distribution where most individuals cluster around the mean, with fewer individuals at the extremes.

Central Limit Theorem:

The normal distribution in quantitative inheritance is underpinned by the central limit theorem. This statistical principle states that when multiple independent factors (such as genes) contribute to a trait, the sum of their effects will approximate a normal distribution, regardless of the individual distributions of these factors. Thus, even if the contributions of individual genes are not normally distributed, the overall trait will be.

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \sim N(0,1) \qquad (4.1)$$

The equation represents a standard normal distribution with a mean of 0 and a variance of 1.

The normal distribution allows various statistical tools and methods, such as variance analysis, correlation, and regression, to predict and analyze quantitative traits. By assuming a normal distribution, researchers can make inferences about the genetic architecture of traits, estimate the effects of individual genes, and predict the outcomes of breeding experiments.

4.2 CONCEPT OF HERITABILITY IN BREEDING PROGRAMS AND EVOLUTIONARY STUDIES

In quantitative genetics, the normal distribution helps estimate **heritability**, the proportion of phenotypic variance attributable to genetic variance. This is done by partitioning the total variance of a trait into genetic and environmental components. The normal distribution provides a basis for these calculations and helps us understand how much of the variation in a trait is due to genetic factors versus environmental influences.

In crop and animal breeding, the normal distribution identifies and selects individuals with desirable traits. Analogically, this concept has been implemented in tree improvement. By understanding the distribution of traits in a population, breeders can select individuals from the tails of the distribution (those with extreme phenotypes) to produce offspring with enhanced or reduced trait values.

Heritability vs heredity

Heritability and heredity are related but distinct concepts in genetics, which are crucial for understanding how traits are passed down and vary within populations. Heredity refers to passing traits from parents to offspring through genes, encompassing the entire process of genetic inheritance, including the segregation and recombination of alleles, and explaining why offspring resemble their parents. Heritability, however, is a specific statistical concept

that quantifies the proportion of the total variation in a trait within a population due to genetic differences among individuals, expressed as a **value between 0 and 1**. Key points of contrast include their scope—heredity encompasses all genetic transmission processes, while heritability focuses on genetic contributions to trait variation within a population; measurement—heredity is qualitative, not quantified, whereas heritability is a quantitative measure. When applied, heredity applies to individual inheritance, while heritability applies to populations. The major difference also lies in environmental influence—heredity considers genetic transmission without direct reference to the environment, while heritability considers the interplay between genetics and the environment. In research, heredity is fundamental for understanding genetics and inheritance patterns. At the same time, heritability is used in quantitative genetics to assess genetic factors in trait variation and inform breeding programs and evolutionary studies (demonstrated in BOX 4.1).

BOX 4.1

Heritability vs Heredity

Consider height in humans. Heredity explains why children tend to be similar in height to their parents because they inherit genes that influence height. Heritability quantifies how much of the variation in height within a specific population can be attributed to genetic differences. For instance, if the heritability of height is 0.8 in a population, it means that 80 % of the variation in height among individuals in that population is due to genetic differences, with the remaining 20 % due to environmental factors and random variation.

In summary, heredity refers to the general process of genetic inheritance from parents to offspring. At the same time, heritability is a statistical measure that quantifies the proportion of trait variation in a population attributable to genetic differences.

Heritability is crucial in breeding programs for selecting breeding stock, predicting breeding outcomes, and achieving genetic improvement. When a trait exhibits high heritability, genetic factors significantly influence its variation, allowing breeders to select individuals with desirable traits confidently, knowing they are likely to be inherited by the offspring. Conversely, environmental factors play a larger role for traits with low heritability, making it less certain that selected traits will be passed on. In such cases, breeders may need to manage environmental conditions carefully to achieve the desired outcomes. Heritability also aids in predicting the response to selection, which is the expected improvement in a trait due to selective breeding.

Phenotypic traits such as height and weight vary continuously among individuals in a population. The sample variance for a phenotypic trait is denoted by V_P. Phenotypic variance can arise from environmental variation (V_E), even without genetic differences. This includes variations in the organism's internal and gestational environments. Phenotypic variance can also stem from genetic differences (V_G), even without environmental variation, though this is rare.

Interactions between genes and the environment, called **genotype-environment interactions** (V_{GE}), can also contribute to phenotypic variance. For example, individuals with the *AA* genotype might grow to 8 m in **Environment J**. Still, only 4 m in **Environment K**, while individuals with the *aa* genotype might grow to 4 m in Environment J and 8 m in Environment K. Without considering both genotype and environment, the effects on height seem unpredictable.

The total phenotypic variance is the sum of these components:

$$V_P = V_E + V_G + V_{GE} \tag{4.2}$$

The fraction of phenotypic variance attributable to genetic variance is called the **"broad-sense" heritability (H^2)**:

$$H^2 = V_G / V_P \tag{4.3}$$

V_G can be divided into three components. The first is the additive genetic variance (V_A), which refers to the phenotypic variation arising from the average effects of different alleles at relevant genes. In the simplest scenario, each gene shows incomplete dominance, where the average phenotypic value of the heterozygote falls exactly between the average values for either homozygote. Additionally, there is no epistasis, meaning the genotype at one gene does not mask the genotype at another.

However, in reality, heterozygotes are not always exact intermediates between the two homozygotes, as seen with complete dominance. Genes also interact in epistatic manners. Thus, we must consider two additional components of genetic variance that can mask the effects of specific alleles: dominance genetic variance (V_D) and epistasis genetic variance (V_I).

$$V_G = V_A + V_D + V_I \tag{4.4}$$

Both dominance and epistasis can make natural selection based on phenotype less efficient. The effects of individual alleles can be masked by alleles at the same gene (dominance) or other genes (epistasis). Individuals with different genotypes can have the same phenotype, and natural selection acts on the phenotype. If the effects of every allele could be "unmasked," natural selection would more effectively select among genotypes.

The ratio of additive and total phenotypic variance corresponds to the **"narrow-sense" heritability (h^2):**

$$h^2 = V_A/V_P \tag{4.5}$$

It should be clear that this value ranges between 0.0 and 1.0 and can never exceed broad-sense heritability, as VA (additive genetic variance) can never surpass total genetic variance. Narrow-sense heritability is particularly crucial for breeders because it predicts the rate at which selective breeding can alter the average phenotype of a population. The **response to selection (R),** defined as the difference in the average phenotype between generation 1 and generation 0, depends on **selection differential (S)** (Fig. 4.1), which is the difference between the average phenotype of the selected parents from generation 0 and the average phenotype of all individuals in generation 0, which is multiplied by narrow sense heritability. The so-called breeder's equation gives the relationship**:**

$$R = h^2 S \tag{4.6}$$

The response to selection is proportional to narrow-sense heritability for a given selection differential. Without additive genetic variance for the phenotype, selection will be ineffective. Therefore, adaptation following natural selection in natural populations requires non-zero narrow-sense heritability for the trait.

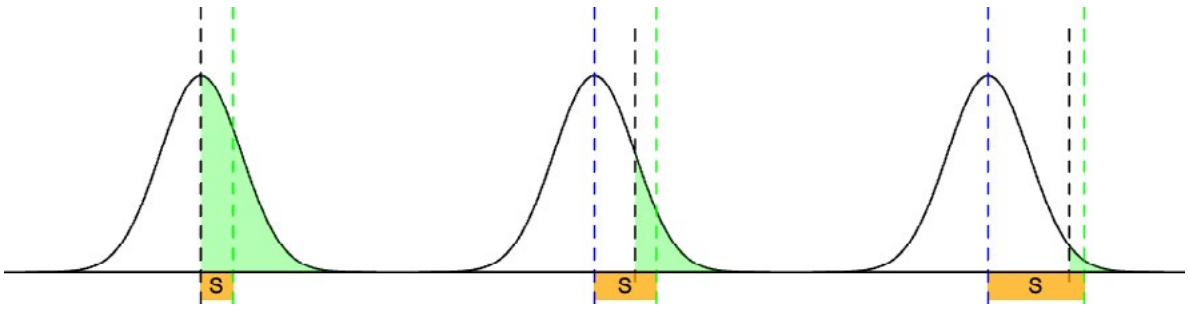


Fig. 4.1 Selection differential is related to selection intensity. Blue dashed line - population mean; green dashed line – mean of the selected parents.

It should be noted that *R* is often substituted by *G*, especially in breeding, as *G* stands for predicted genetic gain. High heritability allows for more efficient genetic improvement because the observed variation is largely due to genetic differences, making selection efforts more effective in enhancing the trait across generations.

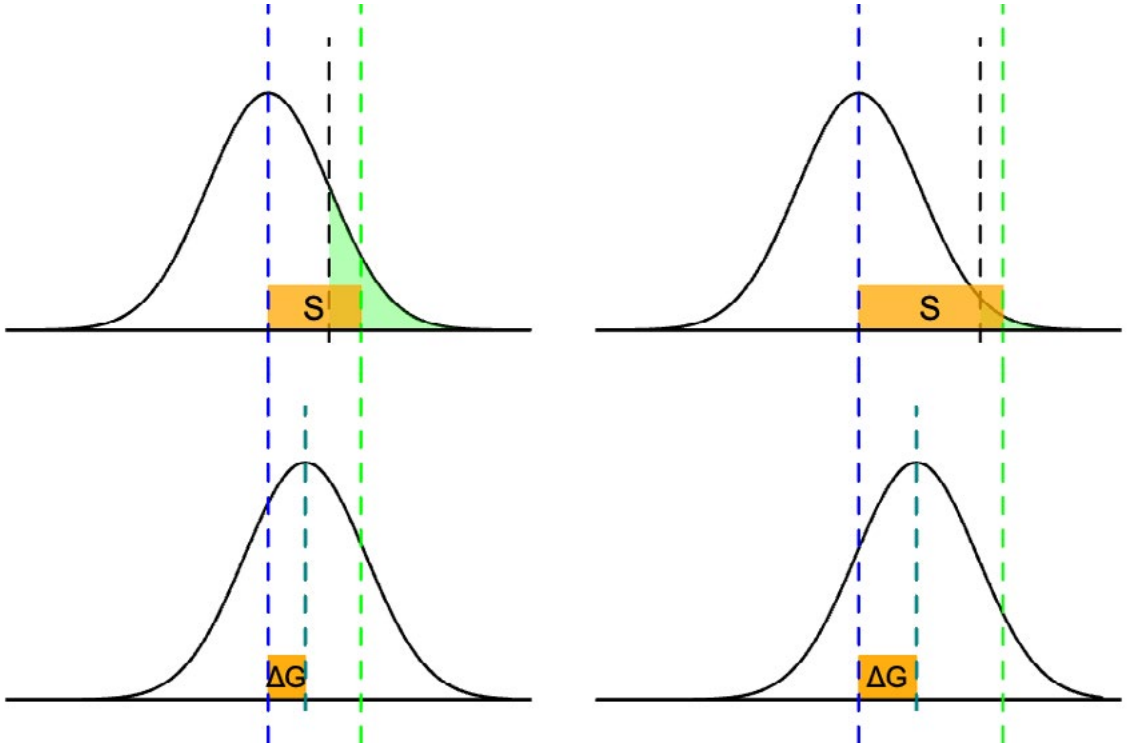


Fig. 4.2 The impact of selection differential on genetic gain in the consecutive generation.

In evolutionary studies, heritability provides critical insights into the evolutionary potential of traits, the estimation of evolutionary rates, and the understanding of genetic variation within populations. Traits with higher heritability are more likely to respond to natural selection, facilitating evolutionary change. By understanding the heritability of a trait, evolutionary biologists can estimate the rate at which the trait can evolve in response to selective pressures, which is essential for understanding how populations adapt to changing environments over time. Heritability studies also contribute to understanding how genetic variation is maintained within populations, which is vital for conservation biology. Maintaining genetic diversity is crucial for species' long-term survival, as it allows populations to adapt to environmental changes and resist diseases. Thus, heritability helps explain the current genetic makeup of populations and predicts their future evolutionary trajectories.

4.3 MAIN METHODS OF HERITABILITY ESTIMATION

Several established methods are used for estimating heritability, each with strengths and specific applications. These methods can be broadly categorized into those that use familial data and those that use population data. We must recognize that heritability estimates are specific to a particular population in a specific set of environments. Different populations have different gene frequencies, and moving a population to a new environment will change their conditions. Consequently, V_P, V_G, V_E, and V_{GE} are all influenced by both the population and the environment.

4.3.1 ESTIMATION METHODS FOR THE NARROW-SENSE HERITABILITY

Parent-Offspring Regression:

The slope of the regression line of offspring phenotypes on parent phenotypes gives the narrow-sense heritability. A more detailed explanation of its historical consequences on genetics and biometry is provided in this chapter.

Breeder's equation:

This equation mentioned above renders the so-called realized heritability:

$$h^2 = R/S \tag{4.7}$$

R is the realized response to selection, and S is the selection differential. It should be noted that the realized heritability can be calculated only after the realized genetic gain (the actual response to selection) can be obtained in the next generation. In growth traits, the actual gains will often occur at rotation age. This information may be available earlier in specific traits, such as resistance.

4.3.2 ESTIMATION METHODS FOR THE BROAD-SENSE HERITABILITY

Twin Studies:

Broad-sense heritability is estimated using the correlation between various phenotypic expressions of monozygotic twins (who share all their genes) and dizygotic twins (who share half their genes).

ANOVA or REML if data is imbalanced:

In experiments with known genetic relationships (e.g., inbred lines, clonal replicates), ANOVA can partition phenotypic variance into genetic and environmental components to estimate broad-sense heritability.

Some other methods are used in breeding, evolutionary biology, or other basic research, which will be discussed further in this chapter. However, some of them are beyond the scope of this book. Further reading can be recommended (Mrode, 2023; White et al., 2007).

Parent-Offspring Regression:

This method involves regressing the offspring's trait values on the parent's trait values. The slope of the regression line (regression coefficient) represents heritability. Commonly used in plant and animal breeding.

Full-Sib and Half-Sib Analysis:

This approach compares trait variance among full siblings (same parents) and half-siblings (one common parent). Variance components are partitioned to estimate the additive genetic variance. They are widely used in breeding programs for livestock and crops and as experimental designs for breeding trials. It is often paired with the Analysis of Variance (ANOVA), which involves partitioning the total phenotypic variance into genetic and environmental components using ANOVA. Heritability is estimated as the ratio of genetic variance to total phenotypic variance.

Restricted Maximum Likelihood (REML):

A statistical method that estimates variance components by maximizing the likelihood of observed data under a linear mixed model. REML provides unbiased estimates of heritability even with **unbalanced** data, making it particularly **relevant in forestry research**. They are commonly used in quantitative genetics and breeding applications across the fields. However, due to its inherent complexity, this method is not further elaborated in the book. For more details, see (Mrode, 2023).

Twin Studies:

It involves comparing the similarity of traits between monozygotic (identical) and dizygotic (fraternal) twins. The difference in trait correlations between the two types of twins is used to estimate heritability. They are commonly used in human genetics. Similarly, clonal trials are used in tree improvement to effectively separate genetic and environmental variance.

Genome-Wide Association Studies (GWAS):

It involves scanning the genome for markers associated with trait variation. In this case, the so-called genomic heritability is estimated from the proportion of trait variance explained by the identified genetic markers. It has been recently used in modern genetics to analyze complex traits. However, complex conifer genomes are currently challenging for GWAS due to their size and the limited availability of dense marker panels.

Genomic Best Linear Unbiased Prediction (GBLUP):

It uses genomic information to predict the genetic value of individuals and estimate heritability. It uses the REML and following BLUP, but both rely on genomic (realized) relationships. Combines genomic and phenotypic data to estimate heritability and subsequent genomic breeding values. They are increasingly used in animal and plant breeding programs.

4.3.3 PARENT-OFFSPRING REGRESSION AND HERITABILITY:

In 1886, Francis Galton published his analyses of the heights of parents and their adult offspring. The principle of calculating the heritability of height involves regression analysis, a term that Galton himself coined. He observed that children of above-average parents tend to be shorter than their parents—a phenomenon he called "regression to the mean." This observation led to the term "regression" in statistical analysis (Galton, 1886).

The slope of the regression line indicates the degree of resemblance between parents and their offspring. A slope of zero means no relationship exists between the parents' and offspring's traits, while a slope of one indicates that offspring, on average, match their parents' traits. This method allows the estimation of heritability—a measure of how much variation in a quantitative trait can be attributed to genetic factors. Key aspects of regression include the slope estimate and the standard error of this estimate (Std. Error).

Galton's work laid the foundation for the biometric school of thought, further developed by his followers, including Karl Pearson (known for Pearson's correlation coefficient). Around 1900, the rediscovery of Mendel's laws initially described a different inheritance system, specifically for qualitative traits.

Researchers often investigate whether the value of one continuous variable depends on another. For example, does running speed depend on leg length? Does the percentage of dead roaches depend on poison concentration? Does storm damage (in millions of dollars) depend on wind speed? In these cases, the direction of dependence is clear.

The relationship between dependent and independent variables can be complex. However, there are instances when it is reasonable to assume a linear relationship, and linear regression becomes a valuable method in these scenarios. The general problem involves measuring the values of both dependent and independent variables for N data points. For example, if $N=10$, you plot the dependent variable on the y-axis (vertical axis) and the independent variable on the x-axis (horizontal axis).

The relationship between these variables can be written in standard algebraic form:

$$y = bx + c \tag{4.8}$$

Where b **is the slope of the line**, and c **is the y-intercept** (the value of y when x=0). We might try to draw an eye-fitted line through these points. However, which of these is the "best" line?

There is too much subjectivity. Linear regression offers an objective solution by choosing a line that minimizes the squared vertical deviations between the data points and the line. This method is often referred to as "**least squares linear regression**."

In a specific population, the phenotype of **offspring** may partly depend on the phenotypes of the **parents** if the parents' genes influence the phenotype. Additive genetic variance is mathematically equivalent to parent-offspring covariance $COV_{parent,\ offspring}$. Using the parents' phenotypic variance (VAR_{parent}) as a measure of VP, we can estimate h^2 using the linear regression formula:

$$T_{offspring} = \frac{Cov_{parent,\ offspring}}{Var_{parent}} \times T_{parent} + c \tag{4.9}$$

$$h^2 = Cov\left(P_{parents}, P_{offspring}\right) / Var\left(P_{parents}\right) \tag{4.10}$$

T represents the trait values for which we aim to estimate narrow-sense heritability.

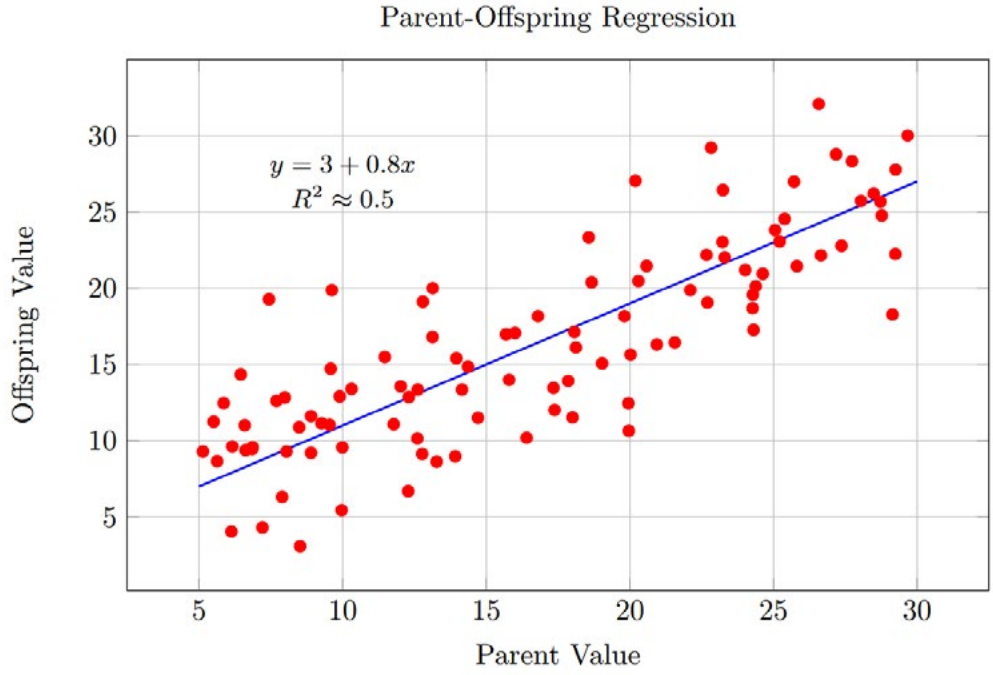


Fig. 4.3 Parent-Offspring regression based on 100 simulated observations in a hypothetical quantitative trait. Under the full-sib experimental design, the estimated narrow-sense heritability is equal to 0.8 – the linear regression slope. R^2 designates the coefficient of determination, quantifying the data points' spread.

The regression of the offspring's score on the mother's score requires calculating the variance of the mother's score and the covariance between the mother's score and the mean score of her offspring. The y-intercept is easily calculated, as described above. However, if the father's contribution to the trait is unknown, we can estimate only half of the narrow-sense heritability from the slope. Obviously, the father's genes (in forest genetics, we usually talk about paternal gametic contribution) also affected the offspring phenotypes. Males and females were paired randomly, so they may have had different scores. Therefore, the regression of offspring on a single (more commonly maternal) parent underestimates narrow-sense heritability by about 50 %.

It should be noted that while linear regression chooses the "best-fit" line through the points, there is considerable room for statistical error. The 95% confidence interval for the slope is based on the variance in the scores. An explanation for calculating this is available in any good statistics book (e.g., Welham et al., 2014). Suffice it to say that the more points you have, the narrower the 95% confidence interval will be.

Therefore, the analysis yielded a biased estimate of narrow-sense heritability, which could be relatively high (in our students' past sample, their height was usually estimated at around 0.8 narrow-sense heritability, a relatively

high estimate). However, considerable statistical error is associated with the estimate, and a narrow-sense heritability as low as 0.3 would fall within the 95% confidence interval.

4.3.4 HERITABILITY ESTIMATED USING ANOVA

Key Assumptions

The environmental variance within groups is the same across groups. There is no interaction between genetic and environmental factors.

This method allows you to estimate how much of the variation in your phenotype of interest is due to genetic differences among individuals.

Including a block effect in the analysis (often substituted by replication) accounts for potential environmental or experimental conditions that might vary across blocks or groups within your experiment. This adjustment helps isolate the genetic variance more accurately by controlling for the environmental variability between blocks.

Balanced Design: The calculation assumes a balanced design, where each block has the same number of groups, and each group has the same number of individuals. In forestry, this may be quite unrealistic. However, slight deviations from these assumptions may not be critical. The exact statistical theory is beyond the scope of this book. For relevant details, see Welham et al. (2014)

Block Effect Significance: If the block effect is insignificant, the contribution to the total variance might be negligible.

Including a block effect helps ensure that the estimate of genetic variance is not confounded by environmental or experimental variability across different blocks.

Heritability Estimate Example Based on ANOVA

Table 4.1. ANOVA using a randomized complete block design

Source of Variation	Degrees of Freedom	Mean Squares	Expected Mean Squares
Replications	$R-1$	MS_4	$\sigma_w^2 + T\sigma_{RF}^2 + TF\sigma_r^2$
Families	$F-1$	MS_3	$\sigma_w^2 + T\sigma_{RF}^2 + TR\sigma_f^2$
Families × Replications	$(F-1)(R-1)$	MS_2	$\sigma_w^2 + T\sigma_{RF}^2$
Trees within Plots	$RF(T-1)$	MS_1	σ_w^2

F, R, and T refer to the number of families, replications, and trees per family plot within each replication, whereas σ^2_w, σ^2_{RF}, σ^2_f, and σ^2_r are the within-plot, replication × family, family, and replication variance components, respectively.

An analysis of variance (ANOVA) allows the tree breeder to partition the observed variation into genetic and environmental components and to assess their interactions if the proper experimental design is employed. A simple ANOVA that is appropriate for an experiment involving open-pollinated families or families representing a polycross mating scheme (see Chapter 6 on Breeding) planted in a randomized complete-block design is shown in Table 4.1. We deliberately used this most applied experimental design, which allows for h^2 and parental breeding values estimation even in low-input breeding.

Several terms used in ANOVA warrant explanation. The term *'source of variation'* is self-explanatory; it denotes which partition of the total variation is accounted for in that table line. *'Degree of freedom'* indicates the number of independently variable classes, whereas *'mean squares'* for any source denote all the variation that has contributed to the observed differences for that effect (it is a sum of calculated and divided by the respective degrees of freedom in the given line). The term *'expected mean squares'* denotes the relative contributions of each source of variance to each mean square.

For example, in Table 4.1, the family mean square is composed of variation among trees within plots, interactions of replications and families, and family variation. The mean squares are used to test for statistical significance of

an effect, but further analysis is required to estimate the variance components for within-plot (σ_w^2), family (σ_f^2), replication x family interaction (σ_{RF}^2), and replication (σ_r^2). The variance attributable to the interaction of families with individual replications can be estimated in the following way:

$$\sigma_{RF}^2 = \frac{MS_2 - MS_1}{T} = \frac{\sigma_W^2 + T\sigma_{RF}^2 - \sigma_W^2}{T} \tag{4.11}$$

The family component of variance is estimated from the family mean squares and mean square of the interaction:

$$\sigma_f^2 = \frac{MS_3 - MS_2}{TP} = \frac{\sigma_W^2 + T\sigma_{RF}^2 + TP\sigma_f^2 - \sigma_W^2 - T\sigma_{RF}^2}{TP} \tag{4.12}$$

Individual-Tree Heritability Calculations

Variance components such as those derived from the analysis shown in Table 8.1 can be used to estimate heritability. The term individual-tree h^2 was coined to determine the parameter based on individual observation. Contrastingly, family heritability is used for estimation based on family means. Family heritabilities are usually higher than individual tree heritabilities because they are based on averaged estimates with a sample of many progenies. Consequently, the environmental effects in the test are averaged out for the family means. Family heritability is much less commonly reported. Thus, we tend to assume that most narrow-sense heritability estimates were based on individual observations.

The heritability calculation involves equating the *statistical* components of variance to their *genetic* counterparts, such as the components for additive genetic variance, nonadditive genetic variance, and phenotypic variance.

For the analysis in Table 8.1, where half-sib families are being tested, the **family component of variances equals one-fourth of the additive genetic variance**. With the testing scheme used in this example, the phenotypic variance is estimated by the sum of the within-plot variance component, the replication × family variance component, and the family component:

$$\sigma_P^2 = \sigma_w^2 + \sigma_{RF}^2 + \sigma_f^2 \tag{4.13}$$

Individual-tree narrow-sense heritability, which is the actual ratio of additive genetic variance to total phenotypic variance, is calculated as:

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} = \frac{4\sigma_f^2}{\sigma_W^2 + \sigma_{RF}^2 + \sigma_f^2} \tag{4.14}$$

Knowledge of the genetic meaning of the family component allows one to interpret the meaning of the other components. The before-mentioned interaction term results from the failure of families to behave the same way relative to each other in different replications, and the within-plot variance is composed of the remainder of the genetic variation plus environmental variation within plots. Therefore:

$$\sigma_W^2 = \frac{3}{4}\sigma_A^2 + \sigma_{NA}^2 + \sigma_E^2 \tag{4.15}$$

Where:

- σ_A^2= Additive genetic variance
- σ_{NA}^2= Nonadditive genetic variance
- σ_E^2= Environmental variance

It should be noted that the mating design used for the example in Table 4.1 **cannot be utilized to calculate broad-sense heritability (H^2)**, the ratio of the total genetic variation to the total phenotypic variance, because the half-sib design cannot partition the nonadditive genetic variation.

4.3.5 SUMMARY

The established methods for estimating heritability include familial data methods such as parent-offspring regression, full-sib and half-sib analysis, and twin studies; population data methods like ANOVA and REML; and genomic methods such as GWAS and GBLUP. Each method has its applications and strengths, making them suitable for different genetic and breeding studies.

4.4 GENERAL COMBINING ABILITY

General Combining Ability (GCA) is the value of an individual judged by the mean value of its progeny. Suppose an individual is mated to many randomly selected individuals in a population. In that case, GCA equals the deviation of the mean of its progeny from the overall mean of the entire population. High GCA usually implies the presence of genes with additive effects affecting the evaluated quantitative trait.

As with heritability, GCA can be estimated in progeny trials raised from systematic matings. When data from such trials are analyzed, we often determine that a parent independent of mating partners gives rise to well-performing progenies. This is an example of a parent with good GCA. More precisely expressed, the average deviation of this parent's progeny from the total mean of the trial is an estimate of the GCA of that parent. To illustrate this, hypothetical values are given in exercises of this subject for a hypothetical trial in which all possible matings between parents are involved except for selfing. When parents serve both as females and males, the mating design is designated as diallel. (A more detailed description of mating designs is carried out in Chapter 6 on Tree improvement in connection with tree breeding since mating designs are very important in breeding.)

Estimating GCA is one of the main objectives in forest tree breeding and enables the identification of the genetically most valuable trees. However, the general combining ability of an individual tree is a relative estimate depending on which parents are being tested and the testing environment. It will be further explained in a chapter on tree improvement that in multiple-generation tree breeding, recurrent selection for general combining ability is applied.

4.4.1 BREEDING VALUE

The genotypic value of an individual is judged by the mean value of its progeny. If an individual is crossed with many randomly selected individuals in a population, its breeding value equals the double deviation of its mean from the grand mean of this population. The breeding value is effectively the doubled GCA of the individual parent.

The breeding value is defined as 2 x GCA because only half of the genetic information comes from one of the parents, the other half coming from other parents. It is evident that the breeding value depends on the other parents tested and the environmental conditions under which the testing occurred. This means breeding value, additive variance, and heritability are relative estimates. All the parameters are relative to the concrete population and the specific environmental conditions of the test sites.

However, the more rigorous definition of breeding value is nested within the concept of **average effect.** The breeding value of a parent for a specific trait quantifies the phenotypic superiority of the parent that is passed on to its offspring by random mating with other parents in the population. Only average allele effects (additive effects) are passed on to offspring from a parent, and a parent contributes, on average, half of its breeding value to its offspring.

4.5 SPECIFIC COMBINING ABILITY

Specific Combining Ability (SCA) labels the specific pairs of parents after their respective cross yields a progeny that strongly deviates from what is expected based on the parents' general combining abilities. High SCA for the given parental combination usually indicates the presence of dominance or epistasis. Similarly, high SCA could imply the effect of heterosis.

The progenies of specific crosses may have values that deviate conspicuously from the mean values of both parents, which may be close to the total mean of the trial. Such a deviation is a Specific Combining Ability (SCA). The family and parental deviations from the grand mean are given in exercises. These latter deviations are estimates of

the parental GCAs on the assumption that the experimental error is = 0 (a major simplification!). Under the same assumption, it is possible to estimate the specific combing abilities with the aid of the following general equation:

$$y_{ij} = m + GCA_i + GCA_j + SCA_{ij} \quad (4.11)$$

In SCA, the non-additive interactive effects are recreated anew in each generation. However, the diversity of unique genotypes produced in the offspring is constrained to the alleles present in the two parents, unlike offspring from a half-sib family, where the male parent's alleles represent a larger, theoretically population-wide sampling. In full-sib families, one-quarter of all offspring will share the same genotype at any locus due to identity by descent. Suppose one allele at a locus is dominant. In that case, all trees with this genotype will share a genotypic value influenced by the dominant allele's action rather than solely the average effects of the alleles.

This phenomenon illustrates the concept of SCA, where the mean performance of a full-sib family can differ from the expectation based solely on the average breeding values of the parents. If all gene loci exhibit only additive effects (with no dominance), the SCA contribution to the family mean is zero. While general combining ability (GCA) values are intrinsic to individual parents, SCA is specific to the offspring resulting from a particular cross. Consequently, SCA values are expressed as deviations that sum to zero across all possible parental pairs in the population, making the linear approach used for deriving SCA effects appropriate.

Additionally, GCA and SCA values are uncorrelated. For example, if a female parent is crossed with 100 male parents to produce 100 full-sib families, approximately half of these families will have positive SCA values, and half will have negative SCA values. Importantly, these 100 SCA values will not correlate with the 100 GCA values of the male parents.

The estimations of both described types of combining abilities in realistic breeding programs take place using complex statistical models in which mating design and experimental design are crucial components. Various effects and their errors can be distinguished in an experimental series planted at multiple test sites containing more than one population. Such complex experiments with full-sib crosses increase our possibilities of identifying different effects compared to the situation for open-pollinated families. All experimental trials containing the same crosses are designated as an experimental series.

Estimating GCA is one of the main objectives in forest tree breeding and enables the identification of the genetically most valuable trees. Again, It should be noted that the general combining ability of a tree is a relative estimate and depends on which parents are tested and the testing environment.

4.6 GENETIC CORRELATIONS

When two different traits are measured in a forest tree population, there may be an association between the measurements. For example, suppose that both height and diameter are measured on the 100 trees. We might expect trees with superior height to be above average in diameter. This would imply a phenotypic association between the two traits, which could be quantified by the **phenotypic correlation**, r_p, which measures the strength of the association.

However, the phenotypic correlation between two polygenic traits may be due to genetic or environmental causes, and just as with phenotypic variances, there is a need to understand these underlying components that give rise to a phenotypic correlation. The main cause of a **genetic correlation** between two traits is **pleiotropy**, which means that a given gene locus influences the expression of more than one trait (Mode & Robinson, 1959). Consider two polygenic traits that are both influenced by many gene loci. If any of these gene loci are pleiotropic and affect the expression of both traits, then there is a measurable phenotypic correlation between the traits caused by the underlying genetic correlation. Similarly to heritabilities, we can define a broad-sense genetic correlation (r_G) and a narrow-sense genetic correlation (r_A). The broad-sense genetic correlation is the correlation of true clonal values for the two traits X and Y. In contrast, the narrow-sense genetic correlation is the correlation of true breeding values for the traits X and Y. These correlations measure the association of clonal values and breeding values for the two

traits in the reference population. For example, if a value of *rA* is negative, it indicates a strong tendency for a tree with an above-average breeding value for one trait to have an inferior breeding value for the second trait.

Calculating the size or sign (+/-) of the underlying genetic and environmental correlations from a phenotypic correlation is impossible. This has two major implications: The phenotypic correlation may have little utility in making genetic inferences; thus, we must estimate the genetic correlation. Randomized, replicated field experiments employing offspring from many parents are required to estimate the underlying causal genetic correlations. Genetic correlations are even more difficult to estimate precisely than heritabilities and require extremely large experiments (large sample size). There are three types of genetic correlations relevant in forest trees: (1) The **genetic correlation between two distinct traits**, such as between growth and wood specific gravity; (2) The genetic correlation between the same trait expressed at two different ages, called **age-age correlations** or juvenile-mature correlations; and (3) The genetic correlation of the same trait expressed in two distinct macroenvironments – **type B genetic correlation** as the best-known proxy of GxE.

For tree improvement, it is particularly important to disclose how other traits are affected when selecting one specific trait. It would not be surprising if the same sets of alleles affect two consecutive stages during bud burst in Norway spruce. On the other hand, it is not certain that the bud burst during spring and growth cessation during late summer are affected by the same alleles. To answer these complex questions, genetic correlations are calculated.

The genetic correlation is a correlation of two traits' breeding values. In agreement with estimates of other important genetic parameters of populations (breeding values, heritability, etc.), the genetic correlations are valid for the population and the conditions under which it is tested. In the equation for the genetic correlation, the covariance between traits *x* and *y* is one part; this covariance estimates the covariation between the two traits. The genetic correlation is frequently designated as r_A and it is equal to:

$$r_A = \frac{Cov_{xy}}{\sqrt{Var_x \times Var_y}} \tag{4.12}$$

To enable high precision in the estimates of genetic correlations, the experiments must contain progenies of numerous parents. Genetic correlations are even more difficult to estimate precisely than heritabilities and require extremely large experiments (large sample size). Knowledge of genetic correlations among traits can be important both for the natural populations of forest trees and in tree improvement. If two traits have a strong positive or negative correlation, then natural or artificial selection on the first trait affects the other trait. This is called a **correlated response** to selection or **indirect selection**. Consider the narrow-sense genetic correlation and remember from earlier in this chapter that only additive allele effects are passed on to offspring following random mating. With a strong positive genetic correlation (*rA*) between two traits, a parent with a high breeding value for the first trait tends to have a high breeding value for the second trait, and this parent produces superior offspring for both traits. In tree improvement programs, genetic correlations between two traits can be important for several reasons: (1) If two traits have a strong favorable correlation, then selection and breeding on the first trait results in genetic gain for both traits; (2) If two traits have a strong unfavorable correlation, then it is more difficult to progress in both traits simultaneously; and (3) If genetic correlations are unknown, it is possible for a selection program to produce random results.

Age-age correlations are very valuable for helping to identify the best age at which to make selections from the genetic and economic perspective. It is rather rare that tree breeders can afford to wait until rotation age to select the best trees for a tree improvement program. Let's consider volume growth, in which selection is done early, while the goal is to improve volume production at a mature age near the full rotation. These may be two different polygenic traits (growth at younger ages versus growth at older ages). If nearly the same set of gene loci influence volume growth across all ages, then the genetic correlation between the two traits is high. Suppose a major fraction of loci important at younger stages is less influential or new ones become important at older ages (when the trees dominate within a stand and begin to flower). In that case, the genetic correlation between the two ages decreases. Consequently, the success of selection at early ages in producing volume gain at rotation age depends at least

partially on the age-age correlation of the two ages. Recently, age-age correlations have been estimated using bivariate mixed models. Similarly to trait-trait genetic correlations, age-age experiments are prone to high standard errors (high demands on sample size) and require precise experimental design.

In simplified terms, the rank change or stability of genetic entries across environments is a widely recognized indicator of genotype × environment interaction (GxE). Essentially, GxE reflects inconsistencies in the relative performance of genotypes when grown in different environments. These inconsistencies may manifest as rank changes of genotypes across environments (referred to as *rank change interaction*) or as variations in performance differences without rank changes, where differences are inconsistent across environments (known as *scale effect interaction*).

An example of a scale effect interaction would be when one species consistently exhibits greater bole volume growth than another. Still, the magnitude of this difference varies depending on the soil type. For instance, the difference might be substantial in one soil type but minimal in another. Practical examples of rank change interactions have been reported in many species, such as southern pine species, which exhibited rank changes in early height growth depending on the silvicultural management intensity in which they were grown.

In GxE interaction discussions, the terms *genotype* and *environment* are used broadly. Genotypes may include different species, seed sources, provenances, families, or even clones of the same species. Similarly, environments can encompass a range of factors such as soil types, elevations, climates, fertilizer treatments, planting densities, or any combination of environmental or silvicultural conditions.

Statistically, GxE interaction has traditionally been analyzed as a two-way interaction in the analysis of variance (ANOVA), occurring when the true differences between genotypes for a given trait are not consistent across environments.

More recently, using linear mixed models, GxE has been quantified and reported by estimating site × site covariance for random genetic factors (which may range from clones to provenances). In such cases, type B genetic correlation is often estimated to explore further and report these interactions.

4.7 INBREEDING AND HETEROSIS

Various types of inbreeding in cross-pollinating organisms may cause a decrease in the vigor of the affected individuals. This is called inbreeding depression. Nils Sylvén's pioneered this type of research by establishing the oldest progeny trial with selfed Norway spruce in 1916. This trial was established before statistics were considered; thus, it has no replications. Despite these flaws, the results were persuasive with a stem volume of the selfed trees amounting to less than 50 % of the stem volume of the outbred trees. Even worse performances of selfed Douglas fir, noble fir, ponderosa pine, and Scots pine have been recorded in replicated experiments. Somewhat lower inbreeding was noted in one comparable experiment with *Pinus radiata*.

Several studies on tree species in nurseries have resulted in an inbreeding depression of about 20%. These results may suggest that the inbreeding depression is less pronounced under good conditions than under adverse conditions. Is there any genetic explanation for this dramatic inbreeding depression due to selfing?

Quantitative genetics methods revealed the relationship between the size of the inbreeding depression and the degree of inbreeding has been derived. This relationship is most commonly visualized as gradually increasing relatedness between the parents of the inbred individual. To explain the concept, the **inbreeding coefficient** must be defined and clarified. In trees with both female and male flowers (**monoecious** is the correct term in contrast to **dioecious** genera such as *Taxus*), a high degree of inbreeding can be obtained via repeated selfings. Full-sib, half-sib, and first cousin matings are other causes of inbreeding with a decreasing degree of relatedness in that order.

In quantitative genetics, the inbreeding coefficient F estimates identity by descent of alleles. The identity by descent (IBD) concept relies on the probability that two copies of the same allele at a common ancestor have been brought together in an offspring. This allele is also called autozygoes. It must be noted that it is not equal to plain homozygosity, where the same alleles at a homozygote must not originate from one common allele at an ancestor. Thus, these are not identical by descent, although they are identical by state.

It is undoubtedly true that the degree of homozygosity increases with inbreeding. The inbreeding coefficients for various types of relatedness are visualized in Eriksson et al., (2008) and Hartl & Clark (2006).

The above-mentioned probabilistic model of IBD is most commonly used to calculate inbreeding coefficients, and it dates back to Sevall Wright, who also postulated the effective population size. It is not a coincidence that these two crucial concepts of population genetics are closely related.

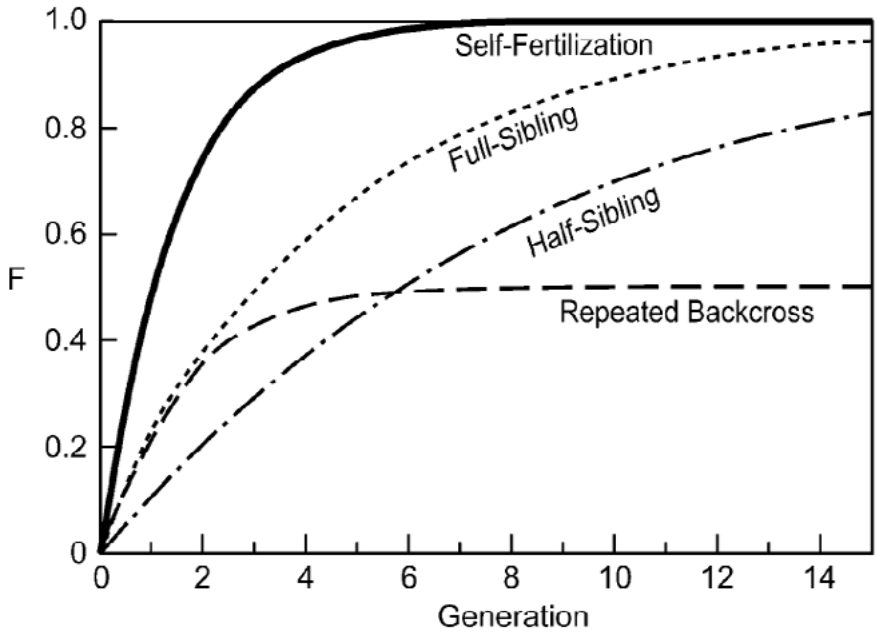


Fig. Inbreeding coefficient (F) plotted over multiple generations under variable inbreeding; modified from (Falconer & Mackay, 2009; White et al., 2007).

Heterosis is the opposite of inbreeding depression; thus, it increases vitality after mating between inbred individuals. The best-known example is hybrid maize breeding, in which several generations of inbreeding were carried out before mating occurred between individuals from different inbred lines. This method has achieved spectacular results in maize, rice, and many other important crops.

In tree species, heterosis has had a major impact on breeding for the so-called hybrid vigor. Here, **hybrid vigor** and **heterosis** are used interchangeably to describe hybrid superiority over parental species. This positive effect has been recorded in many natural and artificial hybrids of various species. In this case, the inbred lines are substituted by hybridized species and the subsequent burst of heterosis. Similarly, to breeding for heterosis, this effect will dissipate in subsequent generations due to complex introgression patterns. Hybridization may lead to heterosis or hybrid vigor due to non-additive gene action, analogically to superior crosses with high SCA. Hybrids may exhibit greater phenotypic stability due to higher levels of heterozygosity, and hybrids may allow the planted range to be extended to sites where one or both parental species are marginal because they can strive on specific ecological niches. Regardless of the mechanism, producing artificial hybrids brings alleles together that would rarely occur in pure species or natural hybrids. This sometimes creates a hybrid taxon superior to either single species involved. Efficient means of vegetative propagation greatly facilitate the deployment of hybrids. Thus, there are hybrid breeding programs in several countries for species of *Eucalyptus*, *Populus* (famously, *Populus tremola* x *Populus tremuloides* was excellent for the soft wood), and *Salix*.

Also, in this case, there is a quantitative genetic equation that describes the heterosis HF_1 expected in the first filial generation after mating between two inbred lines:

$$H_{F1} = \sum d(p_1 - p_2)^2 \qquad (4.13)$$

In this case, p_1 is the frequency of one allele at one of the inbred lines, and p_2 is the frequency of the same allele at another inbred line. Summation of the effects in loci affecting the trait takes place. The equation reveals the larger the difference in gene frequencies between lines; the larger H_{F1} will be. The largest effect is obtained when the allele frequency is 0 in one line and 1 in the other line, i.e., one line is homozygous a_1a_1 and the other homozygous a_2a_2. Also, in this case, the dominance deviation d is involved, and there will be no heterosis if d at all loci involved is 0. Another condition for heterosis is that d at most loci is positive. Mating between individuals from different inbred lines immediately restores the vitality lost by inbreeding, which is important for conservation genetics.

4.7.1 THE DERIVATION OF INBREEDING COEFFICIENT:

BOX 4.2

The calculation of inbreeding coefficients is best illustrated through a simple example involving full-sibling mating. Consider a homozygous a_1a_1 individual mating with two unrelated males, resulting in two offspring (F1). These F1 individuals are then mated, producing an individual in the F2 generation for whom we aim to calculate the inbreeding coefficient.

The probability that the a_1 allele is passed from the a_1a_1 female to its daughter is = ½, following the fundamental principle that offspring inherit 50% of their genes from each parent. Similarly, the probability that the a1 allele is transmitted from F1 to F2 is also ½. The probability of the a1 allele being passed from the original a1a1 female to the F2 individual along this path is therefore ½ × ½ = ¼. On the analogous path (right side), the probability is also ¼. To determine the likelihood of obtaining a homozygous F2 individual, the probabilities for both paths (¼ × ¼) must be multiplied, yielding 1/16.

An alternative approach, commonly used in more complex pedigrees, starts at the F2 individual and traces paths back to the source. For each step along the pathway, the probability is multiplied by ½. This method can simplify calculations for intricate lineages.

The other homozygous possibility, where the F2 individual inherits both alleles as a_1a_1 has the same probability as the original a_1a_1 homozygote. To calculate the overall probability of identity by descent, the probabilities for the two pathways are summed, resulting in a total of 1/8. This value represents the inbreeding coefficient for half-sibling mating. It is important to note that the unrelated males at the top of the pedigree do not contribute to identity by descent. Using similar reasoning, the inbreeding coefficient for full-sibling mating can be shown to be ¼.

4.8 INTENSITY OF SELECTION

To explain the importance of the selection intensity parameter, we must return to the already defined breeder's equation. In this section, we shall discuss the effects of different strengths of artificial selection. The meaning of selection differential (S) was illustrated (Fig. 4.1). As may be seen from this figure, the selection differential depends on the trait distribution. If the same proportion of individuals is selected, the selection differential is larger if the distribution is wider. The **selection intensity (i)** has been introduced to compare different selection scenarios. The selection intensity is obtained by dividing the selection differential by the phenotypic standard deviation (square root of the phenotypic variance for the trait measured).

$$i = \frac{S}{\sigma_{ph}} \qquad (4.14)$$

The selection intensity is non-linearly related to the proportion selected. Effectively, this non-linearity means that increasing the selection intensity from 2 to 3 requires a much larger population than raising it from 1 to 2. To make it possible to calculate the result of a certain selection for a particular trait, it is necessary to know the genetic variation in this trait. If we aim to select several individuals as parents for a new generation, the additive genetic effects are of importance. Thus, the narrow-sense heritability is of interest, and the improvement is equal to the narrow-sense heritability multiplied by the selection differential. The result of this product is usually referred to as genetic gain, ΔG. As an equation, we get:

$$\Delta G = h^2 \times S \tag{4.15}$$

Or if we substitute the selection differential by selection intensity:

$$\Delta G = h^2 \times i \times \sigma_{ph} \tag{4.16}$$

To visualize the crucial impact of the additive variance for breeding progress by artificial or natural selection, the narrow-sense heritability can be expressed as the ratio $\sigma_a{}^2 / \sigma_{ph}{}^2$.

If this ratio is included in the previous equation, we get the genetic gain calculated in the following way:

$$\Delta G = \frac{i \times \sigma_a^2}{\sigma_{ph}} \tag{4.17}$$

In which i express the selection intensity for the selected trees, $\sigma_a{}^2$ is the additive variance, and σ_{ph} is the phenotypic standard deviation in the entire population (despite the complex term, it is solely a measure of the total variation of the population in the given trait). The effects of the selection confirm what we said before: the larger the selection differential, the larger the gain by selection. In many cases, it is useful to predict gain from selection before making the selections. For example, this may be important for planning an effort in the selection process or evaluating different selection strategies. Before realizing the selections, neither the S nor i are known (since nothing has been measured yet), but we can still predict expected gains when four conditions are met (Falconer & Mackay, 2009): (1) The phenotypic values in the base population follow a normal distribution; (2) Truncation selection is employed - only phenotypes above a certain truncation value, will be selected; (3) There is a known proportion of selected trees; and (4) there are available estimates of the heritability and phenotypic standard deviation for the base population.

Implications for phenotypic mass selection

Selection intensity is a crucial factor that breeders can control to enhance genetic gain—the smaller the fraction of the population selected, the higher the selection intensity and the greater the expected gain. However, there's a point of diminishing returns as the effort and cost to measure more candidates increase. For instance, selecting one tree from 2.62, 20, and 10,000 candidates yields selection intensity (i) values of approximately 1, 2, and 4, respectively. If a breeding program aims to select 100 trees, it would need to measure 262, 2,000, and 1,000,000 trees to achieve these intensities. Doubling the selection intensity from 1 to 2 requires measuring about 7.6 times more trees (2,000/262) but doubling it again from 2 to 4 demands 500 times more trees (1,000,000/2,000), illustrating significant diminishing returns and escalating costs.

Furthermore, when selecting multiple uncorrelated traits, the selection intensity for each trait decreases because the probability of a candidate excelling in all traits is the product of the individual probabilities. For example, selecting the best tree out of 100 candidates for a single trait has an intensity of 2.66, but for two traits, the intensity drops to 1.76 per trait since the tree must be in the top 10 % for each trait ($0.10 \times 0.10 = 0.01$ probability). This means the predicted genetic gain for each trait is reduced by about 50% when selecting for two traits instead of one ($2.66/1.76 \approx 1.51$). With six uncorrelated traits, the intensity further decreases to 0.87 per trait because the selected tree must be in the top 46% for each trait ($0.466 \approx 0.01$). Consequently, genetic gain per trait is approximately three times higher when selecting a single trait compared to six traits ($2.66/0.87 \approx 3.05$). This demonstrates that to maximize genetic gains effectively and economically, breeders should focus on a limited number of high-priority traits—typically no more than three to five—while considering the trade-offs between selection intensity, genetic gain, and the costs associated with measuring additional candidates and traits (White et al., 2007).

4.9 REFERENCES

Eriksson, G., Ekberg, I., & Clapham, D. (2008). *An Introduction to Forest Genetics*. https://res.slu.se/id/publ/22529

Falconer, D. S., & Mackay, T. (2009). *Introduction to quantitative genetics* (4. ed., [16. print.]). Pearson, Prentice Hall.

Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246. https://doi.org/10.2307/2841583

Hartl, D. L., & Clark, A. G. (2006). *Principles of Population Genetics* (4th edition). Sinauer Associates is an imprint of Oxford University Press.

Mode, C. J., & Robinson, H. F. (1959). Pleiotropism and the Genetic Variance and Covariance. *Biometrics*, *15*(4), 518. https://doi.org/10.2307/2527650

Mrode, R. A. (with Pocrnic, I.). (2023). *Linear Models for the Prediction of the Genetic Merit of Animals* (4th ed). CAB International.

Welham, S. J., Gezan, S. A., Clark, S. J., & Mead, A. (2014). *Statistical methods in biology: Design and analysis of experiments and regression*. CRC press. https://books.google.com/books?hl=cs&lr=&id=UpVBBAAAQBAJ&oi=fnd&pg=PP1&dq=STATISTICAL+METHODS+IN+BIOLOGY+Design+and+Analysis+of+Experiments+and+Regression&ots=CAgF9B6Irx&sig=LCkbWZzJUZDB_srngC7W7kEzmRE

White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI.

5. GEOGRAPHIC VARIATION

5.1 THE CHALLENGES AND HISTORY OF PLANT DOMESTICATION

The Neolithic Revolution emerged independently in approximately seven different regions from around 12,000 BCE to about 4,000 BCE. The levels and methods of domestication and how humans influenced domesticates have changed over the ages. Initially, humans practiced unconscious selection for useful traits. Around the turn of the era, people began purposefully cultivating plants with desired traits. Intentional breeding to improve the traits of domesticated organisms appeared in the 18th century. A systematic scientific approach to breeding, supported by genetic knowledge, developed in the 20th century, and in the 21st century, we can directly read and modify the genomes of organisms.

Finding archaeological evidence for plants, especially vegetables, is somewhat more challenging. Nikolai Vavilov identified the centers of crop domestication in the 1920s as areas with the greatest diversity of wild relatives of the given crop.

In the early stages of domestication, issues such as the shattering of mature seeds—a typical trait of wild progenitors—surely arose. Early farmers may have addressed this problem by harvesting immature seeds—a gradual selection of variants that retained seeds in the ear or pod until harvest likely occurred over time.

Typically, a plant can be successfully domesticated when a desirable trait is achievable by altering a few genes. For example, the wild ancestor of almonds contained bitter substances to deter herbivores, but changing a single gene turned off the synthesis of this bitter substance. In contrast, tannins in oaks are controlled by many genes, likely preventing their domestication. Other typical changes from wild to domesticated forms include reduced growth height, increased seed size, and loss of thorns.

About 10,000 years ago, the following crops were domesticated in the Fertile Crescent: barley, wheat, lentils, peas, chickpeas, olives, dates, grapes, and flax—the use of wild flax dates back to 35,000 years ago. In the Far East, millet, rice, soybeans, mangoes, citrus fruits, taro, and bananas were domesticated. The interesting story of the domestication of lagenaria, a shrub providing woody gourd-like fruits that can be hollowed out to serve as containers—jugs or bowls—occurred in Asia about 12,000 years ago, with its subsequent transport to the American continent 10,000 years ago.

About 5,000 years ago, sorghum, yams, hibiscus, cotton, and coffee were domesticated in Africa. In Mesoamerica, maize, beans, peanuts, chili peppers, tobacco, cocoa, pineapples, and pumpkins were domesticated. In South America, potatoes, quinoa, and amaranth were domesticated.

5.1.1 BEGINNINGS OF BREEDING

Following the initial domestication event, crops' cultivation underwent gradual and unsystematic improvement—early signs of what would later become formal breeding. Ancient farmers might have chosen to preserve seeds from the healthiest or highest-yielding plants or to breed the best individuals from their herds while culling those that posed the most problems.

Selection also occurred indirectly; agricultural groups with the best crops had higher survival rates and thus spread their superior crops more effectively. However, significant and rapid progress was hindered by a lack of understanding of heredity principles and genetic variation.

One of the earliest recorded examples of plant breeding emerged in the early 17th century. European settlers in the New World discovered that their traditional varieties were not well-suited to their new environment. They had to cultivate crops from the native populations and selectively propagate the ones that yielded some produce, thus adapting imported crops through simple breeding. This period also saw the Columbian Exchange, where crops were transferred between the Americas, Europe, and Asia.

5.1.2 THE BRITISH AGRICULTURAL REVOLUTION

The British Agricultural Revolution marked another milestone in agricultural development. Innovations and mechanization in farming led to a significant increase in production. This era saw the first systematic attempts at breeding, notably led by Sir Robert Bakewell. He kept detailed records of animal traits to enable objective selection, eventually leading to the establishment of pedigree books. Bakewell also introduced progeny testing, evaluating the performance of the first group of offspring to select the best sires for future generations.

One of the earliest programs in plant breeding aimed to increase the sugar content in fodder beets. Between 1786 and 1830, sugar production doubled thanks to the efforts of Franz Karl Achard. The 19th century saw the rise of hybridization, where hybrid offspring often exhibited superior traits compared to either parent. This practice inspired Mendel's experiments with peas.

Breeding using statistical evaluation spread thanks to the biometric school founded by Sir Francis Galton. Around 1900, several commercial companies emerged, conducting breeding programs and selling improved varieties.

5.1.3 THE GREEN REVOLUTION AND AGRICULTURE 4.0

The Green Revolution between 1950 and 1970 marked a significant technological and varietal transfer to developing countries, resulting in an approximate 50% increase in crop yields. This era introduced high-yielding varieties of staple crops, advanced irrigation techniques, synthetic fertilizers, and pesticides, significantly boosting agricultural productivity and food security in many regions.

Current Progress in Genomics and Precision Agriculture

In recent years, the impact of genomics on breeding has become increasingly evident, affecting crops, livestock, and even forest tree populations. Genomics allows for precisely identifying and manipulating genes responsible for desirable traits such as disease resistance, drought tolerance, and improved nutritional content.

Agriculture 4.0

Integrating advanced technology into agriculture—Agriculture 4.0—is revolutionizing farming practices. The new progressive methods include:

Autonomous Drones: These are used to monitor crop health, apply pesticides or fertilizers, and even plant seeds. Drones equipped with multispectral or hyperspectral sensors can provide detailed insights into plant health, helping farmers and tree breeders make informed decisions.

Sensors and Satellite Technology: Sensors placed on drones or satellites collect real-time data on soil moisture, nutrient levels, and crop growth. This data can create precise maps of field conditions, enabling targeted interventions.

Machine Learning Models: Machine learning algorithms analyze vast amounts of data from various sources to predict crop yields, optimize irrigation schedules, and detect early signs of pests or diseases. These models help in making precise, data-driven decisions.

Environmental Sustainability: These technologies can lead to more environmentally friendly practices. For example, the precise application of inputs like pesticides and herbicides reduces their overall use, minimizing environmental impact. Mechanical systems driven by drones can sometimes remove weeds individually, potentially eliminating the need for chemical herbicides.

Robotics and Automation: Advanced robotics are being developed to perform tasks such as harvesting, weeding, and planting with high precision and efficiency. These technologies reduce the need for manual labor and increase operational efficiency.

5.1.4 FUTURE PROSPECTS

The future of agriculture lies in further integrating these technologies to create a highly efficient, sustainable, and productive agricultural and forestry system. Research is ongoing to improve genomic editing techniques, enhance sensor accuracy, and develop more sophisticated machine-learning models. The goal is to balance maximizing yields and maintaining ecological health plus genetic diversity, ensuring food security and sustainable wood production for future generations while protecting the environment.

In summary, the advancements stemming from the Green Revolution are amplified by the innovations of Agriculture 4.0, paving the way for a new era of precision, efficiency, and sustainability in agriculture.

5.2 GEOGRAPHIC VARIATION IN FOREST TREES

Geographic variation in forest trees is shaped by a combination of climatic, soil, genetic, and ecological factors that influence the characteristics and distribution of tree species across different regions. Trees adapt to their local environments, with differences in temperature, precipitation, and seasonality leading to distinct adaptations. For example, trees in colder climates, such as boreal forests, often have needle-like leaves to minimize water loss. In contrast, those in tropical forests are typically broadleaf species adapted to high rainfall. Elevation and soil composition also play significant roles, as trees at different altitudes or on varying soil types develop unique traits to thrive in those conditions.

Stydying geographic variation within the native range is a logical first step in genetics research and potential domestication of any tree species because understanding the magnitude and patterns of this variation is important for: (1) Learning about the evolutionary forces at play that caused the observed patterns; (2) Making reforestation decisions about safe seed transfer from the site of collection to a remote planting site that still ensures adaptation; (3) Delineating breeding and deployment zones (**seed zones** – see chapter on tree improvement) for operational tree improvement programs; (4) Deciding which provenances will give the highest yields in a target planting region; (5) Designing selection and genetic testing programs that span the appropriate edaphoclimatic regions; and (6) Optimizing gene conservation strategies that capture the natural genetic diversity existing within species.

Genetic variation is critical, as trees from different geographical areas often display differences in growth rates, disease resistance, and reproductive timing due to limited gene flow and local adaptations. Studies, such as **provenance trials**, where seeds from different regions are grown in a common environment, help reveal these adaptations' genetic basis. Biotic interactions, including pests, pathogens, and mutualistic species like mycorrhizal fungi, also vary geographically, further influencing tree health and distribution.

Historical and evolutionary factors, such as past glaciation events and human activities like deforestation, have shaped the current tree species distribution patterns and genetic diversity. Additionally, latitudinal gradients play a role, with tropical regions hosting more diverse and larger tree species than higher latitudes, where trees might grow more slowly but live longer.

Understanding geographic variation in forest trees is crucial for conservation efforts and forest management, particularly in climate change. As environmental conditions shift, trees highly specialized to their current environments may struggle, leading to changes in forest composition and ecosystem services. This makes the study of geographic variation vital for predicting and mitigating the impacts of climate change on forest ecosystems.

Phenotypic plasticity in forest trees is the ability of a species to alter its physical traits in response to varying environmental conditions, such as changes in light, water availability, and temperature. This adaptive flexibility allows trees to survive across diverse environments by adjusting their growth patterns, leaf morphology, and reproductive strategies. Plasticity is crucial in enabling trees to cope with stressors like drought and herbivory, contributes to geographic variation, and is vital for forest resilience in climate change.

5.2.1 THE IMPACT OF GLACIATION EVENTS ON TREE SPECIES

To show one example, the impact of glaciation events on the distribution of tree species in North America and Europe differed significantly due to variations in geography, climate, and the extent of ice coverage during the last Ice Age.

The Laurentide Ice Sheet covered much of Canada and the northern United States during the last glacial period in North America. As the ice advanced, tree species were forced to migrate southward to refugia, or areas that remained ice-free and suitable for survival. The geography of North America, with its large north-south axis and broad coastal plains, allowed many species to migrate southward without encountering major physical barriers. This migration resulted in a relatively rich species diversity that could recolonize northern areas as the glaciers retreated.

However, this migration also caused significant bottlenecks, where populations were reduced to small sizes in limited areas, leading to genetic drift and reduced genetic diversity in some species. After the glaciers retreated, many species expanded northward again, recolonizing the landscape but often with less genetic variation than before the glaciation.

In contrast, the impact of glaciation in Europe was more severe due to the continent's geography. During glacial periods, the Alps, Pyrenees, and the Mediterranean Sea formed significant barriers to southward migration. As a result, many tree species could not find suitable refugia and faced local extinctions. The ice sheets and harsh conditions of the tundra environment limited the survival and movement of species, leading to a more significant loss of diversity than North America.

European species that did survive often did so in small, isolated pockets of refugia, such as in the Iberian Peninsula, the Balkans, and parts of Italy. This isolation led to even greater reductions in genetic diversity and a slower, more limited recolonization of northern Europe once the glaciers retreated. As a result, the current flora of Europe is less diverse than North America's, with fewer tree species and a more homogenous genetic structure within species.

The main contrast between North America and Europe lies in tree species' post-glacial biodiversity and genetic diversity. North America's broader, uninterrupted north-south axis allowed for more successful migration and re-colonization, preserving a greater variety of species. In Europe, geographic barriers and more severe glaciation led to greater extinctions and reduced genetic diversity, resulting in a less diverse and more homogenous post-glacial tree population. However, in North America, there are notable examples of species that underwent a series of bottle-necks, including western red cedar (*Thuja plicata*).

5.2.2 MAIN TYPES OF GEOGRAPHIC VARIATION IN FOREST TREES

Geographic variation in trees can be categorized into several main types, each reflecting how tree species differ across different regions due to environmental, genetic, and evolutionary factors. The main types of geographic variation in trees include:

Clinal Variation

Gradual changes happen in specific traits across geographic gradients, often associated with environmental gradients like temperature, precipitation, or elevation. Alternatively, we speak about the continuous change of population means along an ecological gradient attributed to changes in allele frequency. For example, tree height might decrease with increasing altitude or latitude as environmental conditions become harsher. This definition of cline dates back to the Langlet study in 1959 (Langlet, 1958).

The most plausible genetic explanation of clinal variation is that differential natural selection acts along the continuous environmental gradient and changes allele frequencies continuously at loci, controlling traits associated with adaptation and fitness. Therefore, in the *Tsuga heterophylla* example, allele frequencies at loci influencing growth cessation and cold hardiness presumably vary continuously from north to south along the Pacific Coast at the loci influencing these two traits of interest. If the genetic variation is clinal, the distinction of geographic races is rather arbitrary. The most northerly and southerly provenances of *T. heterophylla* are distinct races, but such a definition is

more blurry for two provenances separated by only a degree of latitude. Still, clinal variation is an important subset of geographical variation, explaining the underlying evolutionary causes of variation.

It can be shown in growth rhythm, e.g., budburst during spring and onset of dormancy at the end of the vegetation season; there is a continuous variation in Norway spruce, Scots pine, and many other prominent tree species. Such a variation is described as **clinal,** and instead of ecotypes, we may observe **ecoclines** in Norway spruce and Scots pine across main environmental gradients. For most well-studied tree species of the northern hemisphere, the photoperiod was determined as the primary trigger for the onset of growth cessation. If the ambient conditions are adverse, e.g., due to drought, this might also induce growth cessation. The regulation of dormancy by the night length causes a population transferred north to enjoy a longer growth period than at its original site. Similarly, transfers southwards reduce the growth period since the critical night length for growth cessation occurs earlier than at the original site.

In Douglas fir, Interior Rocky Mountain populations (variety *glauca*) exhibit elevational and geographical clines for early frost tolerance strongly associated with the length of the frost-free periods, e.g., with growing elevation, the frost-free period declined. An early budburst and a late budset occurred in populations from high altitudes.

When introducing exotic species, knowledge about clinal variation is often the main decision driver. Norway spruce and Scots pine show pronounced clines, so we expect that introduced species originating from climatic conditions similar to the Scandinavian (or European in the broad sense) also show clinal variation. Suppose the knowledge about provenance differences in a domestic species is as good as for Norway spruce and Scots pine in Scandinavia. In that case, it is easy to select the provenances of a domestic species for species trials. In such situations, one or two provenances might be sufficient. A larger number of provenances of the exotic species that should be tested should be selected. They should be selected from areas with climate and edaphic conditions similar to those in the test area.

Ecotypic Variation

Ecotypic variation refers to distinct forms or populations within a species that are adapted to specific local environmental conditions. These locally adapted forms, known as "ecotypes," arise due to differences in factors like climate, soil, and altitude. For example, coastal trees may develop adaptations for salt tolerance, while the same species inland might not, resulting in different ecotypes.

The term ecotype may be used to describe a specific race made up of genotypes adapted to a particular habitat or ecological niche. Unlike the term cline, which refers to a single trait, ecotype refers to many traits or characteristics of the race that distinguish that particular ecotype. As with clines, natural selection is implied as the cause of genetic differences between ecotypes. Ecotypes are most likely to develop when environments change abruptly and populations are more isolated from each other (i.e., gene migration among populations is limited).

A well-known example of ecotypic variation is seen in Norway spruce (*Picea abies*). These ecotypes are often based on elevation, where trees at different altitudes develop distinct traits to adapt to varying temperatures, moisture levels, and growing seasons. In some cases, these ecotypes can occur abruptly in a fragmented landscape or due to the inversion of vegetation types, where species normally found at higher elevations might be found at lower altitudes due to local microclimates.

The locally adapted forms, known as "**ecotypes**," arise due to differences in factors like climate, soil, and altitude. In contrast to forms, which are generally ephemeric, true ecotypic features are inherited in the offspring. For example, a narrow crown in Norway spruce is passed on to the next generation.

Examples of ecotypic variation can be found in many species. In many cases, this type of variation has been observed by practitioners, but scientific studies may be scarce in recent times. These ecotypes are often based on elevation, as tree populations at different altitudes develop distinct local adaptations to varying temperatures, moisture levels, and length of growing season. Distinct soils underline the specific type of ecotypic variation. In this case, the term **edaphotype** has been coined. One of the examples could be maritime pine in Spain.

Göte Turesson was probably the first scientist to report adaptation to different site conditions more than 100 years ago in the early 1920s. He coined the term ecotype for this specific type of adaptation. Turesson mainly studied

perennials growing under distinct site conditions, such as rocky sites contrasted with beach meadows and dunes. After cultivation at different site conditions, the "rocky" and the "meadow" ecotypes retained their morphology, proving that their characteristics were genetically conditioned due to the local adaptations.

Some advocates of the ecotype concept have claimed that what we observe as continuous variation is a stepwise variation that should be designated as ecotypic. The prerequisite for detecting stepwise variation is that there is no gene flow among populations growing under different site conditions. The pollination and seed dispersal pattern may decide whether there will be an ecotypic or an ecoclinal variation along the specific environmental gradient. It is highly unlikely to find ecotypes in wind-pollinated species with broad and continuous distribution. Suppose plants are exposed to extreme stress, like the grass species growing on mining wastes, that can lead to significant ecotypic variation, as was shown earlier. In summary, the lengthy and animated controversy among scientists whether forest trees show continuous variation is best resolved by trying to identify the evolutionary factors of significance for each species separately.

On a finer geographical scale, **common garden experiments** (a more general term used for provenance trials) show that the strong effects of differential natural selection are sometimes manifested by steep, local clines (i.e., rapidly changing performance of different populations with environmental gradients across small distances) or ecotypes that have arisen across short distances. Two examples of steep, local clines were given by White et al. (2007): (1) An elevational cline in *Abies balsamea* showing strong genetic differentiation for CO_2 assimilation rate among seed collection sites separated by a total distance of 3.2 km; and (2) An elevational cline in *Pseudotsuga menziesii* showing strong differentiation for several traits among provenances located in a 10 x 24 km area. A striking example of local ecotypes was also observed in the *Pseudotsuga menziesii* study, where populations from north and south slopes separated by as little as 1.6 km were found to differ genetically, reflecting the considerable gene migration that must be occurring among seed collection sites through both pollen and seed transfer. Therefore, differential selection can overcome the counter-balancing effects of migration when environmental differences are particularly strong.

If environmental gradients exist, then differential selection pressures cause populations in distinct localities (e.g., higher and lower elevations) to become differentiated from each other as they adapt to local environments. If environmental gradients vary continuously, clinal patterns of genetic variation are most commonly observed. Steep ecological gradients (e.g., large elevational differences) tend to produce steep clines (meaning large genetic differences between provenances). If environmental differences are discrete (such as dramatically different soil types), then ecotypes are more likely to form. **Clines are more common than ecotypes** because most environmental characteristics vary continuously over a region.

Recent studies utilized SNP markers to investigate the underlying genes in putative ecotypes. Korecký (2021) studied Norway spruce elevation ecotypes, and the SNP markers yielded clear separation by the genetic distance. The data analysis revealed a partition into two groups—low- and medium-elevation ecotypes being the one, and the high-elevation forming the second group. Despite the absence of any apparent biological (reproductive phenological asynchrony) or geographical (mountain ridges, long-distance separation) barriers, they assumed the observed distinctive genetic structure of the high-elevation ecotype is a consequence of natural selection.

Discontinuous (Allopatric) Variation

Sharp differences may occur in traits between populations that are geographically isolated. Two populations of the same species separated by a mountain range or a large body of water might develop distinct characteristics due to their isolation.

Phenotypic Plasticity

As described above, variation within the same genetic population is due to trees' ability to change their physical traits in response to environmental conditions. For example, a single tree species may exhibit different leaf shapes or growth forms depending on whether it grows in a shaded forest or an open, sunny area. **Phenotypic plasticity** is also related to the amplitude of a trait of a genotype studied in at least two different environments. The term **reaction norm** is closely related and is used to describe the trait value change of a genotype studied along an environmental gradient.

Plants have a great capability to change their phenotypes depending on the outside conditions. A genotype tested in two or more environments may have different heights, stem forms, wood densities, etc., in the specific environments. The amplitude of such a variation in a trait is a measure of the genotype's phenotypic plasticity. Many genetic textbooks do not regard phenotypic plasticity. Its role in evolution is somewhat ambiguous: On the one hand, phenotypic plasticity can be viewed as a disguise of the genotype, which means that natural selection will not be as efficient as it would be without this disguise.

On the other hand, phenotypic plasticity may contribute to the fitness of a genotype, especially in trees as long-lived organisms with wide distribution ranges encompassing many different site conditions. In this case, natural selection might increase the frequency of genotypes with a large phenotypic plasticity. However, with unpredictable events related to climate change, phenotypic plasticity may not be sufficient to cope with these adverse effects.

Historical or Relictual Variation

Variation may also arise due to historical events such as glaciation or migration, leading to species' survival in isolated refugia with distinct traits.

Example: Certain European tree species that survived glaciation in isolated southern refugia might exhibit distinct genetic characteristics compared to their counterparts in other regions.

Hybrid Zones

Areas where distinct populations or species meet and interbreed often result in a mix of traits from both parent populations. Where the ranges of two oak species overlap, hybrid trees may display traits from both parent species. Similarly, natural hybrid species arose along major hybridization zones in several fir and spruce species.

These main types of geographic variation help explain the diversity of tree forms and functions across different environments and contribute to the adaptation and survival of tree species in a wide range of ecological settings.

5.3 HISTORY OF THE PROVENANCE RESEARCH

Provenance research in forest trees refers to studying the geographic origin and genetic makeup of trees and their populations. This research is crucial for understanding tree species' adaptability, growth, and survival in various environments, particularly in the face of climate change and ecological restoration efforts. The history of provenance research in forest trees reveals its evolution from basic botanical observations to a sophisticated science that informs global forestry practices and conservation strategies.

Early observations and the beginnings of provenance research

The origins of provenance research in forest trees can be traced back to the early botanical explorations of the 18th and 19th centuries. European explorers and botanists began documenting tree species distribution across different continents during this period. They noticed that trees from various geographic regions exhibited variations in growth patterns, form, and hardiness. These early observations laid the groundwork for the concept of provenance, although it was not yet systematically studied.

One of the early pioneers in this field was Carl Linnaeus, whose classification system in the 18th century provided a framework for identifying and categorizing plants, including trees. Linnaeus and his contemporaries recognized that the geographic origin of a plant species could influence its characteristics, an idea that would later become central to provenance research.

During the height of Britain's naval supremacy, overseas colonies drove the demand for shipbuilding—both commercial and combat vessels. Constructing a single ship required approximately 4,000 oak trees, including highly suitable mast trees, and warships had an average lifespan of 12 years. By the 1790s, Britain maintained a fleet of around 300 warships, consuming an estimated 1,200,000 oaks every 12 years. This dependency on oak forests underscored wood as the military's most vital raw material, prompting diplomatic and military efforts in the Baltic to secure timber supplies. Propaganda of the time, such as an illustration featuring Britannia nurturing an oak seedling, emphasized reforestation to ensure Britain's enduring maritime power.

Patrick Mathew (1790 – 1864)
In 1831 he published *On Naval Timber and Arboriculture*, where he described the principles of producing timber suitable for shipbuilding for the Royal Navy.
He suggests removing the worst trees at the same time, which he says would lead to the gradual improvement of the next generations of trees. Matthew's chief argument is given in the Appendix to his book, which attracted no attention, although it is also mentioned in the main text. Matthew claimed, and Darwin agreed, that he had anticipated the theory of evolution by natural selection, although he had not developed his ideas and, indeed, wrote many years later that there was evidence of design and benevolence in nature, and, in particular, that beauty could not result from natural selection. Consequently, he is not given much credit today for his ideas.

Milestones in genecology

Year(s)	Event
1745–1755	Duhamel du Monceau: inventory provenance trial with pine
1759	Swedish Admiralty: advises transfer from north of oak and pine seeds
1787	von Wangenheim: introduction of North American trees to Germany
1823	de Vilmorin: begins his pine provenance trials
1827	Gloger: discusses climatic and other varieties in birds
1831	de Vilmorin: points out the principles of comparative cultivation
1833	Gloger: proposes the rule of gradual geographic diversity
1847	Bergmann: proposes his rule
1862	Schübeler: studies on crop plants transferred to colder climates
1864–1876	Jordan: studies "espèces d'affines" or jordanons
1865	von Nägeli: suggests chemical-physical varieties
1867–1869	Linssner: discusses plant diversity and sums of temperature above 0°C
1872–1878	de Candolle: discovers physiological races in herbaceous species
1875	Naudin and Radlkofer: reciprocal transfer of herbaceous species
1875	Kerner von Marilaun: experiences differences between plants transferred from northern and southern native habitats
1879	Kienitz: correlation between development and temperature of a number of diverse origins of some tree species at their native habitats
1884	Hoffmann: difference between high and low altitude origins
1887	Allen: proposes his rule
1887–1907	Cieslar: states existence of physiological varieties within acknowledged taxonomic species
1904–1907	Schott: reviews older literature and tries European pine provenances
1905–1913	Engler: shows the importance of physiological conditions in pine and spruce
1907	First International Provenance Trial with Scots pine
1910–1912	Brand and Waldron: study a number of strains of alfalfa
1910	Semenov-Tian-Shansky: proposes the concepts of natio and morpha
1913	Engler: demonstrates the inheritance of differences between spruce from high and low altitude
1918	Sumner: studies continuous versus discontinuous variation
1922	Turesson: proposes the term ecotype for a hereditarily adapted ecological race
1923	Turesson: proposes the name genecology for race ecology
1923	Schotte: shows that differences between pine provenances are correlated with mean annual temperature at their native habitats
1925–1930	Turesson: demonstrates climatic ecotypes in herbaceous species
1934–1936	Langlet: studies development and physiology of pine plants and correlates them with latitude and length of growing season; describes the pattern of diversity as depending on environmental factors
1938	Huxley: proposes the term cline for graded diversity
1939	Gilmour and Gregor: introduce the deme terminology
1942	Huxley: adopts the term deme for the local population

Fig. 5.1 Main milestones related to provenance research (Langlet, 1958).

The rise of scientific forestry in the 19th century

The 19th century marked the rise of scientific forestry, particularly in Europe, where the need for sustainable timber production led to the establishment of systematic forest management practices. This period saw the first formal provenance trials, where foresters collected seeds from different geographic locations and grew them in controlled conditions to compare their growth and adaptability.

One of the earliest and most influential provenance trials was conducted by the Prussian forester Julius Adolph Stöckhardt in the mid-19th century. Stöckhardt's work in Germany involved testing different provenances of Norway spruce (*Picea abies*) to determine which regions produced the best growth and timber quality. His findings highlighted the significant impact of geographic origin on tree performance and laid the foundation for future research.

Expansion of Provenance Trials in the 20th Century

The early 20th century witnessed an expansion of provenance research as forestry practices became more scientific and global in scope. This period saw the establishment of large-scale provenance trials, often involving international collaboration. Researchers sought to identify the best tree provenances for reforestation and afforestation projects, especially in degraded or deforested landscapes.

One of the most significant efforts during this time was the International Union of Forest Research Organizations (IUFRO) provenance trials, initiated in the 1920s and 1930s. These trials involved collecting and testing seeds from multiple provenances of key commercial species such as Scots pine (*Pinus sylvestris*) and Douglas-fir (*Pseudotsuga menziesii*). The results of these trials provided invaluable data on the adaptability of different tree populations to various climates and soils, guiding reforestation efforts worldwide.

Provenance Research and Climate Change

The latter half of the 20th century and the early 21st century brought new challenges and opportunities for provenance research, particularly in climate change. As global temperatures rise and weather patterns become more unpredictable, the adaptability of tree populations to changing conditions has become a critical concern for foresters and conservationists.

Provenance research has played a crucial role in identifying tree populations resilient to climate change. Studies have shown that trees from certain provenances may be better suited to future climate conditions than others, leading to "assisted migration" or "climate-smart forestry." This approach involves selecting and planting tree provenances that are expected to thrive under future climate scenarios, ensuring the sustainability of forests in a changing world.

Genetic Advances and Molecular Provenance Research

In recent decades, advances in genetics and molecular biology have revolutionized provenance research. Traditional provenance trials, which relied on observable traits such as growth rate and form, have been supplemented by genetic studies that provide deeper insights into the diversity and adaptability of tree populations.

Molecular markers and DNA analysis have enabled researchers to trace the genetic origins of trees with greater precision, identifying distinct populations and even individual trees that may have unique adaptive traits. This genetic approach has expanded the scope of provenance research, allowing for more targeted conservation strategies and improving the success of reforestation and afforestation projects.

Provenance Trials in the era of landscape Genomics

Provenance trials have long been a foundational tool in forestry and ecological research, used to assess the performance and adaptability of tree populations from different geographic regions. With the advent of advanced genomic technologies, these trials have been integrated into landscape genomic studies to provide deeper insights into the genetic basis of adaptation to environmental gradients and climate change. This section explores how provenance trials are utilized in these emerging fields, enhancing our understanding of genetic diversity and informing conservation and management strategies.

Provenance Trials: A Foundation for Genomic Studies

Provenance trials involve planting tree populations from different geographic regions in a common environment to assess their growth, survival, and adaptability. These trials generate valuable data on how different provenances respond to the same environmental conditions, providing insights into the influence of local adaptation and genetic variation on tree performance.

In landscape genomics, provenance trials offer a unique opportunity to link phenotypic data (observable traits) with genetic data across different environments. By analyzing trees' genetic makeup from different provenances in these trials, researchers can identify specific genes or genomic regions associated with adaptive traits, such as drought tolerance, cold hardiness, or disease resistance. This genetic information is crucial for understanding how tree populations have evolved in response to historical environmental pressures and how they might respond to future changes.

Landscape Genomics: Connecting Provenance Trials to Environmental Gradients

Landscape genomics studies how environmental gradients and spatial factors shape genetic variation across populations. It combines genomic data with environmental and geographic information to identify patterns of local adaptation and predict the potential of populations to cope with changing environments.

Provenance trials play a pivotal role in landscape genomics by providing controlled environments where the influence of specific environmental factors on genetic variation can be isolated and studied. By comparing the genetic data from different provenances grown under the same conditions, researchers can identify genetic markers linked to traits that vary along environmental gradients, such as altitude, temperature, or precipitation.

These insights can help understand the genetic basis of local adaptation, revealing which tree populations are likely to thrive under specific environmental conditions. Additionally, landscape genomics can inform the selection of provenances for reforestation and restoration projects, ensuring that the chosen populations are well-suited to the target environments.

Genomic Offset: Predicting Future Adaptation Potential

Genomic offset is a concept that quantifies the mismatch between the current genetic makeup of populations and the conditions they are likely to face under future climate scenarios. This concept is particularly relevant in climate change, where the environment is shifting rapidly, and tree populations may struggle to adapt.

Provenance trials contribute to genomic offset studies by providing empirical data on how different provenances perform under various climatic conditions. Combined with genomic data, these trials allow researchers to model the potential for future adaptation. By projecting the future climate conditions of a given region and comparing them with the genetic traits of current tree populations, researchers can estimate the "genomic offset"—the degree to which populations will be maladapted to future conditions.

This information is crucial for conservation and management strategies. It can guide the selection of tree provenances that are more likely to succeed in future climates, a practice known as assisted migration. By choosing provenances with lower genomic offset, forest managers can enhance the resilience of ecosystems to climate change.

Integration and Application in Forestry and Conservation

Integrating provenance trials with landscape genomics and genomic offset transforms forestry and conservation practices. These combined approaches provide a comprehensive understanding of how tree populations are adapted to their environments and how they might cope with future changes.

In practical terms, this integration allows for more informed decision-making in forestry management, particularly in selecting provenances for reforestation and restoration projects. It also supports the development of climate-resilient forestry practices by identifying tree populations with the genetic potential to withstand future environmental changes.

Moreover, these approaches contribute to conserving genetic diversity, ensuring that important adaptive traits are preserved within tree populations. This progress is particularly important for maintaining forests' long-term health and sustainability in the face of global environmental changes.

Provenance trials are critical to landscape genomics and genomic offset studies, providing the empirical data needed to link genetic variation with environmental adaptation. By integrating these trials with advanced genomic techniques, researchers can better understand the adaptive potential of tree populations, predict their response to future climate conditions, and inform forest conservation and management strategies. As the field continues to evolve, the synergy between provenance research, landscape genomics, and genomic offset will play an increasingly vital role in addressing the challenges of climate change and biodiversity conservation.

5.4 ESTABLISHMENT OF PROVENANCE TRIALS

Provenance trials are a cornerstone of forestry research, offering valuable insights into tree populations' adaptability and growth performance from different geographic origins. Across the world, forestry practices have transitioned from the unregulated exploitation of natural forests to sustainable systems that prioritize natural and artificial regeneration in plantations. As noted previously, genetic variation among provenances often significantly determines a tree population's capacity to adapt to specific environments. Consequently, selecting the most suitable provenance for a given location is vital to ensure planted forests' survival, productivity, and overall health.

One of the primary applications of knowledge gained from geographic variation within tree species is guiding the selection of appropriate provenances for reforestation efforts. To clarify, the *seed collection zone* (also referred to as the provenance, seed procurement zone, or donor zone) describes the area where seeds are harvested. In contrast, the *planting zone* (also known as the plantation zone, receptor zone, or deployment zone) refers to the area where these seeds are eventually planted. When the seed collection and planting zones are ecologically similar and geographically close, the provenance used is classified as "local." However, if the seed collection and planting zones are geographically or ecologically distinct, the provenance is considered "non-local." Sometimes, the collection and planting zones are merged into a single unit, known as a *seed zone* or *seed transfer zone,* where seeds collected and planted within the same zone are deemed "local."

This section focuses on selecting provenances for operational reforestation within a species' native range when tree improvement programs have not yet provided genetically enhanced seed. It emphasizes the importance of defining seed transfer guidelines that outline the conditions under which seeds can be moved from their collection zones to planting zones, ensuring ecological compatibility. Key topics covered include:

1. **Setting Objectives:** Establish goals that guide the selection of provenances for reforestation efforts.
2. **Insights from Past Trials:** Analyzing lessons learned from previous provenance studies.
3. **Decision-Making Tools:** Introducing a decision tree to assist in selecting provenances and making seed transfer decisions.
4. **Seed Transfer Guidelines:** Examining different types of seed transfer frameworks and their practical implementation.

Considerations related to choosing species and seed sources for planting in non-native environments and defining base populations for tree improvement programs are addressed in more detail in other books (Eriksson et al., 2008; White et al., 2007).

Traditionally, these trials are established through a systematic and rigorous process:

1. Seed Collection:

Selection of Provenances: The first step involves selecting tree populations from different provenances geographic areas where the trees have naturally grown and evolved. These provenances are chosen based on the research objectives, such as testing adaptability to specific climates or assessing genetic diversity.

Seed Harvesting: Seeds are collected from a representative sample of trees within each selected provenance. Care is taken to ensure genetic diversity by harvesting from multiple parental trees. The well-thought sampling design in the natural stands is critical for future comparability. In some cases, individual provenances retain a **half-sib structure**, which can be later utilized in genetic evaluation.

2. Site Selection and Experimental Design:

Environmental Considerations: The trial sites (common gardens) represent a range of environmental conditions, such as variations in soil type, altitude, and climate. These deliberate gradients help in assessing how different provenances perform across diverse conditions.

Uniformity: Within each site, efforts are made to ensure uniformity in soil preparation, planting conditions, and management practices to minimize extraneous variables that could affect the results.

5.4.1 EXPERIMENTAL DESIGNS IN PROVENANCE RESEARCH:

Randomized Block Design:

Provenance trials often use a randomized block design, where each provenance is planted in multiple blocks to account for site variability. This design allows for more robust statistical analysis by controlling for environmental heterogeneity within the site. Multiple replicates of each provenance are planted to ensure the results are statistically significant and not due to random variation.

Alternatives to the Randomized Block Design in provenance trials offer different approaches to controlling environmental heterogeneity and ensuring robust statistical analysis. Here are a few commonly used options:

Completely Randomized Design (CRD): In a completely randomized design, each provenance is randomly assigned to plots without specific grouping or blocking. This design is simpler and often used when the experimental area is homogeneous, meaning minimal environmental variation across the site exists. CRD is quite easy to implement and analyze. Thus, it is suitable for small trials or highly uniform sites. At the same time, it could be less effective at controlling for environmental variability, which can lead to confounding factors influencing the results.

Split-Plot Design: A split-plot design is used when there are two levels of experimental factors, with one factor applied to large plots (main plots) and the other to smaller sub-plots within each main plot. Different provenances could be assigned to sub-plots in the context of provenance trials, while the main plots could represent different site treatments or environmental conditions. This design allows for testing multiple factors and their interactions, making it suitable for more complex trials. Designing and analyzing is consequently more complicated, requiring more sophisticated statistical methods.

Latin Square Design: The experimental area is divided into a square grid in a Latin square design, with rows and columns representing different factors (e.g., soil type, slope). Each provenance is planted once in each row and column, ensuring that every provenance is tested under various conditions. The main benefit is that it controls for two sources of variation simultaneously (e.g., along both axes of the trial area), reducing the impact of site variability. Still, it requires a square layout and is less flexible regarding the number of treatments compared to randomized block designs.

Augmented Design: Description: In an augmented design, some provenances are tested in more plots than others. This is particularly useful when new provenances are introduced into an existing trial. The new provenances are compared with check provenances that are replicated more frequently. Advantages: Efficient use of resources, including many provenances, without needing full replication. Disadvantages: Statistical analysis can be more complex due to the unbalanced nature of the design.

Honeycomb Design: In a honeycomb design, trees are planted in a hexagonal pattern, with each tree surrounded by six others. This design minimizes competition and environmental effects by ensuring equal spacing and reducing edge effects. Advantages: Reduces the impact of competition between trees and provides a more uniform experimental setup. Disadvantages: Requires more space and is less commonly used, making it harder to compare with traditional designs.

Factorial Design: A factorial design involves testing all possible combinations of two or more factors. This could mean testing different provenances under different environmental treatments (e.g., different moisture or temperature levels) for provenance trials. This approach allows for examining interactions between factors, providing a deeper understanding of how various variables influence provenance performance. Disadvantages: It can become very complex with multiple factors, requiring many plots and sophisticated analysis.

Nested Design: In a nested design, one factor (e.g., site) is nested within another (e.g., provenance). This design is useful when hierarchical relationships exist between the factors, such as different sub-regions within a larger region. Replicated trials in a nested structure are useful for understanding variation within and between groups, such as within-provenance versus between-provenance variation. Disadvantages: Complex to set up and analyze, with specific requirements for replication and data structure.

While the Randomized Block Design is widely used for its simplicity and effectiveness in controlling environmental variability, these alternative designs offer different advantages. They can be more suitable depending on the specific objectives, site conditions, and constraints of the provenance trial. Selecting the appropriate design involves balancing the need for statistical rigor with practical considerations such as site variability, resource availability, and the complexity of the trial.

5.4.2 PROVENANCE TRIALS IN THE LONG TERM

Standardized Planting: Seeds or seedlings are planted simultaneously using standardized methods. The spacing between trees is carefully planned to prevent competition and allow for accurate growth and survival assessment.

Ongoing Maintenance: The trial sites are maintained with consistent care, including weed control, pest management, and monitoring for any environmental changes that might impact the trial.

Data Collection and Analysis:

Long-term Monitoring: Provenance trials are typically monitored over several years or even decades to gather data on growth rate, survival, disease resistance, and other traits of interest.

Data Analysis: The collected data is analyzed to compare the performance of different provenances. This analysis helps identify which provenances best suit specific environments or management objectives.

Recent advancements in technology and methodology have significantly enhanced the establishment and analysis of provenance trials. These innovations include:

5.4.3 CURRENT ADVANCES - GENOMIC TOOLS AND CLIMATIC MODELING

Molecular Markers: Advances in molecular biology allow DNA markers to assess genetic diversity and relatedness among provenances before trials are even established. These techniques help select diverse provenances likely to exhibit various adaptive traits.

Genome-Wide Association Studies (GWAS): GWAS can identify specific genes associated with adaptive traits by correlating genetic data with phenotypic performance in provenance trials, enabling a more precise selection of provenances for specific environmental conditions.

2. Climate Modeling and Genomic Offset:

Predictive Modeling: Climate models predict future environmental conditions, and genomic offset models assess the potential for different provenances to adapt to these changes. This information is used to establish trials designed to test the adaptability of provenances to future climates.

Assisted Migration Trials: With the knowledge gained from genomic offset models, provenance trials can be established in regions where the climate is expected to change, testing the potential of different provenances to thrive in these predicted future environments.

5.4.4 COLLABORATIVE AND GLOBAL NETWORKS:

International Provenance Trials: Collaboration across countries and institutions has led to establishing global provenance trials covering a wide range of environments and provenances. These trials provide valuable data that can be applied across different regions and climates. Current research still capitalizes on the major effort of the 1969 series of IUFRO provenance experiments with Norway spruce.

Open Data and Repositories: The data from provenance trials is increasingly being shared through open-access databases, enabling researchers worldwide to contribute to and benefit from existing knowledge.

The traditional establishment of provenance trials has provided a solid foundation for understanding the adaptability and performance of different tree populations. However, integrating new technologies and methodologies has transformed these trials into more powerful tools for predicting and managing the response of forests to changing environments. These advances are helping to ensure that forests remain resilient and sustainable in the face of global challenges such as climate change and biodiversity loss.

The large majority of wide-ranging tree species exhibit complex patterns of geographic variation with clines, ecotypes, and other patterns occurring within the same species over broad and fine geographic scales. These patterns reflect adaptation to current and past edaphoclimatic conditions and distinct biogeological histories, resulting in genetic drift and discontinuities in the species' distribution, slowing gene flow.

Genetic patterns vary among tree species, so proper genetic studies are critical. Results from short-term seedling studies and genetic marker studies can be useful for understanding evolutionary forces, planning gene conservation programs, and defining preliminary seed transfer guidelines. However, definitive seed transfer guidelines should be based on long-term data from field provenance tests established in the target planting zone. When these data do not exist, lessons learned from previous experiments indicate that (1) Planting the local provenance is the safest choice but may not result in maximum growth or product yield, and (2) Transfer of seed from a seed collection zone with a slightly milder climate to a planting zone with a slightly harsher climate may result in increased growth above the local source with little risk of increased maladaptation.

If the goal is to maximize product yield and sufficient data from long-term provenance trials are available, then three options exist for seed transfer, depending on the results: (1) When no significant differences among provenances are exhibited in the planting zone, the forester can deploy any provenance to any planting site; (2) When strong provenance differences are found, but no provenance x site interaction exists, the best provenance(s) can be deployed to all planting sites; or (3) When strong provenance x site interactions are observed, seed transfer guidelines must match provenances to the specific types of planting sites in which they strive.

5.5 ASSISTED MIGRATION

In the face of rapid environmental changes, forest ecosystems face significant risks due to the insufficient natural migration rates of trees, potentially resulting in adaptation lags and weakened forest health. This threatens the role of European forests as long-term carbon sinks. Assisted migration has emerged as a potential management strategy to mitigate these impacts by intentionally relocating forest reproductive material, such as seeds or seedlings, to more suitable regions.

Historically, concerns about maladaptation and genetic disruption discouraged assisted migration, favoring the **"local is best"** paradigm. However, climate change has highlighted the limitations of local seed sources, particularly those affected by bottlenecks or founder effects. Research now focuses on understanding tree populations' responses to climatic variables to identify productive forest reproductive materials for specific zones.

Recent studies, such as (Aitken et al., 2008; Chakraborty et al., 2024; Chludil et al., 2025) have assessed the potential of this strategy. Chakraborty et al. (2024) have modeled assisted migration strategies for seven major European tree species using advanced climate data from publicly available datasets. By incorporating species distribution models (SDMs) and universal response functions (URFs), they evaluated scenarios for carbon storage in juvenile forests under different seed-sourcing approaches. Two scenarios were tested: 1) "local seeds," using geographically proximate and climatically matched seed sources, and 2) "adapted seeds," selecting seed provenances optimized for carbon sequestration regardless of geographic or climatic similarity to the planting site.

The results emphasize the potential of assisted migration to enhance forest resilience and carbon storage under changing climatic conditions, demonstrating the importance of integrating genetic diversity and environmental suitability into reforestation strategies.

Assisted migration, while addressing climate change by relocating species to areas with favorable climatic conditions, may overlook non-climatic factors like soil properties and photoperiod cues, potentially leading to maladaptation and reduced fitness. Soil nutrient availability, interactions with soil organisms, and seed predation can hinder

plant establishment despite optimal climate conditions. Additionally, the practice risks introducing non-native pests and pathogens, as international plant trade inspections are limited and often ineffective.

Economic and logistical challenges further complicate assisted migration. Forest nurseries face increased costs and complexities in sourcing seeds suited to future climates, along with competition from larger foreign suppliers. Many nurseries prioritize commercially valuable species, limiting the ecological diversity needed for climate adaptation. Regional variations in demand for non-native species or resilient local provenances add to uncertainties.

Nurseries require significant investment to meet rising demands for diverse, climate-adapted seedlings, but financial risks and declining numbers of nurseries, particularly in Europe and North America, pose barriers. Legal frameworks, including the Nagoya Protocol and national seed zone restrictions, also constrain assisted migration. A flexible policy framework is crucial to ensure long-term forest health, accommodating diverse scenarios and enabling effective reforestation strategies.

5.6 REFERENCES

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evol Appl*, *1*, 95–111. https://doi.org/10.1111/j.1752-4571.2007.00013.x

Eriksson, G., Ekberg, I., & Clapham, D. (2008). *An Introduction to Forest Genetics*. https://res.slu.se/id/publ/22529

Chakraborty, D., Ciceu, A., Ballian, D., Benito Garzón, M., Bolte, A., Bozic, G., Buchacher, R., Čepl, J., Cremer, E., Ducousso, A., Gaviria, J., George, J. P., Hardtke, A., Ivankovic, M., Klisz, M., Kowalczyk, J., Kremer, A., Lstibůrek, M., Longauer, R., … Schueler, S. (2024). Assisted tree migration can preserve the European forest carbon sink under climate change. *Nature Climate Change*, *14*(8), 845–852. https://doi.org/10.1038/s41558-024-02080-5

Chludil, D., Čepl, J., Steffenrem, A., Stejskal, J., Sagariya, C., Pook, T., Schueler, S., Korecký, J., Almqvist, C., Chakraborty, D., Berlin, M., & Lstibůrek, M. (2025). A Pollen-Based Assisted Migration for Rapid Forest Adaptation. *Global Change Biology*, *31*(1), e70014. https://doi.org/10.1111/gcb.70014

Korecký, Y. A., J. ,. Čepl, J. ,. Stejskal, J. ,. Faltinová, Z. ,. Dvořák, J. ,. Lstibůrek, M. ,. El-Kassaby. (2021). Genetic diversity of Norway spruce ecotypes assessed by GBS-derived SNPs. *Scientific Reports*, *11*, 23119.

Langlet, O. (1958). *A cline or not a cline-a question of Scots Pine.* https://www.cabidigitallibrary.org/doi/full/10.5555/19596602808

White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI.

6. TREE IMPROVEMENT

Tree improvement combines breeding and silvicultural activities that are recurrently applied to exploit the existing natural genetic variation in forest tree populations. The aim is to increase the economic value of artificially regenerated forests and their products.

Modern, deliberate, and organized tree improvement as we know it today started in the 1950s. Compared to other crops, this substantial delay allowed tree breeders to utilize the wealth of knowledge and experience accumulated over centuries in different fields. Initially, tree improvement activities aimed to satisfy the immediate demand for high-quality timber products without long-term planning. Reasonable gains were quickly achieved through intensive selection, selecting only a small fraction of the best individuals. However, this short-term approach was abandoned as it resulted in a shallow genetic base that needed to be revised for future improvements when transitioning to long-term programs.

Most long-term tree improvement programs use a recurrent selection breeding cycle framework comprising two independent but compatible phases: breeding and production. These phases are tightly linked, with the main objective being to obtain and retain a broad genetic base while combining desired characteristics into high-quality, well-adapted trees valuable for the future.

The **breeding phase** of the cycle consists of:

1. **Selection**: Identifying the most desirable individuals in natural forest stands, artificial plantations, or advanced generations of improvement according to specific criteria aligned with breeding goals.
2. **Breeding**: Producing controlled crosses among selected individuals to extract genetic parameters for subsequent breeding rounds.
3. **Testing:** Evaluating the genetic quality of selected trees through testing their progenies (i.e., backward and forward selection – see next sections).

In the **production phase**, the breeding efforts materialize in seed crops delivered to produce new, genetically improved forest stands or plantations.

Tree improvement programs are resource-intensive and long-term endeavors; therefore, exploring less demanding alternative approaches that deliver acceptable and satisfactory gains is vital. Alternative strategies, often coined as low-input breeding, will be elaborated in the next sections.

A **base population** is a source from which individuals are selected to enter a breeding program. The term was used for any starting material of a breeding program irrespective of its genetic background. The alternative terms **founder** and **recruitment population** distinguish between different generations of breeding. The founder population consists of individuals selected in "wild" forests with no previous breeding history, using individual tree selection methods. In Europe, we would probably use the term natural forest stands, but similarly to the term "wild stands," it must be taken with a grain of salt. The recruitment population, functionally identical, originates from advanced generations of breeding. Founder populations can be considered special, first-generation recruitment populations containing assumed unrelated and non-inbred individuals. However, modern genomic methods challenge this assumption as they consider the realized relationship based on SNP markers (see the section on genomics). A gene resource population is defined in natural stands or clonal archives, primarily for gene conservation rather than improvement.

Each generation establishes a **candidate population** from a subset of phenotypically superior individuals drawn from the base or recruitment populations. This population usually has a census number significantly lower than the base population but is expected to generate gains for desirable traits in the next generation. The genetic superiority of these selected trees is evaluated through progeny tests, which are essential for any breeding program. These tests help exclude low-ranking parents from both breeding and production populations.

A **breeding population** is a subset of selected individuals proven to have desirable genetic qualities through progeny testing, serving as parents for subsequent generations. Establishing a breeding population involves balancing high gains through a strong selection differential (if this term is new or misleading, go back to the chapter

on quantitative genetics) and maintaining sufficient genetic variation to ensure future progress. At any stage of any breeding generation, the breeding population can be supplemented by material from external sources, called an infusion population, after selection to fit the improvement program's objectives.

The **production** or **deployment population** is independent of all breeding cycle populations. This highly selected subset of the breeding population includes only the very best individuals. It is dedicated to producing genetically improved seeds (seed orchards) or vegetative propagules (stool beds/hedges) for operational reforestation programs, including clonal forestry. **Seed orchards** will be thoroughly covered in the next section as they are the most important and abundant global production populations. Although establishing production populations is often seen as the final step in each tree improvement cycle before operational plantations, considering and moderating other factors such as seed harvest, handling, storage practices, variance in seed germination, and nursery management practices is crucial. These factors significantly impact the regeneration material's genetic gain and diversity. For more details, please refer to (Eriksson et al., 2008; White et al., 2007; Zobel & Talbert, 1984). Relevant management practices are outlined in later sections. The breeding cycle we just described is visualized in Fig. 6.1.

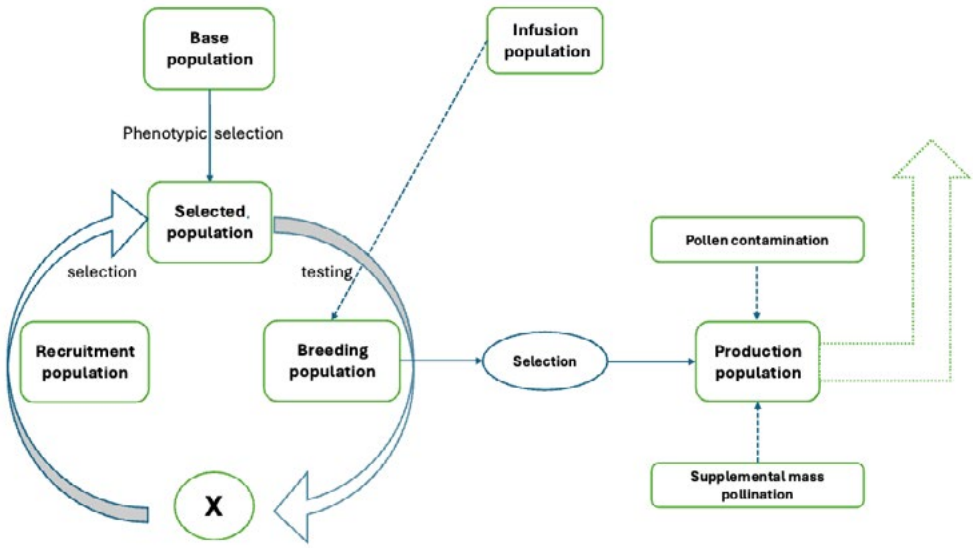


Fig. 6.1 Basic types of population in the breeding program, modified according to (Funda & El-Kassaby, 2012).

6.1 SEED ORCHARDS

Seed orchards serve as the critical link between tree breeding and reforestation efforts. For breeders and forest geneticists, these orchards are essential for deploying improved and genetically superior material. Interestingly, a single seed orchard can function both as a breeding and production population. As a breeding population, the orchard facilitates crossings, producing seeds that result in seedlings planted in experimental layouts within progeny trials. As a production population, the orchard generates seeds for sowing in nurseries and subsequent reforestation activities.

Globally, seed orchards are the most widely used sexually reproducing forestry production populations. Their efficiency is crucial, as it determines the genetic gain and diversity of future forest plantations. This chapter delves into various topics, beginning with the role of seed orchards in the tree improvement cycle. It examines the population genetics model of seed orchards and the biological factors influencing this model, including reproductive investment and success, reproductive phenology, inbreeding, gene flow/migration, and reproductive biology.

Additionally, we review management practices, focusing on those implemented during the establishment phase (such as orchard size, design, number of parents, and their representation) and those aimed at enhancing crop

management (including crown management, supplemental mass pollination, bloom delay, and selective seed harvesting). The impact of these practices on genetic structure is also discussed.

Seed orchards consist of selected, genetically superior parent trees. For forestry practitioners, the primary objective of a seed orchard is to produce abundant and easily harvestable yields of genetically improved seeds. While the ease of seed harvest is a key consideration, this book aims to demonstrate that seed orchards offer much more. As the vital link between tree breeding and silvicultural activities, they translate breeding efforts into improved seed crops utilized in operational forestry as genetically enhanced seedlings. This connection underscores the importance of seed orchards in the broader context of forest management and genetic improvement.

Dutch colonists in Java established the first seed orchards as early as 1880 to propagate the indigenous plant *Cinchona ledgeriana* (Funda & El-Kassaby, 2012). However, the widespread adoption of this concept only occurred about seven decades later. In the 1950s, seed orchards became integral to breeding programs worldwide.

Seed orchards can be categorized based on the breeding generation and the type of plant material used, whether **seedling** or **clonal**. Another classification considers the type of material included in the seed orchard. These options may consist of interspecific seed orchards, which contain clones from two species; inter-provenance seed orchards, which include clones from two different provenances; and intra-provenance seed orchards, where the clones originate from a single provenance.

First-generation seed orchards are established with parents selected solely based on their phenotypes in forest stands or unimproved plantations (founder populations), with their genetic worth being speculative. These orchards can be improved by removing inferior parents, a process known as genetic thinning or roguing. After this, the seed orchard is called a **1.5-generation seed orchard**. This improvement is guided by the results of progeny trials, utilizing backward selection to enhance the genetic quality of the orchard (Fig. 6.2).

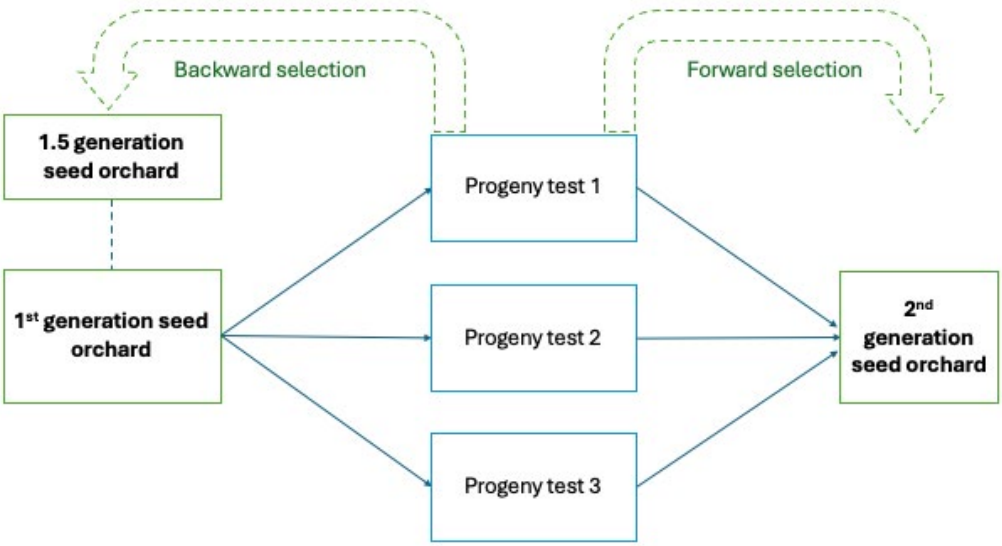


Fig. 6.2 Sees orchard establishment in the breeding strategy context.

Progenies are raised and planted in progeny trials, which typically include multiple progenies. These progenies are often called **full-sib** or **half-sib families**, depending on the type of cross used to create them (detailed further in the section on **mating designs**). A primary goal of progeny trials is to estimate the genetic value of the parental trees. The genetic quality of the parent trees is assessed based on the performance of their offspring in well-designed, replicated experiments.

Parental selections for genetic thinning or for creating new re-grafted 1.5 generation seed orchards are based on these progeny trial evaluations. However, this selection process can only be conducted once or twice before

the pool of available parents becomes too small, leading to an unsatisfactory breeding population size in terms of effective population size (N_e). To address this issue, the best trees in the best families are selected (**forward selection**). Scions for establishing **second-generation seed orchards** are collected from these selected trees within tested full-sib families. The whole scheme is graphically depicted in Fig. 6.2.

As previously explained, advanced-generation seed orchards are established using individuals from controlled crosses between elite genotypes. An alternative method involves using their open-pollinated progenies and subsequently reassembling a full pedigree, thereby achieving a similar level of information as testing full-sib progenies. This alternative approach has its pros and cons, supported by empirical findings.

One of the most ambitious practical experiments on European larch involved pedigree reconstruction, assembling 491 full-sib families (1,260 offspring), representing 35% of the 176 possible 53-parent half-diallel crosses (detailed further in the section on mating designs). This study describes the largest known forest tree pedigree assembly (Fig. 6.3). Such an elaborate crossing scheme could not be accomplished in a single breeding season.

The primary disadvantage of this method is the lesser control over background pollination, the consequences of which will be further discussed. Traditional controlled crosses and pedigree reconstruction aim to increase genetic gain in the next generation through forward selection.

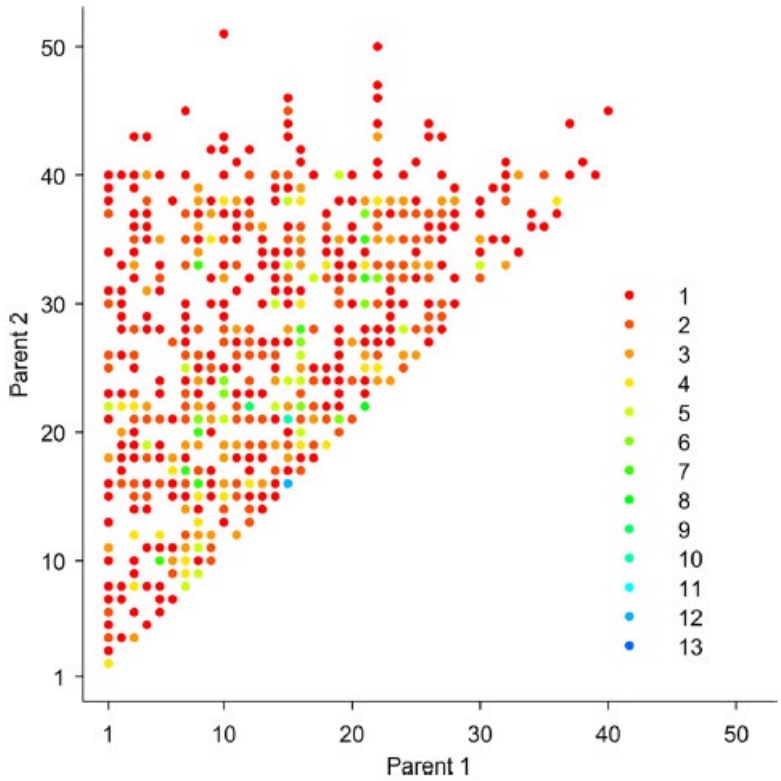


Fig. 6.3 Pedigree reconstruction in full-sib families concerning individual family sizes (Lstibůrek et al., 2020)

Seed orchards are further distinguished based on the material used for their establishment:

Clonal Seed Orchards comprise vegetative propagules such as grafts, cuttings, or tissue-culture plantlets collected or cultivated from plus trees. These orchards involve fewer individuals than seedling seed orchards but offer the advantage of higher genetic gain due to the known pedigrees of the clones. Clonal seed orchards enable earlier flowering and more controlled breeding, thus enhancing the predictability and efficiency of genetic improvements.

Seedling Seed Orchards are established from the generative progeny of plus trees, which can be maternal half-siblings from open pollination or full-siblings from controlled crosses. These orchards tend to have a broader

genetic base due to the larger number of individuals involved in mating, often with unknown male parents in open-pollinated progenies. However, they are expected to provide relatively lower genetic gain because their pedigrees are unknown. Seedling seed orchards are particularly prominent for early flowering species such as Eucalyptus spp., black spruce, some pine species, and hardwoods like oaks in Europe. In these orchards, flowering typically starts later than in clonal seed orchards with grafts. To stimulate flowering, seedling seed orchards are often deliberately established in warmer climates. Additionally, they are used for species that suffer from scion-rootstock incompatibility, such as Douglas-fir. Seedling seed orchards are usually established as progeny trials, serving the dual purpose of seed production and progeny testing, provided the trees are not pruned.

However, most seed orchards are clonal and typically consist of 30–50 parents (the original plus trees), each represented by one or more vegetatively propagated copies known as **ramets**. In most intensive breeding programs, plus tree candidates undergo rigorous scrutiny during selection. A significant issue is that selection often occurs in stands originating from natural regeneration, where quality may differ from the planted stands where the bred material will grow in the future. Additionally, the risk of selecting related plus trees is higher in naturally regenerated stands than artificially regenerated ones.

During the initial selection of plus trees, their growth is compared with neighboring trees in the same stand to account for micro-environmental effects. Wood samples may be taken to correct for age differences between comparative standards and plus trees. Despite these measures, achieving unbiased selection in uneven-aged stands remains challenging.

Seed orchards are established assuming they will function as closed, perfect, panmictic populations (see the chapter on population genetics). In such populations, all individual ramets are expected to have an equal chance to mate with all other ramets, including selfing. If this assumption holds, the allele and genotype frequencies in the offspring population would mirror those of the parental population without evolutionary forces such as migration, selection, mutations, and random genetic drift. Consequently, the genetic gain achieved in previous breeding cycles would be effectively transmitted to the next generation. However, this assumption is overly optimistic, and actual outcomes often differ significantly, making the genetic quality of seed crops unpredictable.

In tree improvement programs, genetic quality is evaluated from the breeders' perspective and can be quantified and compared with founder populations. Genetic quality includes two components: **genetic gain** and **genetic diversity**. These components often act in opposite directions, making their balance complex in breeding programs.

6.1.1 GENETIC GAIN AND DIVERSITY IN SEED ORCHARD CROPS

Most of the phenotypic traits undergoing breeding are polygenic and follow the quantitative mode of inheritance, where the environment could play a major role. The driver of genetic gain for quantitative traits is the response to selection (R), which is the shift between the mean phenotypic value for the offspring of the selected parents and that of the parental generation due to selection. It can be expressed as $R=b_{op}\,S$, where bop is the regression of offspring on parents and S is the selection differential, i.e., the average superiority of the selected parents (see the chapter on quantitative genetics). Under certain assumptions, the regression can be replaced by narrow-sense heritability (h^2), which is the ratio of additive genetic variance to total phenotypic variance, and the previous equation suddenly gets more familiar: $R=S\,h^2$. To quantify the success of the improvement program, the superiority of selected parents must be evaluated relative to the original population using the concept of **breeding values**. A breeding value is the sum of the **average effects** of all quantitative trait loci that an individual carries (summed over the pair of alleles at each locus and over all loci affecting the trait of interest) or, more practically, it is a value of an individual based on the field performance of its progeny. Breeding values can be expressed in the same absolute units as the trait in question or, after standardization, as a deviation from the population's mean.

Genetic gain then equals the average breeding value of all individuals in a production population, which is transmitted to the next generation under the assumption of random mating. The potential or expected genetic gain is calculated similarly in seed orchards, but the breeding values are weighted by parental representation (i.e., the number of ramets representing a particular parent). However, the genetic gain realized using seed crops is still being determined. It may substantially differ from the expected value, depending on the variation in parental

reproductive success and the magnitude and quality of pollen contamination. Genetic gain (G) is calculated using the Breeder's Equation (Chapter 4 on Quantitative Genetics).

Genetic diversity is a key parameter to consider in breeding programs. It is defined as the total amount of genetic variation individuals carry within or among genetic units because of their evolution. It forms the basis for their response towards biotic and abiotic influences; thus, genetic diversity is regarded as a prerequisite for adaptability and further evolution. According to Fisher's fundamental theorem of natural selection, the rate of increase in the mean fitness of any population at any time is directly related to its genetic variance in fitness at that time. Genetic diversity can be interpreted in several different ways. It is commonly expressed as heterozygosity, the average number of alleles per locus, or the average proportion of polymorphic loci. It can be estimated from several different types of genetic markers. Another option is to relate genetic diversity to the concept of common ancestry (co-ancestry), which is defined as the probability that any two randomly sampled alleles (one from each parent) at a given locus are identical by descent (IBD) (see the chapter on population genetics – calculation of inbreeding. Genetic diversity is then calculated as:

$$GD = 1 - \Theta \tag{6.1}$$

Where Θ is the average co-ancestry of all pairs of population members, including themselves (also termed group co-ancestry) or, in other words, the accumulated loss of genetic diversity.

Most often, genetic diversity is approximated by effective population size, a concept introduced earlier in this book that quantitative and population geneticists widely use and include in most animal and plant breeding programs. Effective population size accounts for factors that reduce a population's census number to a size relevant to evolution. These include situations compatible with typical seed orchard development when: (i) breeding individuals only represent a portion of the entire population because some individuals have not reached sexual maturity or participated in mating, (ii) there are unequal numbers of male and female breeding individuals, (iii) parental gametic contributions vary (i.e. there is unequal reproductive success among parents), (iv) there is spatial and temporal variation in population size and (v) individuals that mate are related and/or inbred. A ratio between the effective and census population sizes is often calculated so that the impact of the factors mentioned above can be directly quantified. The related phenomenon of inbreeding depression, despite being nearly universal, varies across species and even between populations within species in its impact. To provide a tool for breeders that quantifies the reduction in diversity due to co-ancestry and/or inbreeding (for instance, in a population of parents in a seed orchard), Lindgren et al. (1995, 1997) developed a measure called status effective number (N_s), which can be interpreted as the size of an idealized population consisting solely of unrelated and non-inbred individuals. It is estimated as follows:

$$\mathrm{N_s} = \frac{1}{2\Theta} = \frac{1}{2\sum_{\mathrm{i=1}}^{\mathrm{N}}\sum_{\mathrm{j=1}}^{\mathrm{N}}\mathrm{p_i p_j c_{ij}}} \tag{6.2}$$

Where N is the population size, p_i, and p_j are parental representations of parent i and j, respectively (i.e., their expected or actual reproductive successes), and c_{ij} is coancestry between them. Consequently, status number can be derived for any population linked by a pedigree to an initial population with known co-ancestry and inbreeding. In addition to the parental population, this concept can be applied to estimate the genetic diversity of seed crops, whereby parental representations pi from the previous equation (which are equal to the expected gametic contributions of the respective parents) can be replaced by actual parental reproductive success determined from parentage analyses using DNA markers (Funda et al., 2008) or approximated by visual assessment of their fecundities. If all parents are unrelated and non-inbred, then the co-ancestry matrix with elements c_{ij} takes the form $0.5I$ (where I is an identity matrix), and N_s simplifies to:

$$\frac{1}{\sum_{\mathrm{i=1}}^{\mathrm{N}}\mathrm{p_i^2}} \tag{6.3}$$

A form is known under the term 'effective number of parents' but used and interpreted by many authors as effective population size because it reduces the census number of parents to that of an ideal population in which all parents have an equal chance of mating with all other parents.

Furthermore, if all parents have equal reproductive success, the status number reaches its maximum and is equal to the actual census number of the parental population. This feature is particularly useful because it facilitates simple interpretations of the expected impact of relatedness and inbreeding present in the population and the variance in parental reproductive success on the reduction in genetic diversity experienced by the next generation (i.e., a seed crop). Unlike the effective population size, this concept makes no comparison with Wright–Fisher's ideal population; rather, it simply compares against the reference population.

Seed orchards may possess an even greater allelic richness than is, on average, found in natural populations. This phenomenon is probably the result of the plus tree selection being commonly conducted over a wider geographic area, thus covering individuals representing many subpopulations, and stresses the positive role of seed orchards as they associate individuals who would normally not be able to interbreed due to the presence of physical barriers. This feature is particularly important in insect-pollinated species with a scattered distribution, such as the highly valuable wild cherry or wild service tree (*Sorbus aucuparia*) in Europe, which otherwise have very limited access to pollen from more distant and thus (probably) less related sources.

A multi-step selection approach for establishing first-generation seed orchards that begin with selecting a large pool of plus trees exhibiting desired quantitative traits within a base population was proposed. After the initial phenotypic selection, these trees undergo genetic screening to assess marker-based diversity, which enables further refinement of the selection pool. This final group of genotypes is chosen to meet breeding objectives and genetic diversity standards—achieving a balance between homozygosity for advantageous alleles and the conservation of rare alleles critical for resilience.

Originally, this selection process relied on isozyme markers; however, advances in genomics now make it possible to enhance this approach using genome-wide scans. By incorporating data from single nucleotide polymorphisms (SNPs), often associated with traits of interest, breeders can more precisely select individuals that meet phenotypic and genetic diversity targets. This refined, SNP-informed selection process can improve orchard establishment or be applied to thinning, offering a more robust and nuanced method for achieving genetic diversity while enhancing desirable traits.

6.1.2 INNER AND OUTER FACTORS AFFECTING SEED ORCHARDS

Both genetic gain and diversity are crucial measures of the overall genetic quality of orchard seed. Genetic gain estimates a seed lot's average improvement for a desired trait, while genetic diversity quantifies its potential to cope with unpredictable events such as extreme weather or pest problems. However, several factors hinder random mating among seed orchard parents, reducing the effective transmission of desired genes from parental to offspring populations and causing deviations from Hardy–Weinberg equilibrium. The following section addresses these factors and their interactions when relevant.

Reproductive success

Reproductive success is the relative gametic contribution of a parent to the next generation, often measured by the number of offspring produced. Several methods estimate parental reproductive success in seed orchards. The simplest assumes it is a function of parental representation, with gamete production reflecting the number of ramets representing a parent. Other methods approximate reproductive success based on reproductive investment, assuming a high positive correlation between gamete production (fecundity) and actual reproductive success. These include ocular scoring of generative buds, strobili, or flowers, measuring seed-cone volume or mass, counting seeds per cone, and estimating the proportion of filled seeds per parent. Variation in parental reproductive phenology has also been proposed. Studies show substantial variation in reproductive investment among parents of nearly all forest tree species. The '20:80 rule' asserts that 20% of an orchard's parents produce 80% of the seed-cone crop, with minor deviations observed in black spruce, radiata pine, loblolly pine, and Douglas-fir.

With the rapid development and spread of various molecular genetic markers in recent decades, conclusions based on scoring parental fecundities in seed orchards could be verified. Many studies report a substantial male imbalance, confirming the over-representation of a few highly fecund male parents in tree seed crops. However, few studies have used molecular markers to investigate the relationship between reproductive investment and reproductive success in seed orchards. Studies using microsatellite DNA markers in Nordmann fir, Japanese black pine, sugi, and *Eucalyptus nitens* seed orchards found that male fecundity explained 76%, 43%, 15%, and 58% of the variance in reproductive success, respectively.

Reproductive phenology

Depending on the species and breeding program, seed orchard parents often originate from large geographic areas, adapting to different climatic conditions. Synchrony in reproductive phenology is crucial for random mating and balanced reproductive success among orchard parents. Reproductive phenology involves life cycle events like bud burst, pollen shedding, and seed-cone receptivity, influenced by seasonal and inter-annual environmental variations. Since reproductive phenology is strongly genetically controlled, selected parents retain their characteristics even in new environments. Thus, parents nearby may still form non-overlapping phenological classes, creating temporally isolated subpopulations with limited gene exchange.

Asynchronous phenology has been observed in seed orchards of Douglas-fir, Sitka spruce, radiata pine, loblolly pine, and black pine. At the same time, stronger synchronization occurs in Scots pine, black spruce, and Norway spruce. Female reproductive phenology, especially the initiation of receptivity, is more genetically controlled than male phenology, with pollen shedding largely determined by environmental factors.

Variation in reproductive phenology reduces potential mating combinations, leading to **positive assortative mating** (preferential mating between similar phenotypes) and imbalanced reproductive success. Highly fecund parents may be excluded from reproduction if males shed pollen when females are not receptive or receptive females lack ambient pollen from orchard individuals.

Gene flow/migration

Gene flow indicates genetic exchange between populations through migration. Gene flow is easier to track in seed orchards because trees are sessile, and the focus is on external pollen entering the orchard. By genotyping all male candidates' parents within an orchard and assigning offspring to these candidates, it is possible to distinguish between internal and external pollen sources. Gene flow from unselected, external pollen sources, known as **pollen contamination**, significantly impacts the genetic quality of seed crops.

Pollen contamination levels vary by species, mating system, reproductive phenology, orchard size, isolation degree, and environmental conditions. Studies show contamination rates from nearly 0% to 90%. For example, a Norway spruce seed orchard showed consistent contamination levels of 69-71% over three years, while a Douglas fir orchard varied from 44% to 89% over two years.

The two components of seed crop genetic quality, genetic gain, and genetic diversity are affected by contamination in the opposite direction. Genetic gain decreases with increasing contamination unless the contaminant sources have equal or higher breeding values than the orchard parents. Genetic diversity typically increases with additional parental contributions from pollen contamination, though high contamination can either increase or decrease diversity depending on the genetic characteristics of the source population. In conifers, adapted for aerial pollen transport, contamination can be especially high if nearby stands of the same species exist.

Despite 100% contamination rates, half of all genes will originate from within-orchard parents. However, genetic gain reduction is more pronounced if contaminating pollen comes from poorly adapted trees. Additionally, nursery--stage selection of faster-growing seedlings from contaminant pollen may lead to maladaptation.

Quantifying pollen contamination in clonal seed orchards using paternity analysis with molecular genetic markers is straightforward. In seedling seed orchards, the higher number of genotypes complicates this process. Estimation based on differences in allele frequencies from outcrossing pollen, ovules, and outside seed is suggested. Using highly polymorphic genetic markers allows for high-confidence paternity inferences, and genome-wide scanning with SNPs may soon eliminate the issue of large parental population sizes.

6.1.3 SEED ORCHARD POPULATION

The number of genotypes included as parents in a seed orchard is one of the most fundamental questions, yet there is no universal answer. Theoretically, one single parent with the highest breeding value would generate the highest possible genetic gain in a seed crop. Still, the seed crop's genetic base would be too limited, and the resulting forest stands would lack resilience to unpredictable environmental changes. Furthermore, all the seed produced would be the product of selfing. Despite the relevance and importance of the number of parents in a seed orchard, this subject had yet to receive full discussion in the literature. The North Carolina State University Tree Improvement Cooperative (2001) recommended using 20–30 parents because a larger number would reduce potential gain; however, the actual numbers used vary widely from as few as 5–10 to about 90 parents. Lindgren & Prescher (2005) tried to provide a more generalized way of calculating the optimum number of parents but concluded that the range of optimum values was quite broad. For instance, 20 parents were chosen as the standard for Swedish orchards. This criterion applies only to unrelated, tested parents deployed in equal proportions (which may not be optimal for maximizing genetic gain), and the appropriate number of parents may be case-specific, i.e., when more relevant information is available, higher or lower numbers may be optimal. The census number of parents is a crucial starting point for seed orchard establishment, but as mentioned earlier, the census number may not be fully reflected in seed crops. Therefore, it was suggested that effective rather than census number of parents should be controlled for each seed lot and used as a criterion for estimating genetic diversity in seed crops because this measure better reflects their actual gametic composition. This suggestion was implemented, e.g., in British Columbia, Canada, where the effective number of any seed lot used for reforestation must not drop below 10, a value generally considered to capture the majority (95%) of genetic diversity existing in a base population.

Similarly, Lindgren originally suggested a value of 8 and for Alberta, Canada, however, a more conservative value of 18 has been implemented. Other theoretical and empirical data show that seed orchards with 20 or more parents should represent a similar level of risk as seeds collected from the natural population. Another aspect that should be considered is whether the reforestation material from seed orchards can regenerate naturally. Although it is common practice to clear fallen forest stands and exclude natural regeneration, it was suggested that if a species has a long rotation period and is normally naturally regenerated, more parents should be considered.

Historically, clonal seed orchards have been commonly established using a uniform spatial representation of parents across the population, such that the same number of ramets replicated each parent. When information on parental breeding values became available, parents above a certain threshold value were preferentially utilized, while those below were rejected. Lindgren suggested exploiting this information, such that parents with higher breeding values would be planted in higher proportions to boost the crops' genetic gain. In contrast, inferior parents would only contribute fewer replicates to maintain an acceptably high genetic diversity. Following this idea, Lindgren & Matheson (1986) the linear deployment concept was developed, whereby parental representation during orchard establishment would be linearly related to their respective breeding values. Although this certainly represented a step forward, it still possessed two major limiting assumptions: first, that seed orchard populations consisted solely of unrelated and non-inbred parents, and second, that parental reproductive success is a function of parental representation, i.e., production of viable seed and reproductively successful pollen is proportional to the number of ramets representing a particular parent.

Regarding utilizing the linear deployment concept, it was obvious that kinship and inbreeding considerations became inevitable as breeding programs advanced. Current development has tackled these issues. Please see the relevant literature; these issues must be simplified for our introductory course.

6.2 INBREEDING

Inbreeding refers to mating between genetically related individuals, including self-fertilization, and is usually quantified by the coefficient of inbreeding F, which equals the coefficient of co-ancestry or kinship among individuals in the previous generation. This has similarities with assortative mating; however, while assortative mating affects only a limited number of loci that control or are linked with a trait of interest, inbreeding affects all loci indiscriminately and causes genome-wide excesses of homozygosity. In many species, inbreeding results in a fitness decline due

to increased homozygous genotypes with deleterious alleles, called **inbreeding depression**. Studies on the genetic basis of inbreeding depression indicate that a few genes confer large deleterious effects, whereas most other deleterious effects are minor. Although the genetic consequences of inbreeding depression may vary among species and even among populations within species, it is a critical factor that must be considered in all breeding programs. This is especially true for monoecious species such as forest trees because of their self-fertilization capability; in each cycle of self-fertilization, individuals' homozygosity increases by one-half, thus increasing the chance of producing genotypes with rare deleterious alleles. Rare deleterious alleles arise mainly as unfavorable mutations or are evolutionary relics. Examples of unfavorable rare alleles can be found in the resistance of Douglas-fir to damaging biotic agents or the survival of loblolly pine, in which homozygotes for the rarer allele had lower values than homozygotes for common allele; resistance to pathogens in crops is often connected to such alleles. Some plant species have developed a variety of pre- and postpollination mechanisms to minimize or fully avoid selfing, such as self-incompatibility in angiosperms (i.e. wild cherry), dioeciousness, physical separation of male and female strobili within trees of monoecious species, and asynchrony of male and female reproductive phenology. Apart from self-fertilization in any generation's seed orchards, half-sib, full-sib, and parent-offspring matings are likely in advanced-generation orchards through recurrent selection. Negative effects of inbreeding, such as reduced seed set, survival, growth, and productivity, have been observed in many forest tree species.

Progeny of self-pollinated or otherwise inbred trees is typically substantially smaller than that of outcrossed trees, and this difference can be further exacerbated during stand development as the slowly growing inbred trees are eventually outcompeted for light and other resources by more vigorous outcrossed trees. Western red cedar appears to be an exception from this norm: its selfing rate was relatively high in both natural and artificial populations, but negligible inbreeding depression was determined. In a redwood, no consistent effect of selfing on several seeds per cone was found. However, under stress conditions, survival from self-fertilized seed was relatively much lower than from outcross seed because height was just 65–80% after one year's growth. Still, it was greater and more balanced under good nursery conditions. After 14 years in the field, the difference in growth was further accentuated, and the self-fertilized individuals averaged only 42% of the height of their outcrossed counterparts. It has been speculated that low to moderate levels of inbreeding, such as crossing half-sibs, may increase the production of filled seeds, leading to larger proportions of inbred seedlings that still meet nursery culling standards. Consequently, these inbred seeds may have an even more severe impact on the growth and yield of forestry plantations than self-fertilization. Based on a complex review of more than 50 tree species, conifer species' selfing rates were mostly low in both seed orchards and natural populations.

6.3 SEED ORCHARD DESIGN

Various **seed orchard designs** have been developed and implemented in real-world scenarios, with a common objective to minimize self-fertilization and promote panmixia. These designs assume that pollination efficiency is a function of distance, and thus, most genetic exchange occurs among neighboring ramets, while it decreases between those further apart. Since parents are usually replicated by several ramets in seed orchards, the objectives of all designs (i.e., minimizing selfing and promoting panmixia) have been secured through a specific spatial layout of ramets concerning their parental affiliation.

However, most marker-based analyses have challenged these main objectives and questioned the relevance of the seed orchard designs that are mostly used. According to a Burczyk & Prat (1997) study on a Douglas-fir seed orchard, the proportion of matings that resulted from outcrossing within neighborhoods (30 m radius for each mother tree) was estimated to be 43%. The effect of distance and direction of individual males from mother trees, pollen fecundity and phenological synchronization were all significant in determining patterns of outcrossing within neighborhoods. Additionally, male reproductive success increased with pollen fecundity and proximity to and phenological overlap with a given mother tree. For comparison, Burczyk et al. (1996) also investigated pollen dispersal in a seed orchard of an insect-pollinated species, *Eucalyptus regnans*. They found that approximately 50% of effective pollen gametes were a product of males more than 40 m distant from the respective mother trees, making insect pollinators efficient promoters of cross-fertilization in this orchard. A similar pattern of no

preferential mating among the closest neighbors was found in seed orchards of *Eucalyptus grandis*, as the mean pollination distance was estimated to be 32 and 58 m, respectively. Only about one-half of the assigned pollen traveled < 50 m.

Moreover, long-distance pollen flow has been observed in both wind-pollinated and insect-pollinated species. Thus, it is unsurprising that matings within seed orchards are limited mostly by their physical size, i.e., the maximum possible distance among their members.

Probably the most successful design of the past was the permutated neighborhood design [COOL (Computer Organized Orchard Layouts)] developed by Bell & Fletcher (1978), which effectively separated ramets of a given parents by a predefined minimum number of positions between them.

As advanced-generation orchards became more frequent, it was necessary to consider the issue of relatedness and include it as an additional input factor for designing new orchards. A solution was provided by the minimum inbreeding (MI) seed orchard design incepted by Lstibůrek & El-Kassaby (2010). Layouts are compared using the 'minimum distance' concept for any selected deployment option. Later El-Kassaby et al. (2014) introduced the Randomized, Replicated, Staggered Clonal-Row (R2 SCR) seed orchard design, which combines randomization with row arrangement to permit equal, linear, and custom deployment options. Rows are staggered such that each parent is surrounded by four different parents, and the neighboring parents vary across replications. The most recent work of Lstibůrek secured the ONA algorithm, which is focused on maximizing panmixia but also allows for separating the ramets of the same clone. In addition, it can be combined with the original MI to account for related clones within advanced generations.

6.4 CROWN MANAGEMENT

Irrespective of the seed orchard origin - of whether seed orchard trees are seedlings or graftings – it must be kept in a manageable size by **crown pruning** that promotes adequate seed and pollen strobili development and facilitates easy and safe seed cone harvest. In Douglas-fir and Sitka spruce, higher values of selfing were found in lower than upper parts of the crown. Techniques for physical (girdling) and hormonal **cone induction** are also available. The latter most commonly involves a mixture of gibberellins known as GA 4/7 applied as a foliar spray, stem injection, or a combination of both. **Top grafting** is an important game-changing technique, which, by its inception in several breeding programs, saved a substantial amount of time and thus increased the genetic gain in these programs per unit of time. Top grafting involves grafting young scion material into the branches of existing, reproductively already mature trees, resulting in composites representing three different genotypes: a **rootstock. This interstock** was initially grafted onto the juvenile rootstock and a **top-graft**. The underlying principle of top--grafting is that fitness, physiological stability, and reproductive competence of the interstock are transferred to the top-graft, and the top-grafted scion carries the desirable genetic material. Although top grafting has been widely used for many years in fruit trees to widen the production of different fruit varieties (promoting cultivars), it was adopted in tree improvement programs to accelerate breeding by inducing early flowering in graftings and thus shortening intervals between generations, as well as infusing new selections into mature seed orchards. Although top grafting successfully met these objectives in loblolly pine, slash pine, and other species such as European and Japanese larch to a limited extent, its use to reduce selfing has been surprisingly neglected. Since selfing as the highest level of inbreeding constitutes a major obstruction to both quality and quantity of seed crops (see subsection 'Inbreeding'), then top grafting could also be viewed as a possible way of reducing it, especially for trees with larger, because pollen grains from one part of a tree fertilizing ovules on other parts of the same tree could effectively still be outcrossing. Results from a study of selfing in one top-grafted western red cedar seed orchard suggest that this technique may have promoted outcrossing and thus improved the overall genetic quality of the seed crop. The average selfing rate was estimated to be 7.3%, a value substantially smaller than all others reported for this species in natural and artificial populations. Unfortunately, direct evidence for this statement is unavailable because offspring from the top-grafted and control ramets were not directly compared.

6.5 SUPPLEMENTAL MASS POLLINATION, BLOOM DELAY, AND INDOOR SEED ORCHARDS

Pollen is a crucial component of all seed orchards as it contributes one-half of gametes to the generation of offspring, provided there is zero or negligible pollen contamination from background pollen sources. Seed orchardists have devised management techniques to intervene with natural pollination to lessen the effect of the above-mentioned deviations from expectations. These techniques also require scrutiny and evaluation, so their application should be weighed against the associated costs. While pollen is relatively easy to store and transport, it represents a powerful tool for seed orchard managers to manipulate the genetic composition of seed crops and specifically increase genetic gain. Apart from breeding activities where pollen is applied in small quantities on isolated cones to create controlled crosses among selected individuals, it is also commonly applied in larger quantities on unprotected receptive female strobili either for an entire orchard population or for a selected subset of parents or ramets. This latter technique, called pollen augmentation or **supplemental mass pollination** (SMP), was introduced to increase seed crop genetic quality, and it is widely utilized in seed orchards to improve yields, balance male reproductive success, introduce specific genotypes into the seed orchards and reduce selfing and pollen contamination. With the advent of molecular markers, SMP effectiveness could be evaluated in detail, and it was proven to facilitate mating among reproductively asynchronous parents and reduce pollen contamination while increasing seed set and outcrossing, particularly for early and late reproductive phenology classes. It has been proven that SMP effectiveness greatly depends on timing as well as the method of pollen application. Still, other factors such as pollen contamination rate, complementary management practices such as bloom delay, or actual environmental conditions at flowering may also influence its effectiveness. When individual strobili were pollinated, average success rates ranged between 66 and 84%, but in an operational study where whole trees were pollinated, the success declined to 7–26%. Mimicking SMP, honeybees were successfully introduced to improve seed production in some insect-pollinated genera, such as *Eucalyptus* and *Acacia,* as they lowered flower abortion and selfing and increased seed set per capsule. Similar benefits were observed after bee hives were situated in the wild cherry seed orchards.

Bloom delay is a method based on the work by (Silen & Keane, 1969) which fine water mist is applied in early spring to seed orchards to cool their immediate environment relative to oncoming spring conditions. The gradual evaporation of water ensures that heat-sum, a cue triggering actions pertaining to reproductive phenology, including reproductive bud development, is accumulated more slowly within seed orchards. As a result, treated trees are temporally isolated from exposure to background pollen contamination. This technique was proved to effectively delay and unify reproductive phenology, helping to increase panmixia within the studied orchards' populations by balancing male gametic contributions, as well as reducing pollen contamination and selfing. Another positive by-product reported for bloom delay was decreased insect infestation in orchard trees.

Indoor seed orchards are a costly but very effective solution for many of the problems mentioned before. The experimental use of greenhouse-based indoor seed orchards is well-known in southern Sweden.

6.6 SEED ORCHARD ESTABLISHMENT

Before establishing new seed orchards, finding sites with a warm local climate to stimulate flowering is important. Based on a detailed analysis of many seed orchards, the ideal location of a seed orchard can be determined. Before a new seed orchard is established, it is necessary to verify the local climate of the candidate locality to get any flowering. Topography is the most important factor. Cold air or fog should not remain in the seed orchard or run through the orchard. Therefore, we prefer a somewhat elevated position that reduces the cold air stream. A natural barrier should ensure protection against dominating winds from north or southwest. Good light conditions with open to sun radiation are strongly preferred, which points towards a slight southwest slope. Regarding soil conditions, light river sediments with a satisfactory fraction of fine minerals (25% fine sand or finer) are desirable with good drainage owing to the elevated location and the light sediment. Seed orchard soil should not be too nutritious, as this stimulates the overall growth of the graftings, often at the expense of fructification. On the other hand, any opposite extremes with poor soil, high soil acidity, or tendencies to build peat sediments must be avoided!

Conclusion / Takeaways

Seeds from existing seed orchards contain considerable genetic gain. The great weakness of conventional seed orchards worldwide is that the theoretically possible gains are not reached owing to pollen contamination and the variable pollen and seed production of the seed orchard clones. Differences in the time points for receptivity and pollen dispersal also contribute to deviations from theoretical expectations.

6.7 PROGENY TESTING AND MATING DESIGN

Progeny testing plays a crucial role in forest tree breeding, primarily for identifying parents with good general combining ability, as discussed in the chapter on quantitative genetics. This method, known as backward selection, involves selecting parents based on the performance of their progeny. Another key objective of progeny testing is estimating variance components, which is essential for determining heritability in broad and narrow senses. These estimates are vital for future breeding programs and predicting potential gains. Additionally, progeny trials provide a basis for forward selection, where the best trees from the best families are chosen for the next generation of the breeding population. There are three main types of mating designs used in progeny testing:

- **Diallel matings**
- **Factorial matings**
- **Nested matings**

In addition, **polycross** and **open pollination** may be used.

The various types of mating designs must have advantages and disadvantages. All artificial mating work is labor-intensive and thus expensive. It is crucial to clarify the objective of the mating work before deciding which mating design to use.

Diallel mating involves parents serving both as female and male. To obtain comprehensive information about the genetic quality of a set of trees, all possible (reciprocal) crosses among all parents should be carried out, such as 1x2, 2x1, 1x3, 3x1, and so on. This design provides the best estimates of **additive** and **non-additive effects**, as well as selfing and reciprocal effects. The progeny trial from such a complex mating design is also ideal for selecting the best individuals within the best families. A significant advantage of the complete diallel is that all families are represented.

However, the major disadvantage of the complete diallel is its impracticality when the number of parents is high. For example, testing 50 parental trees would require 50 x 49 = 2,450 crosses, excluding selfings. This makes the mating work and field trials excessively large and unrealistic for applied breeding. Additionally, large numbers of families require a large, homogeneous area of forest land, which is difficult to find. For instance, a complete diallel mating with 50 parents at a spacing of 2 x 2 meters would only allow for three plants per family in a 3-hectare field trial.

	1	2	3	4	5	6	7	8
1		x	x	x	x	x	x	x
2	x		x	x	x	x	x	x
3	x	x		x	x	x	x	x
4	x	x	x		x	x	x	x
5	x	x	x	x		x	x	x
6	x	x	x	x	x		x	x
7	x	x	x	x	x	x		x
8	x	x	x	x	x	x	x	

Fig. 6.4 Complete diallel mating scheme

To avoid such unrealistic test sizes, the number of families must be reduced, which can be achieved through **partial diallel matings**.

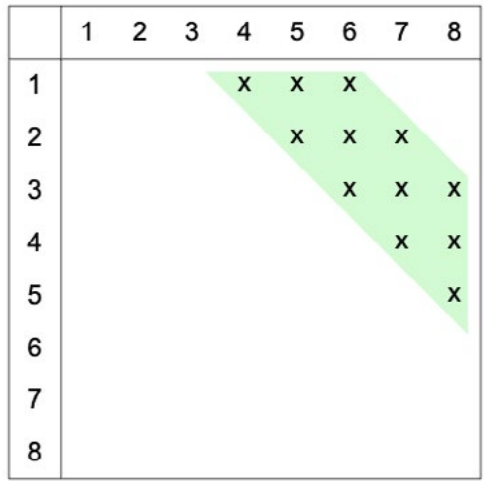


Fig. 6.5 Partial diallel mating scheme

After reducing a complete diallel mating scheme, the most generous mating design regarding the number of remaining families is the **half-diallel**. As the name implies, half of all possible crosses are carried out. This trade-off is typically achieved by excluding reciprocal crosses, assuming maternal effects can be neglected. A commonly used modification of the partial diallel also excludes selfings, resulting in an empty diagonal in the crossing matrix. This type of mating design strikes a good balance, allowing for achieving various objectives in progeny testing, such as identifying parents with good general combining ability, estimating variance components, and performing forward selection.

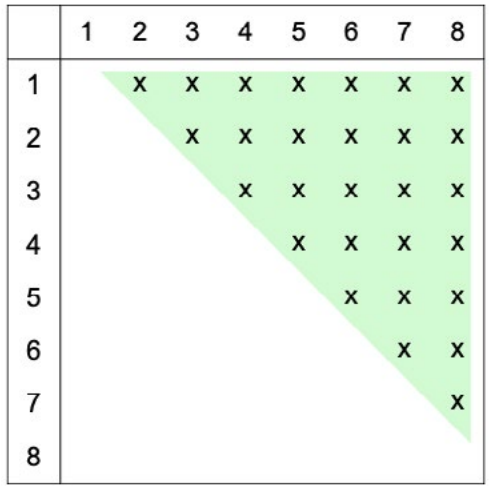


Fig. 6.6 Half-diallel mating scheme

In **factorial mating**, a parent serves either as a female or a male. When a factorial mating design involves a few male clones and numerous females, it is called **common tester** mating. The primary advantage of the common tester design is that it provides fairly accurate estimates of female general combining abilities. However, the number of unrelated families is low and does not exceed the number of males, resulting in an unbalanced design concerning the number of females and males. Historically, this was the first systematic mating design used globally in tree breeding programs.

	1	2	3	4
5	x	x	x	x
6	x	x	x	x
7	x	x	x	x
8	x	x	x	x
9	x	x	x	x
10	x	x	x	x
11	x	x	x	x
12	x	x	x	x

Fig. 6.7 Common tester mating scheme

In the early days of tree breeding, comparing the performance of plus tree progenies with ordinary seed lots was essential. However, systematic matings were nearly impossible due to erratic flowering in young seed orchards. Early tree breeders, therefore, had to rely on data from unsystematic matings. The greatest efficiency of factorial matings is achieved when there are equal numbers of females and males. Additionally, factorial matings can be simplified by reducing the number of matings, facilitating easier handling of progeny tests.

Disconnected half-diallels, groups of half-diallels without common clones, became popular worldwide around 1980, replacing the common tester mating design in many tree breeding programs. The major advantage of this mating design is that small half-diallels are easy to complete. Parents are selected based on flowering in a particular year, accommodating year-to-year fluctuations. Clones that do not flower one year may flower the next, allowing another half-diallel to be completed that year. Flowering variability has been a significant obstacle for certain species in completing mating designs involving multiple clones.

	1	2	3	4	5	6	7	8
1								
2	x							
3	x	x						
4	x	x	x					
5								
6					x			
7					x	x		
8					x	x	x	

Fig. 6.8 Disconnected half-diallel mating scheme

However, the major disadvantage of mating designs without connections between groups of progenies is that comparing the breeding values of parents from different groups is unreliable.

Single-pair mating involves each parent being mated with just one other parent. This design is optimized for parents with good breeding values to generate families for forward selection. However, the ability to estimate genetic variance components is limited in this case.

Polycross and **open pollination** are effective alternatives for estimating breeding values, though they do not provide non-additive estimates. In polycross, each parent is pollinated with a pollen mix typically consisting of many males. Open pollination involves harvesting seeds from trees without any artificial pollination. During the initial selection of plus trees, simultaneously collecting seeds allows for the early establishment of progeny trials, accelerating the breeding process. Because only one progeny represents each parent, the required trial area is significantly smaller than other mating designs. However, the number of trees per family should be larger in progeny testing using polycross or open pollination than in other designs.

The **nested mating design** involves grouping parents into a series of "nests," with ideally no fewer than 20 parents in each nest. Each female is mated with pollen mixes from each nest in its most complete form. While this may result in some selfing, it is generally considered negligible as selfed seedlings will be outcompeted by outcrossed seedlings. Estimates of parental general combining ability (GCA) are reliable if the pollen mix comprises 20-30 parents. There are also opportunities to modify the complete nested design into less labor-intensive versions.

6.7.1 BASIC PRINCIPLES OF PROGENY TESTING

Selecting an appropriate site for progeny trials is crucial, with breeders typically opting for homogenous ground to ensure consistent results. However, the initial establishment phase of progeny trials is highly sensitive. Planting shocks can affect plants randomly, with some experiencing severe impacts and others being less affected. Juvenile plants also face competition from weeds, which can randomly influence their early growth. Over time, the effect of these environmental factors diminishes, allowing genetic quality to play a more significant role in determining growth.

For tree species with long rotation, delaying the evaluation of progeny trials until the trees are fully mature is impractical. Some breeders argue that evaluating progeny trials after one-third of the rotation period is sufficient to rank parents based on growth. This one-third period represents many years, particularly for progeny trials in high-latitude regions. Evaluating parental growth after 15-20 years is unlikely to result in significant errors. The best indicator of long-term growth potential is often the growth increment observed during the evaluation period's last five or ten years.

As trees in a progeny trial mature, they increasingly compete for water, nutrients, and light resources. Intense competition can lead to a more pronounced differentiation among families, facilitating the selection of the best parents. However, if the competition is excessively strong, it may overestimate the potential genetic gain from the material under such conditions. To determine the optimal timing for selection, breeders can estimate **age-age correlations**—correlations of the same trait at different ages in the same tree.

For instance, in a Swedish progeny trial of *Pinus sylvestris* (Scots pine), high age-age correlations were found for tracheid length between ages 11 and 31 and for wood density between ages 8-11 and 28-33. The genetic gain per year for these traits was two to three times higher when selection occurred at age 11 compared to age 31 or 33, suggesting that the optimal selection age might be even lower than 11 years. Consequently, early testing for these traits can enhance the efficiency of the *Pinus sylvestris* breeding program.

In summary, early evaluation and selection in progeny trials, considering age-age correlations and environmental effects, can significantly improve the effectiveness of tree breeding programs, especially for species with long rotation times.

6.7.2 EARLY TESTS

Early testing in tree breeding programs involves evaluating young plants, such as seedlings or saplings, to predict their future growth and performance as mature trees. This method expedites the breeding cycle, allowing breeders to decide which genotypes to select and propagate faster. Significant expectations have been placed on the ability to predict future growth performance using seedlings or even seeds (BOX 6.1). The advantage of early testing lies in the accelerated completion of the breeding cycle compared to traditional long-term field testing. However, a major challenge with early testing is identifying juvenile traits or combinations of attributes that strongly correlate with

valuable adult characteristics. Up to the end of the 20th century, early tests for growth yielded inconsistent results, with some studies showing strong juvenile-mature (J-M) correlations and others showing none. Analyzing the reasons for weak J-M genetic correlations is essential for improving early testing methods.

Several factors can contribute to weak J-M genetic correlations:

1. **Human Error**: Mislabeling or mixing scions, grafts, and seed lots can lead to incorrect identities. Pollen contamination, especially in wind-pollinated species, is difficult to prevent, potentially compromising experimental integrity.
2. **Additive Variance**: The additive genetic variance may be low at the juvenile or mature stage, affecting the correlation.
3. **Allelic Differences**: Different alleles may regulate traits at juvenile and mature stages. For instance, in spruce, free growth observed in juveniles may disappear as the trees age.
4. **Genotypic Differences**: Different genotypes might produce the same phenotype at juvenile and mature stages. A genotype that causes fast growth in juveniles may differ from one that causes rapid growth in mature trees.
5. **Environmental Differences**: Growth conditions in growth chambers, greenhouses, or nurseries differ significantly from field conditions, leading to genotype x environment interactions.
6. **Non-Genetic Effects**: During the establishment phase, non-genetic factors may dominate. Strong J-M genetic correlations are more likely to emerge once genetic effects become more prominent in field trials.
7. **Field Trial Imperfections**: Field trials may not fully reflect genetic potential due to experimental design flaws or other sources of imprecision.
8. **Complex Traits**: Growth is a complex trait, and individual components may not show strong J-M correlations. Using an index to weigh different growth components could help overcome this issue.

Leveraging results from existing field trials is a practical approach to developing effective early tests. This is feasible when the parents of these trials are available in seed orchards or clonal archives. Crossings can be repeated, or seeds obtained from open pollination, allowing the study of young siblings of mature field trial trees at the juvenile stage. Such early tests, known as **retrospective tests**, utilize existing genetic information to improve early selection accuracy.

Methods to Improve Early Testing

Retrospective Testing: Utilizing data from existing field trials, breeders can evaluate the early performance of young siblings of mature trees. This approach relies on established genetic information to validate early test results.

Index Selection: Combining multiple traits into an index can improve the predictive power of early tests. By weighting various components of growth, breeders can better estimate overall performance.

Genomic Selection: Advances in molecular genetics allow for the use of DNA markers to predict traits. Genomic selection can enhance the accuracy of early tests by identifying genetic markers associated with desirable traits.

Age-Age Correlations: Studying correlations between the same trait at different ages helps determine the optimal timing for early selection. High age-age correlations indicate that early measurements are reliable predictors of future performance.

BOX 6.1
Case Studies and Examples
Eucalyptus spp.: Early testing in eucalyptus has shown promising results, with strong correlations between juvenile and mature growth traits, enabling faster breeding cycles for pulp and paper production (Pinto et al., 2014).
Loblolly Pine (*Pinus taeda*): Genomic selection has been integrated into early testing programs for loblolly pine, improving the accuracy of early predictions and accelerating breeding efforts for timber production (Resende Jr et al., 2012).

Future Directions

Integration of Genomic Tools: Continued advancements in genomic technologies will enhance the precision and efficiency of early testing.

Refinement of Trait Indices: Developing more sophisticated indices incorporating a broader range of traits and environmental interactions will improve early selection accuracy.

Enhanced Phenotyping: Advanced phenotyping methods, such as remote sensing and image analysis, can provide detailed and accurate measurements of early traits.

Early testing is a crucial component of modern tree breeding programs, offering significant advantages in terms of time and cost efficiency. While challenges remain, ongoing research and technological advancements continue to improve the reliability and accuracy of early selection methods.

6.8 CLONAL FORESTRY

Clonal forestry involves using vegetative propagation to produce genetically identical trees, or clones, from a select group of superior genotypes. This practice has a long history, with early examples in China for Chinese fir and Japan for sugi (genus *Cryptomeria*), dating back over a millennium. In modern forestry, clonal propagation is widely used in tropical and subtropical hardwoods like eucalypts and poplars due to their ease of rooting and short rotations. However, its application in conifers has been more challenging due to issues related to maturation and propagation costs.

Improving desirable traits such as growth rate, wood quality, and disease resistance measures genetic gains in forestry. Clonal forestry captures additive and non-additive genetic variances, leading to higher potential gains than traditional family forestry. The estimated genetic gain from clonal forestry over family forestry ranges from 5% to 25% in conifers and 25% to 50% in hardwoods like eucalypts.

Genetic gains in tree improvement are achieved through the advancement of generations and the intensity of selection. Higher genetic gain is generally associated with increased selection intensity and the propagation method used within the same generation. The progression of genetic gain typically follows these stages:

Open-Pollinated Seed Orchards (Unrogued): Seeds are collected from seed orchards without any selection based on progeny testing.

Open-Pollinated Seed Orchards (Rogued): Seeds are collected from seed orchards that have been selectively thinned (rogued) based on progeny testing to remove inferior genotypes.

Control-Pollinated (CP) Full-Sib Families: Seeds are collected from CP families in seed orchards or genetic archives where parent trees have been selected based on genetic tests. This method is referred to as "CP family forestry." CP full-sib family seeds are often used to create donor-plant hedges, which are then vegetatively propagated to produce improved material. This method is known as "family forestry with vegetative propagation."

Clonal Forestry: Propagules are derived from single or multiple well-tested clones created through vegetative propagation. This method selects clones based on their performance in clonal trials, capturing both additive and non-additive genetic variances, thus offering the highest potential genetic gains.

Each stage represents increased genetic gain due to more intensive selection and advanced propagation techniques, culminating in the highest gains achieved through clonal forestry.

6.8.1 RISK ASSESSMENT

There are three primary risks associated with clonal forestry: **(1) Plantation Failure:** The risk of a clonal plantation failing due to biotic or abiotic factors is a significant concern. The failure risk is influenced by genetic uniformity, where a single clone can succumb entirely to specific threats; **(2) Loss of Genetic Diversity:** Clonal forestry can reduce genetic diversity at both the forest and landscape levels, making populations more susceptible to pests and diseases over time; **(3) Propagation Success Rate:** The success of achieving the theoretical genetic gain depends on the propagation success rate. Lower propagation success can significantly reduce the expected gains.

Risk mitigation strategies involve using sufficient clones to balance genetic gain and diversity. Studies suggest that 5 to 30 different clones can provide adequate risk mitigation, with an optimum level of around 18 to safeguard against unknown future threats.

6.82 REALIZED GAINS IN CLONAL FORESTRY

Realized genetic gains from clonal forestry can be substantial. For example, in Radiata Pine in New Zealand and Australia, clonal forestry has shown a 15-20% increase in volume growth and improvements in wood properties such as stiffness and density. In Eucalypts in Brazil, clonal forestry has led to productivity gains from 25-30 m^3/ha/year to 35-45 m^3/ha/year, with some plantations achieving over 60 m^3/ha/year.

Genomic Selection and Clonal Forestry

Genomic selection (GS) is an emerging technology that uses genome-wide markers to select superior genotypes, potentially accelerating the breeding cycle. When combined with somatic embryogenesis (SE), GS can significantly enhance the efficiency and speed of developing clonal forestry programs in some species. This combination allows for early selection of superior genotypes and reduces the need for long-term clonal testing.

Conclusion and Recommendations

Clonal forestry offers substantial genetic gains and economic benefits compared to traditional family forestry. However, careful management is required to mitigate risks associated with reduced genetic diversity and propagation success rates. Recommendations for implementing clonal forestry include conducting robust yield trials to validate genetic gains and economic benefits; estimating propagation costs by comparing different propagation methods to ensure financial viability; maintaining genetic diversity by using a sufficient number of clones to balance genetic gain and diversity; combining GS and SE to accelerate the development and implementation of clonal forestry, especially in conifers. Following these recommendations, clonal forestry can enhance forest productivity and sustainability, provided the associated risks are effectively managed.

6.9 DEPLOYMENT OF BRED MATERIAL IN TREE IMPROVEMENT

In tree improvement programs, the deployment of bred material refers to the strategic planting and utilization of genetically enhanced tree varieties. The primary goal is to achieve specific objectives such as increased yield, improved resistance to pests and diseases, or better adaptation to environmental conditions. The methods of deployment vary depending on the species, the objectives of the program, and operational constraints.

Seedling Deployment

One common method is using seedlings grown from seeds collected from selected trees in seed orchards. These can be either open-pollinated seedlings, which are cost-effective but genetically variable, or controlled-pollinated seedlings, where seeds are produced through controlled pollination between selected parent trees. Controlled-pollinated seedlings offer better control over genetics but are more labor-intensive and expensive to produce.

Clonal Deployment

Clonal deployment involves planting genetically identical copies, or clones, of superior trees. This method is used when uniformity and replicating desirable traits are essential, such as in high-value species or environments where performance predictability is crucial. Clonal mixtures, which involve planting a mix of different clones, are sometimes used to increase genetic diversity while utilizing clonal propagation's benefits.

Family-Based Deployment

Another approach is family-based deployment, where seedlings are derived from selected family lines. These can be half-sib families originating from open-pollinated seeds collected from selected mother trees or full-sib families produced through controlled crosses between selected pairs of parent trees. Family blocks, where different families are planted in separate blocks, allow for the evaluation of family performance under various environmental conditions and inform further selection efforts.

Enrichment Planting

In some cases, improved material is planted in an existing plantation to enhance the genetic quality or introduce new traits, a method known as enrichment planting. This approach is often used in natural forests or established plantations to improve the overall genetic base.

Mixed Deployment and In Situ Deployment

Mixed deployment involves combining improved and local varieties to balance the benefits of genetic improvement with local adaptation. In situ deployment refers to introducing improved genetic material into natural populations to enhance the genetic base and improve resilience without establishing new plantations. This method is particularly important in conservation-oriented programs.

Large-Scale Commercial and Agroforestry Deployment

For commercial objectives, large-scale plantations of improved material are established to maximize productivity and economic returns. In agroforestry systems, improved tree species are integrated with crops to enhance productivity, improve soil quality, and offer additional products like fruits or timber.

Genetic Conservation

Finally, some deployment strategies focus on conservation plantings to preserve the genetic diversity of tree species. These efforts involve planting a diverse range of genotypes in protected environments to ensure the sustainability of forest resources.

Deploying bred material in tree improvement is a critical step that requires careful planning and consideration of genetic diversity, site selection, environmental impact, and socio-economic factors. Proper deployment strategies ensure the success of tree improvement programs and contribute to the sustainability of forestry resources.

6.9.1 SEED PRODUCTION AREAS (SEED STANDS) AND DIRECTED SEED COLLECTIONS

In tree improvement programs, Seed Production Areas, also known as seed stands, are crucial for producing well-adapted seeds to specific environments. Seed stands are usually preferred in species where the seed orchard establishment is problematic. This might be the case for several broadleaved species, such as Silver Beech or Sessile Oak, in Central Europe due to grafting issues. These areas are chosen based on their high genetic quality and suitability to local conditions, ensuring that seeds collected from them will thrive when planted in similar regions. The trees within these stands are typically selected for their superior traits, such as faster growth, resistance to pests, or overall health. Once selected, these areas are carefully managed to maximize seed production and maintain the genetic integrity of the trees.

Seed stand management practices include thinning operations to reduce competition among trees, pest and disease control, and protection from environmental disturbances like fire. The timing of seed collection is carefully planned to ensure seeds are mature and viable, and post-collection processing and storage are conducted to preserve seed quality until planting.

Directed Seed Collections complement seed stands by focusing on the selective gathering of seeds from individual trees with desirable characteristics. This method ensures that only the best genetic material is used for future planting or breeding. Trees are chosen based on visible traits, such as height, form, and resistance to stress, and sometimes through genetic testing for specific traits.

The collection process involves carefully gathering seeds without damaging them, often requiring specialized equipment or techniques. Detailed records of the parent tree and collection conditions are kept, aiding in tracking the performance of seedlings and informing future breeding decisions.

Both Seed Production Areas and Directed Seed Collections are integral to improving tree populations' genetic quality and adaptability. Seed zones provide a broader, regionally adapted seed source, while Directed Seed Collections ensure that specific desirable traits are captured and propagated. Together, these methods support the goals of reforestation, conservation, and the overall success of tree improvement programs.

6.9.2 MANAGEMENT OF GENETIC DIVERSITY DURING SEED COLLECTING, PROCESSING, AND IN THE NURSERIES

Nurseries play a significant role in managing genetic diversity in tree improvement and reforestation programs. How seeds are handled, seedlings are grown, and plants are selected and distributed can either maintain or reduce the genetic diversity of the established tree populations. Here's how nursery practices influence genetic diversity:

Seed Source and Collection

The genetic diversity of the seedlings produced in a nursery begins with the diversity of the seeds collected. If seeds are sourced from a broad range of parent trees across different locations (within a seed zone or from multiple seed zones), the resulting seedlings will have a wider genetic base. Conversely, if seeds are collected from a limited number of trees or from a single source, the genetic diversity will be reduced.

Nurseries that use a diverse seed source help preserve genetic variability, which is crucial for the adaptability and resilience of the planted populations to changing environmental conditions and future threats such as pests, diseases, or climate change.

Seed Handling and Storage

Proper handling and storage of seeds are essential to maintain their viability and genetic integrity. Poor storage conditions, such as exposure to moisture, high temperatures, or pests, can reduce seed viability and lead to the loss of specific genotypes, thus narrowing genetic diversity.

Nurseries that employ the best seed handling and storage practices can better preserve the full spectrum of genetic diversity in the original seed lot.

Germination and Seedling Production

The conditions under which seeds are germinated and seedlings are grown can influence genetic diversity. For example, uniform conditions in a nursery may favor certain genotypes over others, potentially leading to a reduction in genetic diversity as only the most suited genotypes thrive. Additionally, overcrowding or competition in nursery beds may cause some seedlings to die off, reducing diversity.

By providing a range of conditions that mimic natural environments and ensuring adequate space and care for all seedlings, nurseries can help maintain genetic diversity among the plants they produce.

Selection and Culling

During the nursery stage, a selection process often occurs where weaker or less desirable seedlings are culled. If this selection is based solely on observable traits without considering maintaining genetic diversity, it can significantly reduce the genetic pool. The one often cited example would be favoring the pioneering ecotypes that exhibit fast early growth. On the other hand, the slower-growing ecotypes are removed during culling.

Careful selection practices that aim to retain a broad range of genetic traits help ensure that the planted populations will be genetically diverse and resilient.

Vegetative Propagation

In nurseries that use vegetative propagation (such as cuttings or tissue culture) to produce trees, the genetic diversity is inherently lower because all the resulting plants are clones of the parent plant. While this can be beneficial for producing uniform, high-quality trees, it can also limit genetic diversity.

Nurseries relying heavily on clonal propagation must balance this with other practices that promote genetic diversity, such as maintaining a diverse breeding population or using seeds from genetically diverse sources for some of their production.

Distribution and out-planting

Finally, how seedlings are distributed and outplanted can affect genetic diversity. For example, if seedlings from a single genetic source are planted across a large area, the resulting forest will have limited genetic diversity. Conversely, mixing seedlings from different genetic sources within planting sites can enhance diversity.

Ensuring that seedlings from various genetic backgrounds are planted together can increase the genetic diversity of the new forest stands, enhancing their adaptability and resilience.

In summary, nursery practice profoundly impacts the genetic diversity of tree populations. From seed sourcing and handling to seedling production and distribution, each step must be carefully managed to preserve and enhance genetic diversity. This is crucial for the long-term success of reforestation and tree improvement efforts, as genetically diverse populations are more likely to adapt to changing conditions and resist environmental stresses. By adopting practices that prioritize genetic diversity, nurseries can contribute significantly to the sustainability and resilience of future forests.

6.10 METHODS OF VEGETATIVE PROPAGATION

Vegetative propagation refers to the reproduction of plants using parts of the parent plant, such as leaves, stems, or roots, to produce genetically identical offspring or clones. This method is widely used in forestry and horticulture to reproduce plants with desirable traits. Vegetative propagation offers many diverse uses in operational forestry: (1) preservation of genotypes in clone banks/archives; (2) multiplication of desired genotypes for seed orchards and breeding; (3) evaluation of genotypes and their specific interaction with the environment through **clonal testing;** and (4) maximizing genetic gains through its use in deployment (regeneration of clones in operational plantations. Here are the methods of vegetative propagation, including stool beds, rooted cuttings, and somatic embryogenesis:

1. Stool Beds

Stool beds, or hedges, are a traditional method of vegetative propagation, particularly useful for producing rooted shoots, also known as "stools" or "suckers." This technique is commonly used for propagating woody plants, especially those that naturally produce shoots from the base of the stem.

Process:

Establishment: A parent plant, often called a stool plant, is planted in a prepared bed.

Coppicing: After the parent plant is established, it is cut back close to the ground to encourage the growth of multiple shoots from the base.

Mounding: As these shoots grow, soil or other rooting media are heaped around their bases. This practice encourages the formation of roots along the buried portions of the shoots.

Harvesting: After the shoots have developed sufficient roots, they are cut from the parent plant and can be transplanted as independent plants.

Advantages:

- This method is relatively low-cost and simple to execute.
- It is particularly effective for species that naturally produce basal shoots.

Applications:

- Stool beds are widely used in the propagation of fruit trees, certain types of shrubs, and some forest species like poplars and willows. In conifers, stool beds are commonly used to reproduce Norway spruce until a certain age.

2. Rooted Cuttings

Rooted cuttings are one of the most common methods of vegetative propagation, involving rooting a severed section of a plant, such as a stem, leaf, or root. This method is versatile and can be used for a wide variety of plant species.

A specific and not often cited method is rooting needle fascicles in pine. Normal shoot development could be problematic, but chemicals and hormones can fix these initial issues. The original idea appealed to tree breeders for the opportunity to get many propagules from a single tree (Zobel & Talbert, 1984).

Process:

Cutting Preparation: A section of a healthy plant, usually a stem or branch, is cut. The cutting typically includes a few leaves and at least one node (the point where leaves are attached).

Rooting Medium: The cuttings are placed in a suitable rooting medium, including soil, sand, perlite, or a mixture. The medium should be well-draining and moist.

Environmental Control: Cuttings are often kept in a controlled environment with high humidity and appropriate light levels to promote rooting. Techniques such as misting or covering with a plastic dome can help maintain humidity.

Root Formation: Over time, roots develop from the cut end of the cutting. Once sufficient roots have formed, the new plant can be transplanted into a pot or outdoor setting.

Advantages:

- Allows for the rapid multiplication of plants with desirable traits.
- Produces genetically identical plants to the parent, ensuring uniformity in growth and characteristics.

Applications:

- Rooted cuttings are widely used in forestry, horticulture, and agriculture, especially for propagating ornamental plants, fruit trees, and forestry species such as pines, spruces, and eucalypts.

3. Somatic Embryogenesis

Somatic embryogenesis is an advanced vegetative propagation method involving the production of embryos from somatic (non-reproductive) cells. These embryos can then develop into complete plants, later called **emblings**. While this method is not as commonly used as stool beds or rooted cuttings, it is important in certain contexts, particularly in plant breeding and genetic research. This method is used in the propagation of certain forest species and in the commercial production of crops like oil palm, coffee, and some ornamentals. It is also a valuable tool in genetic engineering and conservation biology.

Process:

Cell Culture: Somatic cells, often from leaf, stem, or root tissue, are cultured in a laboratory under sterile conditions.

Embryo Induction: Under the right conditions, these cells can be induced to form embryos, similar to seeds in their ability to grow into new plants.

Embryo Development: The embryos are nurtured in a controlled environment, developing into small plants (plantlets).

Acclimatization and Planting: The plantlets are then gradually acclimatized to normal growing conditions before being transplanted to soil or other growing media.

Advantages: Somatic embryogenesis allows for the mass propagation of plants with specific genetic traits, which is useful in breeding programs. It can produce genetically uniform plant material for large-scale planting programs.

Disadvantages: The disadvantages of somatic embryogenesis in trees primarily revolve around the challenges of acclimating embryos to natural conditions and the risk of introducing genetic variability through **soma-clonal variation**. These factors can complicate the use of this technique in forestry and conservation, where reliability, consistency, and genetic stability are crucial.

Somatic embryogenesis involves growing embryos in a controlled, artificial environment before transferring them to natural conditions. This transition phase is critical and delicate. Many tree species or entire genera can be particularly sensitive during this stage, leading to high failure rates.

The embryos, accustomed to the controlled laboratory conditions, might not cope well with environmental stressors like changes in temperature, light intensity, humidity, or soil conditions. This stress can result in poor survival rates or suboptimal growth once the embryos are transplanted.

Certain tree species may have specific requirements during acclimation that are difficult to replicate or predict, leading to inconsistent results. This challenge can make somatic embryogenesis less reliable or efficient for commercial or conservation purposes for those species.

6.10.1 SOMA-CLONAL VARIATION

Soma-clonal variation refers to genetic changes that occur during somatic embryogenesis. These variations can lead to differences in the genetic makeup of the cloned plants compared to the original donor plant. This variation can result in trees with undesirable or unpredictable traits, such as altered growth patterns, reduced disease resistance, or changes in wood quality. This unpredictability is a significant disadvantage when uniformity and consistency are desired, such as in commercial forestry operations.

Clonal propagation through somatic embryogenesis aims to produce genetically identical plants. Soma-clonal variation undermines this objective by introducing genetic diversity, which can be problematic when the goal is to replicate specific, desirable characteristics of the parent tree.

There is also concern about the long-term stability of these variations. Some traits might not manifest immediately but could appear in later stages of growth or under certain environmental conditions, complicating the evaluation and management of the propagated trees.

6.10.2 SPECIFIC PROBLEMS OF VEGETATIVELY PROPAGATED CULTURES

Several terms, such as maturation, cyclophysis, ontogenetic aging, and phase change, describe woody plants' morphological, anatomical, and physiological changes as they age (Greenwood and Hutchinson, 1993). These changes are linked to differential gene expression at various stages of development, highlighting the importance of gene regulation. Common age-related changes in most tree species include (1) reduced growth rates, (2) decreased branchiness, (3) the onset of flowering and seed production, (4) a decline in rooting ability, and (5) slower growth of cuttings obtained from older trees.

The last two changes are particularly significant in producing rooted cuttings in operational programs. While vegetative propagation is generally easier with juvenile material, rooting becomes progressively more challenging as trees age. The rate of decline in rooting ability differs among species, with some experiencing a decrease as early as one year of age. Additionally, cuttings from older trees often retain maturity characteristics, such as slower growth. For this reason, large-scale rooted-cutting programs for conifers and most angiosperms rely on juvenile material.

Hedging practices, such as severe shearing, topping, and pruning, are commonly employed to delay aging and maintain rooting ability in donor plants. Hedging can extend acceptable levels of rooting ability and the field performance of rooted cuttings for several years longer than non-hedged trees. Effective management of donor plants, including appropriate timing and severity of shearing and proper nutrition, is crucial for the success of large-scale rooted-cutting programs. However, even with appropriate management, the rooting ability and performance of cuttings decline as hedges age. To address this, hedges must be periodically replaced by producing new control--pollinated seeds from superior full-sib families to establish new hedges.

The term **C-effects** includes several environmental, developmental, and atypical genetic effects common to a group of relatives and cause the differences among groups of relatives to be larger than predicted by the classical genetic theory described in Chapter 4 on quantitative genetics.

In the broadest sense, types of C-effects include:

(1) **Topophysis**, positional effects associated with branches located in different parts of the crown such that vegetative propagules taken from the lower and central portions of the tree crown generally possess more juvenile characteristics than those from the upper or peripheral portions.
(2) **Cyclophysis**, maturation effects as described above in this chapter.
(3) **Periphysis**, environmental effects that have caused tissue preconditioning such that donor plants conditioned in different environments produce propagules with distinct characteristics.
(6) **Epigenetic effects**, ex. DNA methylation is altered by the season or different climate of origin.
(4) **Maternal effects**, such that members of the same maternal family share common attributes (e.g., due to seed size differences among female parents).
(5) **Non-nuclear genetic effects**, such as caused by inheritance of mitochondrial and chloroplast DNA.

C-effects can influence the performance of families and clones, often leading to outcomes that deviate from predictions based on classical genetic theory. In families, the last two types of C-effects create greater differences. In comparison, the first three types affect vegetative propagules, such as those produced through clonal forestry or VM family forestry. Biologically, these effects manifest as follows: (1) Propagules (e.g., rooted cuttings) from the same clone may vary in performance due to differences in the environmental and maturation state of the donor plants or the crown position from which cuttings were taken; and (2) distinct clones may exhibit varied responses to the propagation system, resulting in apparent genetic differences partially influenced by C-effects.

C-effects have two major implications for operational programs. Firstly, these differences are inconsistent across different propagation systems or donor plant conditions, as each unique set of environmental or maturation factors introduces a new array of C-effects that impact propagation and subsequent performance. For instance, if a superior clone's hedges age more quickly than others, its rooted cuttings may perform exceptionally when young but lose their growth advantage as the hedges age. Consequently, clones or VM families may change rank depending on how they respond to these varying conditions. To address this, testing and ranking clones or VM families under propagation conditions and donor plant states that align with those used in reforestation operations is essential.

Secondly, C-effects can lead to overestimating heritability and genetic gains because they exaggerate the differences observed among clones or families beyond what classical genetic theory predicts. For example, suppose certain clones respond better or worse to the propagation system (e.g., through rooting or performance). In that case, these clones may appear artificially superior or inferior due to both genetic and C-effect factors. This inflation results in upwardly biased broad-sense heritability estimates and predicted genetic gain when C-effects are significant.

6.10.3 SUMMARY

Stool beds and rooted cuttings are widely used and relatively simple methods of vegetative propagation, suitable for many trees and plants. Stool beds are particularly effective for species that naturally produce basal shoots, while rooted cuttings offer versatility and rapid multiplication for many species. Though more complex, somatic embryogenesis provides a powerful tool for advanced breeding and large-scale propagation, particularly in species where other methods may be less effective. Each method has its specific applications, advantages, and considerations, contributing to the overall strategy of vegetative propagation in forestry and horticulture.

6.11 REFERENCES

Bell, G. D., & Fletcher, A. M. (1978). Computer organised orchard layouts (COOL) based on the permutated neighbourhood design concept. *Silvae Genet*, *27*(6), 223–225.

Burczyk, J., Adams, W. T., & Shimizu, J. Y. (1996). Mating patterns and pollen dispersal in a natural knobcone pine (Pinus attenuate Lemmon.) stand. *Heredity*, *77*(3), 251–260.

Burczyk, J., & Prat, D. (1997). Male reproductive success in Pseudotsuga menziesii (Mirb.) Franco: The effects of spatial structure and flowering characteristics. *Heredity*, *79*(6), 638–647.

El-Kassaby, Y. A., Fayed, M., Klápště, J., & Lstib\uurek, M. (2014). Randomized, replicated, staggered clonal-row (R 2 SCR) seed orchard design. *Tree Genetics & Genomes*, *10*, 555–563.

Eriksson, G., Ekberg, I., & Clapham, D. (2008). *An Introduction to Forest Genetics*. https://res.slu.se/id/publ/22529

Funda, T., & El-Kassaby, Y. A. (2012). Seed orchard genetics. *CABI Reviews*, 1–23. https://doi.org/10.1079/PAVSNNR20127013

Funda, T., Chen, C. C., Liewlaksaneeyanawin, C., Kenawy, A. M., & El-Kassaby, Y. A. (2008). Pedigree and mating system analyses in a western larch (Larix occidentalis Nutt.) experimental population. *Annals of Forest Science*, *65*(7), 1.

Lindgren, D., Gea, L. D., & Jefferson, P. A. (1997). *Status number for measuring genetic diversity*. https://www.cabidigitallibrary.org/doi/full/10.5555/19981601712

Lindgren, D., Jefferson, P. A., & Gea, L. D. (1995). Status number-a measure of genetic diversity. *Evolution of Breeding Strategies for Conifers from the Pacific North West Joint Meeting of the IUFRO Working Parties S*, *20205*, 06.

Lindgren, D., & Matheson, A. C. (1986). An algorithm for increasing the genetic quality of seed from seed orchards by using the better clones in higher proportions. *Silvae Genet*, *35*(5–6), 173–177.

Lindgren, D., & Prescher, F. (2005). Optimal Clone Number for Seed Orchards with Tested Clones. *Silvae Genetica*, *54*(1–6), 80–92. https://doi.org/10.1515/sg-2005-0013

Lstibůrek, M., & El-Kassaby, Y. A. (2010). Minimum-Inbreeding Seed Orchard Design. *Forest Science*, *56*(6), 603–608. https://doi.org/10.1093/forestscience/56.6.603

Lstibůrek, M., Schueler, S., El-Kassaby, Y. A., Hodge, G. R., Stejskal, J., Korecký, J., Škorpík, P., Konrad, H., & Geburek, T. (2020). In Situ Genetic Evaluation of European Larch Across Climatic Regions Using Marker-Based Pedigree Reconstruction. *Frontiers in Genetics*, *11*, 28. https://doi.org/10.3389/fgene.2020.00028

Pinto, D. S., Resende, R. T., Mesquita, A. G. G., Rosado, A. M., & Cruz, C. D. (2014). *Early selection in tests for growth traits of Eucalyptus urophylla clones test*. https://www.cabidigitallibrary.org/doi/full/10.5555/20143309403

Resende Jr, M. F. R., Munoz, P., Resende, M. D., Garrick, D. J., Fernando, R. L., Davis, J. M., Jokela, E. J., Martin, T. A., Peter, G. F., & Kirst, M. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.). *Genetics*, *190*(4), 1503–1510.

Silen, R. R., & Keane, G. (1969). *Cooling a Douglas-fir seed orchard to avoid pollen contamination* (Roč. 101). Pacific Northwest Forest and Range Experiment Station, USDA Forest Service. https://books.google.com/books?hl=cs&lr=&id=igEKp3D_ht0C&oi=fnd&pg=PA5&dq=Silen+RR,+Keane+G.+Cooling+a+Douglas-fir+-seed+orchard+to+avoid+pollen+contamination.+U.S.D.A.+Forest+Service+Research+Note+PNW-101%3B+1969.&ots=3HF2FZZvf_&sig=jbnQhopZSxY5UrIBMwi6l2RNtjE

White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI.

Zobel, B., & Talbert, J. (1984). *Applied forest tree improvement*. Wiley.

7. INTRODUCTION TO GENOMICS OF FOREST TREES

Genomics studies the structure, function, evolution, and mapping of genomes. Genomics extends some of the traditional sub-disciplines of genetics, such as transmission and molecular genetics, to the entire genome. Genomics is a science that depends highly upon technologies that allow rapid analysis of hundreds or thousands of genes in many individuals simultaneously, such as by automated DNA sequencing. The advent of high-throughput sequencing technologies, such as Next-Generation Sequencing (NGS), has enabled rapid and detailed analysis of hundreds to thousands of genes. Genomics encompasses various fields, including structural genomics, functional genomics, and comparative genomics, each focusing on different aspects of genome analysis.

Key concepts tackled by modern genomic methods are, for example, pleiotropy - a single gene influences multiple, seemingly unrelated phenotypic traits.; epistasis - the effect of one gene depends on the presence of one or more 'modifier genes'; and Gene-Environment Interaction (*GxE*) when different genotypes respond differently to changing environmental conditions. Genomics has recently been divided into several branches:

Structural Genomics

Structural genomics aims to map all genes within a genome and place them on chromosomes. This is usually achieved by DNA sequencing of individual genes or entire genomes and by genetic or physical mapping, which involves assembling sequences into contiguous sequences (contigs) and further into scaffolds, larger assemblies that may include gaps where the sequence is unknown.

Functional Genomics

Functional genomics seeks to understand the roles and interactions of genes and how they influence phenotypic traits. This includes the annotation of genomes, where genes and their functions are identified based on homology to known sequences. Mapping all the genes is only an initial step in genomics, as the goal is to understand the function of all genes and their interactions. Techniques such as microarray analysis are used to study the expression patterns of genes. Eventually, all genes in all biochemical pathways will be identified, and the interactions between these genes and gene products will be discussed. Functional genomics also seeks to determine the relationship between the vast allelic variation in genes and the array of different phenotypes within populations.

Comparative Genomics

Comparative genomics involves comparing the genomic sequences of different species to understand the structure and function of genes and their evolutionary relationships. In forest trees, model organisms such as *Populus* (poplars), *Pinus* (pines), and *Eucalyptus* are often studied due to their economic and ecological significance following the development of comparative genetic maps based on orthologous genetic markers. Complete genome sequencing in Poplar has greatly enhanced comparative genomic analysis in these taxa.

Genome Assembly and Annotation

Genome assembly involves reconstructing the original sequence of DNA from short reads produced by sequencing technologies. Annotation identifies gene locations and predicts their functions, often relying on comparisons with previously annotated genomes.

7.1 FOREST TREE GENOMICS

The study of forest tree genomics reveals significant differences between angiosperms (e.g., *Populus*) and gymnosperms (e.g., *Pinus*). These differences include reproductive strategies, genome size, and the timing of reproductive maturity, which influence their genomic research approaches and applications.

One approach is to determine the DNA sequence of the entire genome and infer genes from the DNA sequence. This approach has been applied to *Populus* but is currently not feasible in conifers because of their large genome sizes. An alternative approach is determining the DNA sequences for the gene-coding regions. This can be accomplished by sequencing cDNA derived from the mRNA of genes expressed during the experiment.

The construction of genetic linkage maps is another integral component of genomics. Genetic maps show the position of genes relative to one another on chromosomes and are valuable for understanding genome organization and evolution. Maps are extremely useful tools for identifying genes controlling interesting phenotypes. Qualitatively inherited traits, such as disease resistance genes, can be located on maps and then cloned based on their map position. The map positions of the individual genes controlling quantitatively inherited traits, called QTLs, can also be determined from analyses using genetic maps. QTLs for various growth, wood quality, and other economic traits have been identified. Knowledge of the number and size of effects of QTLs controlling a quantitative trait can assist tree breeders.

Role of genomics in studying genetic variation

In modern days, genetic variation in forest trees is studied by evaluating traits in structured plantings or using molecular markers to identify genes associated with important traits. In many cases, these two main approaches are combined.

Methods like QTL mapping and genome-wide association studies (GWAS) link genetic variants to phenotypic traits, providing insights into the genetic basis of adaptation and diversity.

Genomic science technologies have their roots in advancements made during the Human Genome Project. Significant investments by the pharmaceutical and biomedical industries have driven genomic research forward, primarily to discover new drugs and develop treatments for various diseases. These genomic technologies are also being applied to improve crops, livestock, and forest trees. However, the substantial financial commitments made by private companies often result in the patenting of genomic data and tight control over intellectual property. Consequently, while genomic sciences hold the potential to greatly enhance our understanding of genomes and gene functions, the resulting discoveries and innovations might be restricted by corporate ownership, limiting access to this knowledge and its applications.

7.2 ASSOCIATION GENETICS

As a subset of functional genomics, **QTL (Quantitative Trait Loci) mapping** studies help estimate the number, effects, and approximate locations of genes that influence complex traits. However, pinpointing the exact gene responsible for a QTL remains challenging. Research suggests that most commercially important quantitative traits in forest trees are likely controlled by a polygenic mode of inheritance, involving many genes with small effects. The success of a QTL mapping study partly depends on understanding the genetic basis of the quantitative trait in question. A key initial step is to estimate the number of genes involved in controlling a trait. Although theoretical methods exist for this purpose, they were not commonly used in forest trees before the advent of QTL mapping. The primary question is whether a trait is **polygenic**, controlled by many small-effect genes, or **oligogenic**, controlled by fewer genes with larger effects. Intuitively, QTLs are easier to detect for traits under oligogenic control than those under polygenic control. However, QTL mapping experiments must often proceed without prior knowledge of the genetic architecture of a trait and derive insights about genetic control based on their results. Despite the challenges, since its inception in the 1060s, QTL mapping has greatly enhanced our understanding of the genetic basis of complex traits in forest trees. Recently, more traditional, labor-intensive approaches, such as segregation analysis across multiple generations—which is not feasible for many long-lived tree species—have been replaced by newer strategies.

The updated approach, known as **association mapping**, allows for more precise identification of the specific genes that control complex traits, ultimately pinpointing the mutations responsible for phenotypic differences in quantitative traits among individuals. Each mutation that affects a trait is termed a quantitative trait nucleotide (QTN). The main distinction between QTL mapping and association mapping lies in their methods: QTL mapping depends on genetic linkage from a few generations of controlled crosses, while association mapping utilizes population-level linkage disequilibrium (LD) between genetic markers and QTNs, which has developed over many generations in a large, intermixed population. Originally designed to identify genes associated with complex human traits (Cardon & Bell, 2001; Weiss & Clark, 2002), association mapping has only recently been adapted for use in plants (Neale & Savolainen, 2004; Rafalski, 2002).

There are two basic approaches to association genetics:

(1) Genome scan
(2) Candidate gene

In the genome scan approach, genetic markers are distributed across the entire genome, allowing for a comprehensive search for quantitative trait nucleotides (QTNs). In contrast, the candidate gene approach involves placing only genetic markers within specific genes that influence a particular phenotype. While the genome scan method, now often realized through **whole-genome sequencing**, offers a more thorough investigation, it is also more costly to implement.

Association genetics shares four key components with quantitative trait loci (QTL) mapping: (1) A mapping population; (2) Phenotypic data for the quantitative trait(s) for every individual in the mapping population; (3) Genetic marker information for each member of the population; and (4) A statistical method to link genotypes to phenotypes. Although the issues related to phenotyping are similar for both methods, the other three components differ significantly between them.

The effectiveness of an association genetics study depends on the extent of linkage disequilibrium (LD) between genetic markers and QTNs at the population level. The degree of LD is influenced by the population's history, including past bottlenecks and subsequent recombination events (as discussed in Chapter 3). For instance, a population that has recently experienced a bottleneck or has a low recombination rate will generally exhibit higher LD than a large, randomly mating population with high recombination rates, common in many abundant conifer species.

Most natural forest tree populations with lower LD levels are expected to exhibit the latter characteristics. However, artificially created populations, such as those used in breeding programs, will likely display much higher LD. Higher levels of LD increase the likelihood of detecting associations between genetic markers and QTNs, which is particularly beneficial for breeding purposes, as will be further detailed in subsequent sections. Simulation studies suggest that association mapping populations should comprise at least 500 individuals (Long and Langley, 1999). Appropriate experimental designs for field trials, including spatial analysis and clonal replication, should be employed to enhance the precision of phenotype evaluation.

Association mapping in forest trees is a relatively recent development. For example, (González-Martínez et al. (2007) a study using a candidate gene approach was conducted to identify associations between genes involved in lignin and cellulose synthesis and various wood property traits in *Pinus taeda*. The study found linkage disequilibrium between SNPs within genes, but the LD diminished significantly with increasing distances between base pairs, becoming nearly undetectable beyond 2,000 base pairs. They discovered associations between SNP genotypes in certain lignin biosynthetic pathway genes and wood property phenotypes. This preliminary research illustrates the potential of association mapping to identify genes that control complex traits in forest trees, paving the way for marker-assisted breeding in both within-family and between-family selection scenarios.

7.2.3 PRINCIPLES OF GENOME-WIDE ASSOCIATION STUDIES (GWAS)

Genome-wide association Studies (GWAS) are a powerful tool used to identify genetic variants associated with traits of interest. In forest genetics, GWAS is crucial in understanding the genetic architecture of complex traits such as growth rate, wood quality, disease resistance, and adaptation to environmental stresses. The core principle of GWAS is to scan the genome for single nucleotide polymorphisms (SNPs) that are statistically associated with a specific phenotype. By analyzing large populations of trees, researchers can identify genetic variants that contribute to phenotypic variation, leading to insights that can improve breeding programs and conservation efforts.

Key steps in conducting a GWAS:

Population Selection: A diverse population is selected, ideally encompassing a wide range of genetic backgrounds and phenotypic variations. In forestry, populations often include natural stands, breeding populations, or progeny trials.

Phenotyping: Accurate and consistent measurement of phenotypes is critical. In forestry, phenotypes might include height, diameter, wood density, leaf morphology, or resistance to pests and diseases.

Genotyping: The population is genotyped using high-throughput methods to identify SNPs across the genome. Techniques such as genotyping-by-sequencing (GBS) or SNP arrays are commonly used.

Statistical Analysis: Statistical models are applied to test the association between SNPs and the phenotype of interest. The most common method is a linear mixed model, which accounts for population structure and relatedness.

Interpretation and Validation: Significant associations are interpreted, often leading to the identification of candidate genes. Validation through independent populations or functional studies is essential to confirm the findings.

Application in Breeding: Identified markers can be used in marker-assisted selection (MAS) or genomic selection (GS) to accelerate breeding programs. For example, selecting trees with favorable alleles for growth or disease resistance can improve forest productivity and resilience.

BOX 7.1
Case Study for Norway spruce
Utilizing data from the Swedish breeding programs, this study involved over 5056 elite trees and their progenies, focusing on traits such as budburst stage, frost damage, tree height, and wood quality. The research identified 55 novel quantitative trait loci (QTLs) linked to these traits, with the largest effects observed for budburst stage. Notably, some QTLs exhibited pleiotropic effects, influencing multiple traits like frost damage and tree diameter simultaneously, suggesting a shared genetic basis. Among the candidate genes identified, the MAP3K genes stood out for their role in regulating cold stress responses, which are crucial for adaptation to climate (Chen et al., 2021)

7.3 MARKER-ASSISTED SELECTION

Marker-assisted selection (MAS) is a form of selection using genetic markers associated with a target trait. There are potential benefits to marker-assisted selection in forest trees, but there are also numerous hurdles to its successful application.

Almost any tree improvement program aims to increase the mean genotypic value of the breeding and production populations (defined in the chapter on tree improvement). Traits of long-term interest to tree breeders include growth and DBH, as well as wood properties and disease resistance. While significant genetic gains in these traits have been achieved using traditional selection methods, marker-assisted selection was identified during the 1990s as potentially useful to enhance gains and shorten the generation interval.

One type of MAS is described in Indirect Selection Based on Markers Linked to QTLs. In this method, pedigreed mapping populations are used to identify associations (i.e., linkages) between molecular markers and quantitative trait loci (QTLs) such that selection for specific alleles at the marker locus results in an increased frequency of favorable alleles at the QTL.

All forms of indirect selection involve selecting one trait to improve in a different trait, called the target trait (Chapter on quantitative genetics). Examples of classical indirect selection not involving molecular markers include: (1) Selection of seedling height to improve rotation age bole volume; (2) Selection of pilodyn penetration of trees in the field to improve total bolewood density.

The efficiency of indirect phenotypic selection compared to direct phenotypic selection on the target trait itself depends on the heritability of the two traits (indirect and target) and their genetic correlation (Equations in the chapter on quantitative genetics).

Indirect MAS is an extension of classical indirect selection because selection is based on genetic markers, the indirect traits correlated to target phenotypic traits. In indirect MAS, the correlation between the marker and the target trait normally results from the marker being linked to a region of the chromosome containing a gene affecting the target trait.

A second and developing form of MAS is to select directly on the individual alleles at one or more loci affecting polygenic traits (see Direct Selection on Genes Coding for Target Traits). This form of marker-assisted selection requires knowledge, at the molecular level, of some or all the genes controlling the target trait and can be viewed as direct selection rather than indirect selection since selection is for specific, favorable alleles at those loci. In this case, the markers are causal polymorphisms directly affecting the polygenic target trait (e.g., quantitative trait nucleotide, QTN, or insertion/deletions). To be successful, direct molecular selection would likely need to be applied to many, if not all, of the individual gene loci that code for polygenic target traits. Indirect selection based on genetic markers depends on the QTL mapping methods described before. In contrast, direct selection on gene coding for target traits depends on association genetics.

All forms of MAS can be applied separately or in conjunction with classical selection methods (mass, family, within-family, combined, and index selection) and can be utilized to make selections for selected breeding and/or production populations. Stage and combined index selection are possible ways to use marker and phenotypic information in selection programs (see dedicated texts on breeding strategies). In two-stage selection, a first round of selections is made at a very young age based solely on marker genotypes measured on all potential candidates for selection. Candidates with desirable alleles at the marker loci are chosen, and only the selected individuals are established in field progeny tests. Later, when the tests are old enough, a second selection round is made based on phenotypic measurements. The two stages of selection can be based on the same or different traits (e.g., selection for molecular markers related to disease resistance in the first round, followed by selection for growth and wood density measured in the field tests).

Alternatively, marker data at one or more loci can be included in a combined mixed model analysis with phenotypic measurements on the same or different individuals. Then, selection is based on the individuals' predicted genetic worth, combined with marker and phenotypic data. For example, if two SNP loci are known to affect bole volume growth in a tree species, trees growing in genetic tests could be measured for volume and genotyped at both SNP loci. All these data can be combined to predict aggregate genetic values.

Benefits of MAS might include:

(1) Decreasing the breeding cycle time
(2) Decreasing costs
(3) Increasing selection intensity
(4) Increasing efficiency of selection for low heritability traits.

Challenges of MAS include:

(1) The large task of QTL detection
(2) QTL by environment and QTL by genetic background interactions
(3) Linkage equilibrium conditions between markers and QTLs in tree populations.

Two basic approaches to marker-assisted selection in forest trees are possible. The first is finding linkages between anonymous flanking markers and QTLs in segregating populations. This approach can work for marker-assisted selection within the same segregating families as the linkages were discovered but cannot be applied to new families because of the general lack of strong linkage disequilibrium in forest tree populations.

The second approach uses association genetics to find associations between single nucleotide polymorphisms in candidate genes and the target trait. This approach can be used for family and within-family selection because marker by trait associations is tightly linked and will be in complete or nearly complete linkage disequilibrium in tree populations.

There have been high hopes in the past decades that genetics associated with single nucleotide polymorphisms may lead to applications in which marker data are combined with phenotypic measurements from progeny tests to enhance selection and breeding in tree improvement programs.

In addition to marker-assisted selection, molecular markers have many other potential uses in breeding programs. These include quality control, understanding mating systems in seed orchards, registering and protecting varieties, families, and clones, and 'smart' breeding.

While many of these applications are not yet operational, the future potential is great. As marker technologies evolve, costs decline, and our understanding of underlying mechanisms improves, molecular markers could become integral tools in many phases of tree improvement programs.

7.4 GENOMIC SELECTION

Genomic selection (GS) is an advanced method used in plant and animal breeding to predict the genetic potential of individuals based on their DNA. The effectiveness of GS depends on having a high density of genetic markers across the genome to take advantage of the associations between these markers and the actual genetic variants that affect traits, known as linkage disequilibrium (LD), for use in selection and breeding (Meuwissen et al. 2001). This approach relies on high-density genetic markers, such as single nucleotide polymorphisms (SNPs), distributed across the genome. These markers are associated with specific regions of DNA that influence important traits like yield, disease resistance, or growth rate. By analyzing these markers, genomic selection provides, under certain assumptions, a more comprehensive and accurate prediction of an individual's genetic value than traditional methods based solely on observable traits or pedigree information.

Compared to MAS, GS uses many markers spread throughout the entire genome, not just those associated with known genes or QTLs. This approach does not require prior knowledge about which specific genes or markers are linked to the trait of interest. Instead, GS relies on high-density genetic markers to capture the effects of all relevant genes across the genome, including both major and minor genes that contribute to complex traits. GS provides greater accuracy in predicting the genetic potential of individuals for complex traits because it accounts for the cumulative effects of many genetic variants across the genome rather than focusing on a few major genes. This increased accuracy is especially beneficial in breeding programs to improve polygenic traits. Second, GS enables breeders to perform earlier selection, even before phenotypic traits are expressed, which can significantly reduce the generation interval and speed up the breeding cycle. In contrast, MAS typically requires more time to identify and validate markers and is less effective for traits controlled by numerous small-effect genes.

However, genomic selection also has some disadvantages compared to marker-assisted selection. The initial costs of setting up a GS program can be higher due to the need for high-density genotyping across the genome and the development of robust statistical models. While potentially less powerful for complex traits, MAS can be more cost-effective and simpler to implement when targeting traits controlled by a few major genes. Additionally, for traits with a well-known genetic basis or when the cost of genotyping is a major concern, MAS can still be a viable choice.

The process of genomic selection starts with a training population, a group of individuals for whom phenotypic data (observable traits) and genotypic data (DNA information) are available. This data is used to develop a statistical model that estimates the effects of each genetic marker on the traits of interest. Statistical models, such as **Best Linear Unbiased Prediction (BLUP)** or Bayesian methods, are then applied to calculate each individual's **genomic estimated breeding values (GEBVs)**. These GEBVs combine information from all markers across the genome, allowing breeders to predict the genetic potential of individuals more accurately.

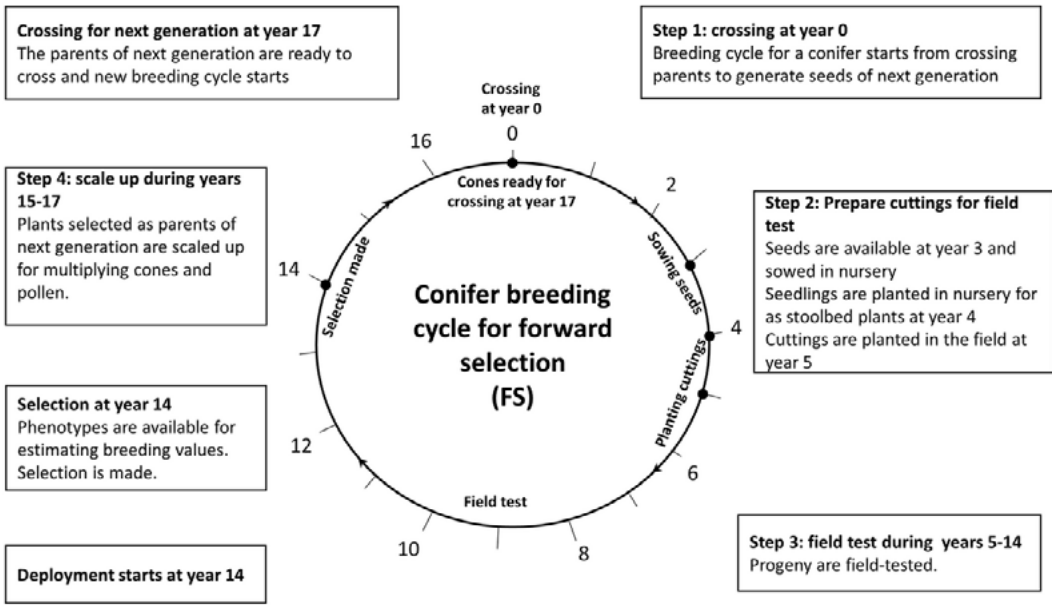


Fig. 7.1 Example of conifer breeding using forward selection; 17-year generation interval; adapted from Li & Dungey (2018); doi: https://doi.org/10.1371/journal.pone.0208232.g001

One of the main benefits of genomic selection is its increased accuracy in predicting an individual's genetic potential. This is achieved by using a dense set of genetic markers spread across the genome, providing a more detailed picture of the genetic basis of traits compared to traditional methods. Additionally, genomic selection allows for faster breeding cycles because it enables early selection of individuals based on their DNA information, rather than waiting for them to express the traits of interest through traditional field trials or progeny testing. This early selection speeds up the breeding process and reduces costs associated with extensive field testing. Over time, genomic selection can improve genetic gain by increasing selection accuracy and minimizing the generation interval, which is the time between generations.

Genomic selection has a wide range of applications across various fields. Plant breeding improves crops for yield, disease resistance, drought tolerance, and quality attributes. This approach is particularly valuable in crops with long generation intervals or when phenotyping is difficult or expensive. In animal breeding, genomic selection is routinely applied to livestock such as cattle, pigs, and poultry to enhance traits like milk production, growth rate, feed efficiency, and disease resistance. It also helps manage inbreeding while increasing genetic gain. Furthermore, genomic selection is used in forestry and aquaculture, where long generations and high phenotyping costs pose significant challenges for traditional breeding methods.

Implementing GS in forest trees presents several unique challenges due to the biological, technical, and economic characteristics of forestry breeding programs. Here are the main hurdles associated with applying GS to forest trees:

- **Long Generation Times:** Forest trees typically have long life cycles, often taking many years to reach maturity and produce offspring. This extended generation time slows the breeding cycle and makes collecting phenotypic data needed for training genomic prediction models challenging. The delay in obtaining mature phenotypic data can hinder the immediate benefits of GS, which relies on a training population with phenotypic and genotypic information (Fig. 7.1).
- **Large and Complex Genomes:** Many forest tree species have large and complex genomes, sometimes with high levels of polyploidy (multiple sets of chromosomes) and extensive genetic diversity. This complexity makes it more challenging to generate high-quality reference genomes and to identify and map the dense genetic markers needed for accurate genomic selection. Tree genomes' diversity and structural complexity can also complicate the development of robust genomic prediction models.

- **High Levels of Genetic Diversity:** Forest trees often exhibit high levels of genetic diversity within and between populations due to their long evolutionary history and adaptation to diverse environments. This genetic diversity can lead to a low degree of relatedness between individuals in breeding populations, which makes it harder to capture linkage disequilibrium (the non-random association of alleles at different loci) and predict the effects of genetic markers across various populations. As a result, the accuracy of genomic selection models may be reduced.
- **High Cost of Genotyping and Phenotyping:** Implementing GS requires extensive genotyping across large populations to capture sufficient genetic variation, which can be expensive, especially for species with large genomes. Additionally, phenotyping forest trees for important traits like growth rate, wood quality, or disease resistance often involves long-term and costly field trials. The combination of high genotyping and phenotyping costs can be a significant financial barrier to adopting GS in forestry.
- **Limited Reference Populations:** In forestry, there are often limited numbers of well-characterized reference populations with detailed phenotypic and genotypic data. Establishing these reference populations is time-consuming and expensive, and their data may not be directly transferable to other populations or species. The limited availability of reference data makes it difficult to develop and validate accurate genomic prediction models across diverse forest tree populations.
- **Complex Trait Architecture:** Many traits of interest in forest trees, such as growth rate, wood density, and resistance to pests and diseases, are complex and controlled by multiple genes with small effects, as well as environmental factors. This polygenic nature of traits adds complexity to developing accurate genomic prediction models, as it requires high-density markers and sophisticated statistical methods to capture the cumulative effects of many small-effect loci.
- **Environmental Variability and** GxE **Interactions:** Forest trees grow in diverse and often variable environments, leading to significant genotype-by-environment (*GxE)* interactions, where the expression of genetic traits depends on environmental conditions. *GxE* interactions can complicate the application of GS, as models developed in one environment may not be predictive in another. This variability necessitates developing environment-specific models or multi-environment trials, which increases the complexity and cost of implementing GS.
- **Lack of Breeding Infrastructure and Expertise**: Forestry breeding programs may lack the infrastructure, computational resources, and expertise needed to implement GS effectively. Many forestry breeding programs are smaller in scale compared to crops, with fewer resources dedicated to genetic research and breeding. Developing expertise in genomics, bioinformatics, and quantitative genetics is essential for successfully adopting GS.

Despite these hurdles, the potential benefits of genomic selection in forest trees—such as accelerating breeding cycles, improving selection accuracy for complex traits, and enhancing genetic gain—make it a promising tool for forestry breeding programs. Addressing these challenges through increased investment in genomic resources, training, and collaborative research efforts will be crucial for the widespread adoption of GS in forestry.

Genomic selection (GS) has been successfully applied to several tree species, particularly those of commercial importance in forestry and horticulture. Notable examples include Eucalyptus and pine species like loblolly pine, where GS has been used to improve traits such as growth rate, wood density, and disease resistance. Spruce species like Norway spruce, Sitka spruce, Douglas-fir, and poplar species have also seen successful GS applications to enhance traits like wood quality, growth, and climate adaptability. Additionally, GS has been used in horticultural species such as apple trees to improve fruit quality and disease resistance. These applications demonstrate the potential of GS to accelerate breeding cycles and improve complex traits across a range of tree species. However, its effectiveness often relies on deployment strategies, as illustrated in Fig. 7.2.

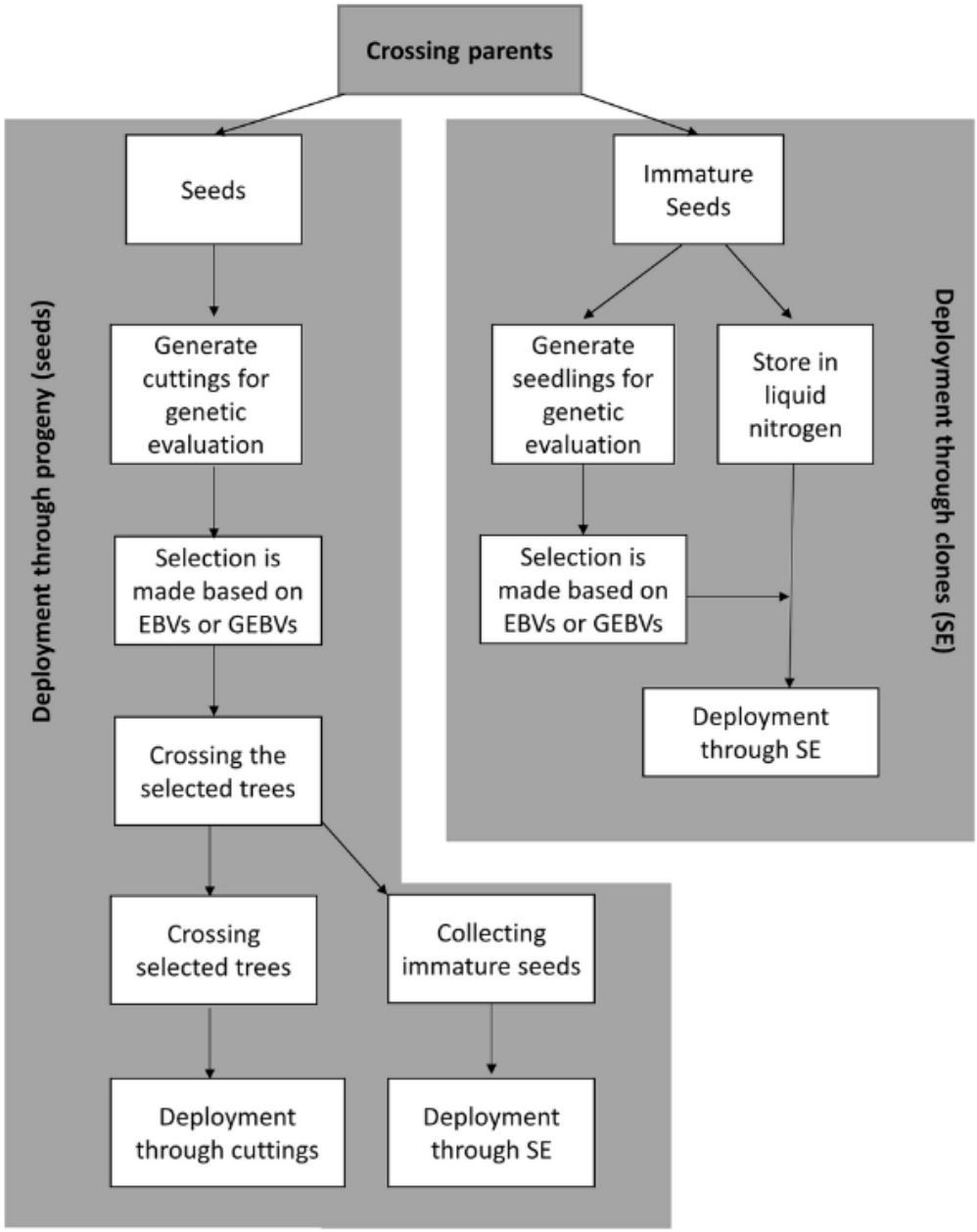


Fig 7.2 Deployment of progeny vs. clones of the selected trees in GS setting. Adapted from Li & Dungey (2018); doi: https://doi.org/10.1371/journal.pone.0208232.g003; GEBVs - genomic breeding values, EBVs, estimated from BLUP (Henderson's mixed linear model equation (Henderson, 1975), SE – somatic embryogenesis.

7.5 HISTORICAL AND EVOLUTIONARY CONTEXT FOR CURRENT GENOMIC STUDIES

Historical climatic events, such as glaciations and interglacial periods, have shaped the distribution and diversity of forest trees. Understanding the genetic consequences of these events helps trace tree species' evolutionary history and migration patterns.

Post-Glacial Migration and Related Marker Studies

Historical climatic changes, particularly glaciation and interglacial periods, have profoundly influenced forest trees' distribution and genetic diversity. The Last Glacial Maximum (LGM), which occurred approximately 31,000 to 16,000 years ago, significantly shaped the biogeography of European forest trees. During glaciation periods, species were forced into refugia—isolated areas where conditions remained favorable for survival.

Refugia and Post-Glacial Expansion

Regions in southern Europe, such as the Iberian Peninsula, Italy, and the Balkans, were refugia for many tree species during glacial periods. These areas provided the necessary conditions for survival and subsequent recolonization.

As glaciers retreated, trees began to recolonize northern regions. This expansion led to new populations and genetic differentiation based on the varying environments encountered during migration.

Genetic Markers in Forest Trees' Migration Studies

Genetic markers, such as mitochondrial DNA (mtDNA), chloroplast DNA (cpDNA), and nuclear DNA markers (e.g., microsatellites and SNPs), have been extensively used to study post-glacial migration patterns. These markers help trace lineage relationships and historical gene flow among populations.

Mitochondrial DNA (mtDNA) traces maternal lineages due to its maternal inheritance pattern. Studies using mtDNA have revealed the presence of distinct haplotypes corresponding to different refugial origins.

Chloroplast DNA (cpDNA): Inherited maternally in most angiosperms and paternally in gymnosperms, cpDNA is valuable for studying plant phylogeography. It provides insights into seed dispersal and historical migrations.

Nuclear DNA Markers: Microsatellites and SNPs offer high-resolution insights into genetic diversity and structure. These markers are used to identify genetic differentiation and adaptation to local environments.

Case Studies of Post-Glacial Migration

Norway Spruce (*Picea abies***)**: Genetic studies indicate that Norway spruce survived the LGM in multiple refugia across Europe. Fossil pollen records and genetic analyses suggest that populations in the Carpathians, the Alps, and the Balkans served as major refugial sources. Post-glacial expansion from these refugia led to the current distribution and genetic diversity observed in Norway spruce populations.

Scots Pine (*Pinus sylvestris***)**: Like Norway spruce, Scots pine also retreated to southern refugia during glaciations. Genetic studies using mtDNA and cpDNA markers have identified distinct genetic lineages corresponding to different refugial origins. Post-glacial migration routes have been mapped, showing recolonization paths from the Iberian Peninsula, Italy, and the Balkans into northern and central Europe.

European Beech (*Fagus sylvatica***)**: The distribution of cpDNA haplotypes indicates that European beech expanded from multiple refugia, including the Apennines, the Balkans, and the Pyrenees. The genetic structure of current populations reflects the historical migrations and subsequent hybridization events.

Implications for Conservation and Management

Understanding forest trees' post-glacial migration and genetic diversity is crucial for conservation and management. It provides insights into the historical processes shaping current genetic structures and helps predict how tree populations might respond to environmental changes. This knowledge is essential for developing strategies to preserve genetic diversity, ensure sustainable forest management, and enhance the resilience of forest ecosystems to climate change.

7.6 SUMMARY AND CONCLUSIONS

Genomics of forest trees provides comprehensive insights into the genetic basis of adaptation, diversity, and evolution in these critical components of forest ecosystems. The advancements in sequencing technologies and genomic analysis tools have significantly enhanced our ability to study and utilize forest tree genomes for conservation, breeding, and understanding ecological interactions. Post-glacial migration studies further enrich our understanding by revealing the historical dynamics that shaped the genetic landscape of forest trees, informing future conservation and management practices.

7.7 REFERENCES

Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics*, *2*(2), 91–99.

González-Martínez, S. C., Wheeler, N. C., Ersoz, E., Nelson, C. D., & Neale, D. B. (2007). Association genetics in Pinus taeda LI Wood property traits. *Genetics*, *175*(1), 399–409.

Henderson, C. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, *31*(2), 423. https://doi.org/10.2307/2529430

Chen, Z.-Q., Zan, Y., Milesi, P., Zhou, L., Chen, J., Li, L., Cui, B., Niu, S., Westin, J., Karlsson, B., García-Gil, M. R., Lascoux, M., & Wu, H. X. (2021). Leveraging breeding programs and genomic data in Norway spruce (Picea abies L. Karst) for GWAS analysis. *Genome Biology*, *22*(1), 179. https://doi.org/10.1186/s13059-021-02392-1

Li, Y., & Dungey, H. S. (2018). Expected benefit of genomic selection over forward selection in conifer breeding and deployment. *PLOS ONE*, *13*(12), e0208232. https://doi.org/10.1371/journal.pone.0208232

Neale, D. B., & Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science*, *9*(7), 325–330. https://doi.org/10.1016/j.tplants.2004.05.006

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in plant biology*, *5*(2), 94–100.

Weiss, K. M., & Clark, A. G. (2002). Linkage disequilibrium and the mapping of complex human traits. *TRENDS in Genetics*, *18*(1), 19–24.

8. GENETIC ENGINEERING

8.1 TARGET TRAITS, TRANSFORMATION AND REGENERATION

Genetic improvement of forest trees has largely progressed since the 1950s using traditional selective breeding approaches. During the past two decades, however, a new strategy involving the transfer of genes by genetic engineering (GE) has added a new option for genetic modification of the tree genome. GE fundamentally differs from traditional or marker-based breeding in that foreign genes coding for traits not normally found in the target plant can be introduced.

Genetic engineering is the process of directly manipulating an organism's DNA to alter its genetic makeup. This technology enables scientists to add, remove, or modify specific genes within an organism's genome, giving it new traits or functions that wouldn't occur naturally. In forestry, GE is often related to transferring genes between organisms from one species to another using molecular fragments, including a desired gene.

The first step often involves isolating the gene of interest, which is then copied or cloned. The DNA fragment (gene of interest) could be inserted into a so-called **vector** that can be a plasmid, a small circular chromosome from a bacterium. This technique relies heavily on discovering the restriction enzymes that cleave the DNA at defined nucleotide sequences, which are characteristic of each enzyme. The plasmid, now a recombinant DNA molecule, is then introduced into a host cell in which it can replicate and produce many copies. Also, these host cells can proliferate if cultivated on a selective medium, allowing only those cells with recombinant DNA molecules to survive and propagate (selective screening by antibiotic resistance). This multiplication of a gene is an example of gene cloning.

Several methods can be used to introduce recombinant DNA into host organisms. In forestry, there are several methods commonly used: i) agrobacterium-mediated transformation, ii) gene gun (biolistic method), iii) fusion of protoplasts, or iv) CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) technology.

Agrobacterium-mediated transformation is gene transfer via biological vectors, most often by the bacterium *Agrobacterium tumefaciens,* which can naturally transfer part of DNA into plant cells. The biolistic method (short for "biological ballistics") is a direct gene transfer method using a gene gun to physically shoot small metal particles coated with recombinant DNA into plant cells. In the protoplast fusion method, plant cells are treated to remove their cell walls, creating protoplasts, which can then fuse with other protoplasts, allowing for the exchange of recombinant DNA.

CRISPR-Cas9 technology is the latest groundbreaking discovery in molecular genetics developed by Jennifer Doudna and Emmanuelle Charpentier in 2012. CRISPR-Cas9 uses a guide RNA to locate a specific DNA sequence in the genome, where the Cas9 enzyme then makes a precise cut. The cell's natural repair mechanisms deactivate the targeted gene or introduce a new sequence during the repair process. CRISPR-Cas9 technology is an accurate, efficient, and versatile tool for gene editing, allowing scientists to modify specific DNA sequences within an organism's genome. Its applications span medicine, agriculture, and research, but its ethical implications, particularly in human gene editing, are subjects of ongoing debate.

BOX 8.1

Genetic engineering relies on several key genetic methods developed since the 1970s; all these specific techniques are described in the initial chapter:

(1) Methods of creating **recombinant DNA** molecules that include a sequence of DNA in which two nonhomologous DNA segments have been combined often from quite different species.
(2) Methods of **DNA sequencing**
(3) **Polymerase Chain Reactions (PCR)**
(4) Methods of producing **synthetic genes** or part of genes
(5) Methods of revealing **gene function**

To produce multiple copies of a specific gene, the creation of recombinant DNA molecules is crucial, a process commonly known as **gene cloning**. This involves isolating the desired DNA fragment that includes the gene of interest, then inserting it into a vector through ligation, forming the recombinant DNA molecule. The vector, often a **plasmid** derived from bacterial sources, may entirely come from a different species. This vector produces many copies of the DNA fragment inside a bacterial cell.

Plasmids are small, circular DNA molecules that coexist with the standard chromosome in bacterial cells (for more information, see the initial chapter). These plasmids typically carry only a few genes, including those for antibiotic resistance. Additionally, plasmids can be engineered with histochemical markers, specific staining properties, or nutritional markers that enable cells containing the plasmid to grow on media lacking essential nutrients. Plasmids replicate independently of the host cell's chromosome, a feature exploited to multiply an introduced DNA fragment. When a bacterial cell, for instance, is grown on a suitable medium, it divides, passing copies of the recombinant DNA molecule to its daughter cells, further amplifying the DNA fragment. As all daughter cells originate from the same parent cell, the resulting colony forms a clone, hence the term gene cloning.

Gene cloning allows the establishment of a DNA library, a collection of clones representing nearly all fragments of an organism's genomic DNA or complementary DNA (cDNA). cDNA, synthesized from mRNA, represents the expressed genes. Depending on the research purpose, the library can be a genomic or cDNA library. The discovery of restriction enzymes, which are specific enzymes that cut DNA molecules at particular nucleotide sequences, has made it possible to "cut" DNA at precise locations. These enzymes act like molecular scissors, cutting only at specific recognition sequences unique to each enzyme. Moreover, they generate two main types of DNA ends, either blunt or sticky, with recognition sequences typically consisting of 4 to 8 base pairs.

Understanding **restriction enzymes (for the most important information, please go to Chapter 1)** is crucial for creating recombinant DNA. Forming a recombinant DNA molecule involves using a plasmid from the bacterium Escherichia coli (*E. coli*) as a vector. This plasmid contains a specific recognition site for the particular restriction enzyme, for example, EcoRI. When treated with EcoRI, the plasmid is cut into a linear DNA molecule with single--stranded sticky ends. Similarly, treating DNA from plant or animal cells with EcoRI results in fragments with the same sticky ends. These fragments can then be joined with the linear plasmid using DNA ligase, an enzyme that naturally repairs single-stranded breaks in DNA. This creates a recombinant DNA molecule because the DNA fragment and the plasmid share complementary sequences. Once formed, the recombinant DNA can be replicated using gene cloning techniques. This allows further study of the gene's structure, function, and expression, including determining the nucleotide sequence through DNA sequencing or introducing the gene into different host cells.

8.2 APPLICATION OF GENETIC ENGINEERING IN FOREST TREES

Genetic engineering presents a powerful tool for enhancing forest trees by allowing the introduction of specific genes into tree varieties that have been improved through traditional breeding methods. This process maintains the integrity of much of the existing genetic makeup. This approach is particularly advantageous for forest trees with lengthy breeding cycles, where conventional crossing and backcrossing methods are impractical for transferring single genes due to the extensive time and effort involved. The forestry industry can achieve more precise and efficient genetic improvement by integrating genetic engineering with traditional breeding programs. Moreover, using gene editing tools has enabled the development of trees with enhanced resilience to environmental stressors, such as drought and extreme temperatures. For instance, genes regulating water use efficiency and stress response pathways have been edited to produce trees that can withstand climate change impacts better.

BOX 8.2

The difference between marker-assisted breeding (MAB) and genetic engineering

Both are used in tree breeding, but they differ significantly in their approaches. MAB enhances traditional breeding by using molecular markers linked to desirable traits, such as disease resistance or drought tolerance, to identify and select individual trees with favorable genetic profiles. This process speeds up the selection of trees with specific traits but relies on natural genetic variation within the species. The breeding process remains natural, with no direct alteration of DNA.

In contrast, genetic engineering allows scientists to directly modify a tree's genome by adding, removing, or editing specific genes. This method can introduce entirely new traits into the tree, even from other species, which is impossible with MAB. For example, genetic engineering might insert a gene from bacteria into a tree to make it resistant to pests or herbicides, something MAB cannot achieve. Genetic engineering offers more precision and can significantly accelerate the development of new traits, while MAB is limited to selecting traits that already exist within the species' gene pool. Thus, MAB supports more traditional breeding methods with improved efficiency, while genetic engineering directly manipulates the genetic makeup of trees to produce desired outcomes more rapidly and precisely.

Practical Applications in Fast-Growing Species

In the near term, genetic engineering applications are likely to first appear in fast-growing hardwood species such as *Populus* (poplar) and *Salix* (willow), which are favored for their rapid growth and short rotation cycles. These species serve as ideal models for genetic engineering due to their well-characterized genomes and established transformation protocols. Efforts are underway to enhance traits like biomass yield, pest resistance, and adaptability to different environmental conditions.

For example, recent studies have focused on modifying lignin biosynthesis in poplar to produce trees with lower lignin content, which is desirable for paper production and biofuel applications. Reducing lignin makes the trees easier to process, leading to more efficient and sustainable production methods.

8.3 GENETIC TRANSFORMATION TECHNIQUES

Genetic transformation involves the transfer of recombinant gene constructs into plant cells, the selection of transformed cells, and the regeneration of these cells into transgenic plants. Achieving this requires advancements in three key areas:

(1) isolating and identifying genes crucial for tree breeding, including discovering promoters that allow appropriate gene expression.
(2) developing reliable gene transfer methods; and
(3) creating efficient regeneration systems to produce and propagate transgenic plants.

Researchers are employing advanced genomic and transcriptomic analyses to improve gene isolation and identification to pinpoint key regulatory genes involved in traits like growth, wood formation, and stress resistance. These efforts are supported by the increasing availability of high-quality reference genomes for forest species, which provide a comprehensive blueprint for targeted genetic modifications.

Promoters used in forest trees are also becoming more refined to achieve tissue-specific or inducible expression of transgenes. For example, promoters activated under stress conditions are being used to enhance the expression of stress-responsive genes only when needed, reducing potential negative impacts on the tree's normal growth and development.

8.3.1 ADVANCED GENE TRANSFER METHODS

Traditional gene transfer techniques include biological vectors, such as *Agrobacterium tumefaciens*, and direct methods like biolistics or "gene guns." The former method can exclusively be used in plants as animals cannot be infected by *Agrobacterium*. The latter method has proven viable both in plants and in animals. However, new techniques are being developed to improve efficiency and reduce unintended effects. One promising approach is using viral vectors that can deliver genetic material more precisely and with higher efficiency than traditional methods. These vectors are engineered to be non-pathogenic and can be designed to target specific cell types within the plant.

Another emerging technique is nanoparticle-mediated gene delivery, where DNA is coated onto nanoparticles that are then introduced into plant cells. This method can overcome some of the limitations of traditional gene transfer methods, such as low transformation efficiency and limited tissue specificity.

8.3.2 BIOLOGICAL VECTORS IN PLANT BIOTECHNOLOGY

The most employed vectors for plant genetic transformation are derived from two bacterial species, *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. Naturally found in soil, these bacteria can infect wounded plants, leading to abnormal growths. Specifically, *A. tumefaciens* causes crown gall disease, resulting in tumor-like growths on the plant stem, while *A. rhizogenes* leads to the development of hairy roots.

When *Agrobacterium* infects a plant cell, it transfers a Ti plasmid into the plant's genome. A specific portion of this plasmid, the T-DNA region, integrates into the plant's chromosomes, triggering the plant to overproduce hormones. This hormonal imbalance causes uncontrolled cell division and the formation of galls in the case of *A. tumefaciens*. *A. tumefaciens* primarily infects dicotyledons, including woody plants such as poplar (*Populus*) and willow (*Salix*). In contrast, monocotyledons are generally resistant under natural field conditions. Although natural infections are rare, both *Agrobacterium* species can transfer genes to conifers. In controlled environments, however, infections can be induced by directly inoculating the bacteria into the plant stem.

The genes responsible for hormone production in these bacteria can be removed from the plasmid and replaced with other genes, which are valuable for purposes such as tree breeding. When this modified Ti plasmid is used to infect conifer cells in vitro, the cells do not produce excess hormones, allowing them to proliferate, mature, and eventually develop into new plants.

8.3.3 DIRECT GENE TRANSFER USING BIOLISTIC METHODS

Direct gene transfer is a technique that introduces foreign DNA into plant cells without relying on biological vectors such as bacteria. One of the most prominent and successful methods in this category is the biolistic approach, also known as particle bombardment or microprojectile bombardment.

This method involves using a particle accelerator, a specialized device that propels tiny particles at high velocities into plant cells. The particles, typically made of gold or tungsten, are first coated with the DNA of interest, such as plasmid DNA, which carries the genes intended for transformation. Upon acceleration, these DNA-coated micro-projectiles penetrate the plant cell walls and membranes, delivering the genetic material directly into the plant cells' nuclei or other cellular compartments where integration into the plant genome can occur.

Biolistic transformation has proven particularly useful in species less susceptible to traditional transformation methods, such as those involving *Agrobacterium*-mediated gene transfer. This method has been successfully employed in conifer species, where it has produced transgenic plants. For instance, in white spruce (*Picea glauca*), biolistic methods have been successfully used to create genetically modified plants. This technique has also been effectively applied to other species within the *Picea* genus, including Norway spruce (*Picea abies*), where it has generated hundreds of transgenic cell lines and plants.

The biolistic method has become vital in plant genetic engineering, particularly for species recalcitrant to other transformation techniques. Its ability to introduce genetic material into a wide variety of plant tissues, including those of economically important conifers, underscores its significance in forestry biotechnology and plant breeding programs aimed at improving traits such as disease resistance, growth rate, and wood quality.

8.4 EMERGING APPLICATIONS AND ADVANCES

Recent advancements in genetic engineering have expanded its potential applications in forestry. For example, CRISPR/Cas9, a precise genome editing tool, has emerged as a transformative technology for forest tree improvement. This method allows for targeted modifications at specific sites in the genome, enabling the knockout of undesirable genes or the introduction of beneficial traits without affecting the rest of the genome. CRISPR has been successfully applied to forest species like Populus and Eucalyptus to enhance disease resistance, wood quality, and growth rates.

CRISPR technology has been successfully applied to forest species like *Populus* (poplar) and *Eucalyptus* to enhance various traits such as disease resistance, wood quality, and growth rates. Here are some key case studies and findings:

Populus: *Populus* species have been a primary focus for CRISPR applications due to their importance in biofuel production, paper industry, and ecological studies. One notable study involved the CRISPR/Cas9-mediated knockout of the phytoene desaturase (PDS) gene in Chinese white poplar (*Populus tomentosa*), which served as a model to demonstrate the effectiveness of CRISPR in forest trees. The knockout of PDS resulted in an albino phenotype, which is useful for confirming successful gene editing. Beyond proof-of-concept, CRISPR has been used to target genes related to lignin biosynthesis in *Populus*, aiming to modify wood properties for more efficient processing in the paper and biofuel industries.

Eucalyptus: CRISPR/Cas9 has been employed to explore gene function related to growth regulation and stress responses in Eucalyptus. One study used CRISPR to knock out genes in the *Eucalyptus grandis* species, which allowed researchers to understand better the genetic basis of traits like growth rate and resistance to environmental stressors. Additionally, CRISPR has been used in conjunction with morphogenic regulators like WUSCHEL and BABY BOOM to enhance transformation efficiencies and improve the regeneration of edited plants, which is particularly challenging in tree species due to their complex genomes and slow growth rates.

Advanced Editing and Regeneration Techniques: Researchers are also exploring alternative Cas nucleases like Cas12a, which may offer more precise editing options for forest trees. Furthermore, efforts to boost regeneration during transformation—such as co-expressing morphogenic regulator genes—have shown promising results in improving the efficiency of producing transgenic and genome-edited trees, especially in species like poplar.

These advancements illustrate the growing potential of CRISPR in forestry, not only for improving commercially valuable traits but also for enhancing the ecological resilience of forest species in response to climate change and other environmental pressures (Bruegmann et al., 2024; Sheng et al., 2023; Yao et al., 2023).

8.5 WHICH TRAITS ARE MOST SUITABLE FOR GENETIC ENGINEERING IN FOREST TREES?

1. Disease Resistance:

Genetic engineering in forest trees has shown great promise in developing disease-resistant varieties, a critical trait for maintaining healthy forests and commercial plantations. Scientists have protected trees from devastating diseases by introducing genes that confer resistance to specific pathogens, such as fungi, bacteria, and viruses. For example, the American chestnut (*Castanea dentata*) has been genetically engineered to resist chestnut blight, a fungal disease that nearly wiped out this species in the early 20th century. Similarly, efforts to develop Dutch elm disease-resistant elms have been successful, offering hope for restoring these iconic trees (Merkle et al., 2007).

2. Insect Resistance:

Insect pests can cause significant damage to forests and plantations, leading to economic losses and ecological disruptions. Genetic engineering has introduced insecticidal proteins, such as those from *Bacillus thuringiensis* (Bt), into trees to enhance their resistance to pests like the spruce budworm. This approach reduces the reliance on chemical insecticides, which can have harmful environmental effects. Trees engineered with Bt genes effectively control pest populations, making them a valuable tool in integrated pest management strategies.

3. Herbicide Tolerance:

Another trait that has been successfully engineered in forest trees is herbicide tolerance. This allows trees to withstand the application of herbicides used to control competing vegetation in commercial plantations. By inserting genes that confer resistance to herbicides, such as glyphosate, trees can be selectively protected while surrounding weeds are eliminated. This trait is particularly useful in establishing new plantations, where controlling competition from other plants is crucial for the successful growth of young trees (Merkle et al., 2007).

4. Growth Rate and Biomass Production:

Improving trees' growth rate and biomass production through genetic engineering has significant economic and environmental benefits. Genes involved in hormone pathways or wood formation can be modified to produce faster-growing trees with increased biomass. For instance, genetically modified poplar and eucalyptus trees have shown enhanced growth rates, making them more efficient for timber production, paper manufacturing, and bioenergy. These improvements also contribute to carbon sequestration, as faster-growing trees can absorb more atmospheric carbon dioxide (Strauss et al., 1995).

5. Wood Quality and Composition:

The quality and composition of wood are important factors for various industrial uses, including pulp and paper production and bioenergy. Genetic engineering allows for modifying lignin content and composition in trees, making the wood easier to process. Reducing lignin levels can lower the environmental impact of pulping processes and improve the efficiency of biofuel production. For example, genetically engineered poplar trees with altered lignin biosynthesis have demonstrated improved pulping efficiency and reduced chemical use in paper production (Baucher et al., 1998).

6. Abiotic Stress Tolerance:

Trees are often exposed to abiotic stresses such as drought, cold, salinity, and poor soil conditions, limiting their growth and survival. Genetic engineering can enhance the tolerance of trees to these stresses by introducing genes that help them cope with adverse environmental conditions. For instance, engineering genes have developed drought-tolerant trees that improve water retention or enhance root growth. These genetically modified trees are better equipped to thrive in challenging environments, making them valuable for reforestation and afforestation projects in arid and semi-arid regions.

7. Reproductive Control:

Controlling the reproductive processes of genetically engineered trees is important to prevent unwanted gene flow to wild populations and manage transgenes' spread. Genetic engineering can induce sterility or delay tree flowering, ensuring they do not crossbreed with wild relatives. This is particularly important for transgenic trees planted near natural forests, where gene flow could have ecological consequences. Techniques such as RNA interference (RNAi) have been employed to achieve reproductive control, making transgenic trees more environmentally sustainable (Brunner et al., 2007).

8. Phytoremediation:

Phytoremediation is using plants to remove or stabilize contaminants from the environment. Genetically engineered trees have been developed to enhance their ability to accumulate heavy metals and other pollutants from polluted soils and water. For example, genetically modified poplars have been engineered to take up and detoxify pollutants such as trichloroethylene, a common groundwater contaminant. This application of genetic engineering offers a sustainable and cost-effective method for cleaning up contaminated sites and restoring ecological balance (Doty et al., 2007).

9. Delayed Flowering and Extended Juvenile Phase:

Delaying the onset of flowering and extending the juvenile phase in trees can benefit forestry and breeding programs. Genetic engineering can modify genes that control the transition from juvenile to adult phases, allowing trees to focus on vegetative growth for a longer period before diverting energy to reproduction. This trait is advantageous in commercial forestry, where maximizing biomass and wood production is often more important than early flowering. Additionally, extended juvenile phases can facilitate controlled breeding and genetic improvement of tree species (Hackett, 2011).

10. Carbon Sequestration:

Enhancing the carbon sequestration capacity of trees through genetic engineering is a promising strategy for mitigating climate change. By engineering trees to grow faster, produce more biomass, or alter their wood composition to store more carbon, they can significantly reduce atmospheric carbon dioxide levels. Trees with enhanced carbon storage capacity could be planted in reforestation and afforestation projects to offset carbon emissions. This approach supports climate goals and contributes to the restoration of degraded ecosystems (Harfouche et al., 2011).

8.6 REGENERATION AND PROPAGATION

An effective system for regenerating and propagating transformed cells into whole plants is essential to develop a transgenic plant. This remains a major hurdle for many tree species, particularly conifers. However, the ability to produce transgenic plants gradually expands to include a wider range of species.

The regeneration and propagation of transgenic cells continue to pose significant challenges in forest tree biotechnology, especially for coniferous species. Nonetheless, effective regeneration systems are already in place for several woody species, such as hybrid aspen, poplars, and spruce species, such as *Picea glauca*, *Picea abies*, and *Picea. mariana*.

Producing transgenic plants from modified cells is still a major challenge in forest tree genetic engineering. Improvements in tissue culture methods, including somatic embryogenesis and organogenesis, have enhanced the regeneration rates for several tree species. Researchers are also investigating species-specific growth regulators and customized culture media to improve regeneration and transformation success rates.

Organogenesis

Organogenesis refers to the development of organs, such as shoots and roots, in plants through the induction of these organs from bud meristems, internodal stem explants, or callus cultures. This process can follow two main pathways: the formation of adventitious buds or shoots from callus tissues and the growth of axillary buds or shoots from existing bud meristems or internodal stem sections. The success of organogenesis depends largely on using an appropriate culture medium, with variations of Marashige and Skoog's (1962) medium, such as Woody Plant Medium, commonly employed for forest tree species.

Explants are typically cultured in sterile containers, such as test tubes, Petri dishes, flasks, or jars, to prevent microbial contamination. They are kept under controlled artificial light conditions (1000-3000 lux) and temperature (24-26°C) to optimize organ growth and differentiation. Cytokinins primarily influence the differentiation of shoots, while auxins such as NAA (naphthaleneacetic acid), IAA (indole-3-acetic acid), IBA (indole-3-butyric acid), and 2,4-D (2,4-dichlorophenoxyacetic acid) are essential for root initiation. The balance between cytokinins and auxins in the culture medium determines whether shoots or roots will develop, with the specific cytokinin-to-auxin ratio is crucial for directing plant tissue growth and differentiation.

To scale up tissue culture technology for the large-scale regeneration of transgenic plants, it is essential to establish a micropropagation method that is straightforward and highly reproducible. Effective regeneration methods via organogenesis have been successfully developed for several angiosperm tree species, including Populus, Betula, and Eucalyptus. For genetic engineering, stock shoot cultures from bud explants are commonly prepared and maintained. Techniques involve co-cultivating leaf discs or internodal stem segments with **Agrobacterium** in vitro, facilitating the transfer of foreign genes from the bacterium to the plant cells.

After the co-cultivation period, plant tissues are cultured in a medium containing a selectable marker, such as kanamycin or another antibiotic. This helps identify and select transformants that exhibit resistance to the antibiotic, indicating successful gene transfer.

Somatic Embryogenesis

Somatic embryogenesis involves the development of embryos from somatic or non-reproductive tissues, which then grow into full plants. These somatic embryos originate from embryogenic cells found in either young or mature embryos and are structurally like zygotic embryos, as they possess both shoot and root poles. Somatic embryogenesis is a one-step process, unlike organogenesis, where shoots and roots typically develop sequentially in different culture media. However, it does not mean somatic embryos form immediately from cultured immature

zygotic embryos; the process often involves multiple stages, including initiation, embryo development, and maturation. Each stage may require different culture media and varying concentrations of phytohormones, such as 2,4-D, BAP, kinetin, NAA, and ABA (abscisic acid).

The key distinction between somatic embryogenesis and organogenesis lies in how the embryos develop. In somatic embryogenesis, embryos grow like zygotic embryos, forming a complete plant structure rather than as sequentially formed organs. As with organogenesis, somatic embryos are cultivated in sterile media under controlled light or dark conditions. This method has been successfully used in several hardwood and conifer species (Becwar, 1993; Dunstan et al., 1995).

Stable genetic transformation using somatic embryogenesis has been achieved in some conifer species through techniques such as co-cultivation of embryogenic cell lines with **Agrobacterium** or biolistic particle bombardment of embryonic tissues.

Somatic embryogenesis is particularly advantageous for producing large numbers of genetically identical plantlets from a single transformed cell, making it an effective method for mass propagating transgenic trees. This technique is especially beneficial for species, like conifers, that are challenging to propagate using conventional methods.

8.7 FUTURE DIRECTIONS AND CHALLENGES

As genetic engineering in forest trees continues to advance, the focus is shifting toward developing trees that are not only commercially valuable but also ecologically beneficial. For instance, researchers are exploring the possibility of engineering trees to sequester more carbon, thereby contributing to climate change mitigation. Additionally, there is interest in creating trees that can remediate polluted environments through phytoremediation, where plants are used to absorb and detoxify contaminants from soil and water.

However, applying genetic engineering to forest trees is not without challenges. Regulatory hurdles, public perception, and potential ecological impacts must be carefully considered. Ongoing research addresses these concerns through comprehensive environmental risk assessments and developing strategies to ensure the safe and responsible use of genetically engineered trees in forestry.

The primary risks associated with transgenic trees include the unintended spread of transgenes to native populations and the instability of gene expression. To mitigate these risks, scientists focus on achieving reproductive sterility and ensuring stable gene expression, though both present significant challenges in genetic engineering. Despite these difficulties, forest biotechnology's potential economic and environmental benefits are substantial. However, there is concern that reliance on genetically engineered trees could reduce the diversity of local varieties, especially in developing countries, if these new varieties outcompete traditional ones.

Although transgenic forest trees have been tested in numerous field trials worldwide, their commercial use is limited, with China being the only country with commercialized genetically engineered poplars. In contrast, genetically engineered crops have been grown commercially in many countries for nearly two decades. The slow adoption of transgenic trees is due to unresolved biosafety and regulatory issues, which are more complex than those for annual crops because trees have longer life cycles and interact differently with their environments. In the United States, the USDA Animal and Plant Health Inspection Service (APHIS) regulates the release of genetically engineered trees, assessing the process of gene transfer and the resulting plant phenotype.

There are concerns about the commercialization of genetically engineered trees centered on the potential for gene flow from transgenic plantations to natural forests, which could negatively impact native tree populations. There are two main strategies to address these concerns: the biosafety approach, which focuses on ensuring absolute sterility before release, and the ecological approach, which assumes some gene escape is inevitable and aims to minimize the invasiveness of transgenic offspring. Both strategies require careful consideration of the long-term impacts on forest ecosystems.

Ultimately, the benefits of genetically engineered trees must outweigh the risks for commercialization to proceed. Current regulatory frameworks are often restrictive, limiting the scope and duration of research on transgenic trees. This regulatory caution reflects the many unknowns about the long-term effects of genetic engineering on forest ecosystems. As a result, there is a need for regulatory reforms that balance safety concerns with the potential for innovation in forestry practices.

8.8 SUMMARY AND CONCLUSIONS

Genetic improvement in forest trees has traditionally relied on selective breeding. Still, the introduction of genetic engineering (GE) over the past two decades has opened new possibilities for modifying tree genomes. Unlike traditional breeding, GE allows for the introduction of foreign genes that can confer new traits not naturally present in the species. This can be achieved through indirect methods, such as using **Agrobacterium** to transfer DNA, or direct methods, like biolistics. Both approaches have been successfully applied to various forest tree species, including angiosperms and conifers.

The potential applications of GE in forest trees include lignin modification, herbicide tolerance, resistance to diseases and pests, and control of flowering times. While field trials of transgenic trees with these traits are ongoing, achieving stable transgene expression over the long lifecycles of trees remains a challenge. Variability in transgene expression due to differences in gene integration patterns and copy numbers among transformants must be addressed to ensure consistent performance.

Given these complexities, optimizing gene transfer technologies to achieve stable and simple integration patterns is crucial. Additionally, regulatory oversight is essential for commercializing GE trees to ensure biosafety for humans and the environment. Concerns such as gene flow to native populations, seed dispersal, and the potential invasiveness of transgenic trees must be carefully managed before they can be released into the marketplace.

In conclusion, while genetic engineering holds promise for advancing forestry practices by introducing desirable traits, significant technical and regulatory challenges must be overcome. Addressing these issues will be key to successfully and safely deploying transgenic trees in forestry.

8.9 REFERENCES

Baucher, M., Monties, B., Montagu, M. V., & Boerjan, W. (1998). Biosynthesis and Genetic Engineering of Lignin. *Critical Reviews in Plant Sciences*, *17*(2), 125–197. https://doi.org/10.1080/07352689891304203

Becwar, M. R. (1993). Conifer Somatic Embryogenesis and Clonal Forestry. In M.-R. Ahuja & W. J. Libby (Ed.), *Clonal Forestry I* (s. 200–223). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-84175-0_13

Bruegmann, T., Fendel, A., Zahn, V., & Fladung, M. (2024). Genome Editing in Forest Trees. In A. Ricroch, D. Eriksson, D. Miladinović, J. Sweet, K. Van Laere, & E. Woźniak-Gientka (Ed.), *A Roadmap for Plant Genome Editing* (s. 347–372). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-46150-7_20

Brunner, A. M., Li, J., DiFazio, S. P., Shevchenko, O., Montgomery, B. E., Mohamed, R., Wei, H., Ma, C., Elias, A. A., & VanWormer, K. (2007). Genetic containment of forest plantations. *Tree Genetics & Genomes*, *3*, 75–100.

Doty, S. L., James, C. A., Moore, A. L., Vajzovic, A., Singleton, G. L., Ma, C., Khan, Z., Xin, G., Kang, J. W., Park, J. Y., Meilan, R., Strauss, S. H., Wilkerson, J., Farin, F., & Strand, S. E. (2007). Enhanced phytoremediation of volatile environmental pollutants with transgenic trees. *Proceedings of the National Academy of Sciences*, *104*(43), 16816–16821. https://doi.org/10.1073/pnas.0703276104

Dunstan, D. I., Tautorus, T. E., & Thorpe, T. A. (1995). Somatic Embryogenesis in Woody Plants. In T. A. Thorpe (Ed.), *In Vitro Embryogenesis in Plants* (Roč. 20, s. 471–538). Springer Netherlands. https://doi.org/10.1007/978-94-011-0485-2_12

Hackett, W. P. (2011). Juvenility, maturation, and rejuvenation in woody plants. *Horticultural reviews*, *7*, 109–154.

Harfouche, A., Meilan, R., & Altman, A. (2011). Tree genetic engineering and applications to sustainable forestry and biomass production. *Trends in biotechnology*, *29*(1), 9–17.

Merkle, S. A., Andrade, G. M., Nairn, C. J., Powell, W. A., & Maynard, C. A. (2007). Restoration of threatened species: A noble cause for transgenic trees. *Tree Genetics & Genomes*, *3*, 111–118.

Sheng, X., Mahendra, R. A., Wang, C.-T., & Brunner, A. M. (2023). CRISPR/Cas9 mutants delineate roles of *Populus FT* and *TFL1/CEN/BFT* family members in growth, dormancy release and flowering. *Tree Physiology*, *43*(6), 1042–1054. https://doi.org/10.1093/treephys/tpad027

Strauss, S. H., Rottmann, W. H., Brunner, A. M., & Sheppard, L. A. (1995). Genetic engineering of reproductive sterility in forest trees. *Molecular Breeding*, *1*(1), 5–26. https://doi.org/10.1007/BF01682086

Yao, T., Yuan, G., Lu, H., Liu, Y., Zhang, J., Tuskan, G. A., Muchero, W., Chen, J.-G., & Yang, X. (2023). CRISPR/Cas9-based gene activation and base editing in *Populus*. *Horticulture Research*, *10*(6), uhad085. https://doi.org/10.1093/hr/uhad085

9. GENE CONSERVATION

9.1 THREATS TO GENETIC DIVERSITY

A variety of human activities and environmental changes currently threaten genetic diversity. Habitat loss and fragmentation due to urbanization, agriculture, and deforestation isolate populations and reduce genetic exchange, leading to inbreeding and reduced diversity. Climate change exacerbates these effects by altering habitats and forcing species to adapt quickly or migrate, often resulting in decreased genetic diversity or even extinction for those unable to cope with the rapid changes.

Overexploitation, such as hunting, fishing, and poaching, poses significant threats by drastically reducing population sizes and creating genetic bottlenecks. Pollution, including chemical contaminants and ocean acidification, further reduces genetic diversity by causing mutations, killing off sensitive species, and altering ecosystems. Additionally, invasive species outcompete native species or cause hybridization, leading to the loss of unique genetic traits.

Agricultural practices like monocultures and selective breeding significantly reduce genetic diversity in crops and livestock, making them more vulnerable to diseases and environmental changes. The spread of diseases, often accelerated by human activity and globalization, can also lead to sharp declines in populations lacking genetic resistance. These factors threaten the genetic diversity essential for species' adaptability and long-term survival.

Genetic variation is crucial for disease resistance because it provides a population with a broader range of genetic traits to help individuals withstand infections and adapt to changing environments. In a genetically diverse population, some individuals are likely to possess natural resistance to specific pathogens due to variations in their immune system genes or other protective traits. When a disease outbreak occurs, these resistant individuals are more likely to survive and reproduce, passing their beneficial genes to future generations. This process helps the population maintain or increase its overall disease resilience.

Without genetic variation, populations become more uniform in their genetic makeup, making them highly susceptible to diseases. If all individuals are genetically similar and a pathogen arises to which they are all vulnerable, the disease can spread rapidly, potentially wiping out the entire population. This lack of variation also limits the ability of a species to adapt to new or evolving pathogens, as there are fewer genetic options available for natural selection to act upon. Consequently, low genetic diversity increases the risk of significant population declines or even extinction due to disease.

In agriculture, maintaining genetic variation is especially important for crops and livestock, as it helps protect against widespread disease outbreaks that can threaten food security. Monocultures, or the cultivation of a single crop variety over large areas, reduce genetic diversity and make crops more vulnerable to pests and diseases. By preserving a wide range of genetic traits in agricultural species, breeders can develop new varieties resistant to emerging diseases, ensuring sustainable food production and ecosystem stability.

9.2 STRATEGIES TO CONSERVE GENETIC DIVERSITY

The primary objective of genetic conservation is to preserve a sufficient level of genetic diversity within targeted species, enabling them to adapt to unpredictable environmental changes. In essence, the conservation of forest tree genes secures the potential for future adaptation. This broad genetic variation is also a critical prerequisite for the long-term success of breeding programs, ensuring that species can respond effectively to selective pressures over time.

There are two main approaches to gene conservation: **in-situ** and **ex-situ** methods. These approaches can be utilized separately or in combination as part of a comprehensive gene conservation program.

In situ gene conservation involves the protection of natural habitats, including forests, wilderness areas, reserves, and national parks, to maintain tree populations within their original environments. This method offers the advantage of preserving numerous species and individual trees relatively cost-effectively. In unmanaged protected areas, natural selection and evolution shape species composition. However, this can sometimes lead to the loss of early successional species, necessitating active management to maintain these species within the ecosystem. One

potential drawback of in situ conservation is the risk of failure in countries/regions where protected areas are not adequately safeguarded, leading to possible loss of genetic diversity.

In contrast, **ex-situ gene conservation** entails the systematic sampling of native populations, followed by the storage of collected germplasm (such as seeds or pollen) under controlled conditions. This germplasm may be preserved in cold storage, within breeding populations (e.g., scions grafted into clone banks or breeding arboreta), or through genetic tests (e.g., seedlings planted in provenance or progeny tests). While ex-situ methods can be more expensive, they are particularly favored in specific situations: (1) when the species is part of an active breeding program that includes gene conservation as an objective, or (2) when in situ gene conservation reserves cannot be sufficiently protected.

Both *in-situ* and *ex-situ* gene conservation require long-term conservation strategies. These strategies should encompass: (1) Sufficient population sizes maintaining genetic diversity; (2) The number and locations within *in situ* populations and sampled *ex-situ* collections to sample the species' diversity properly; and (3) Recurrent cycles of conservation efforts that ensure multi-generational programs.

The number of individuals required for effective conservation can vary depending on the species' genetic characteristics. A relatively small number of individuals (around 50 trees) may be sufficient to ensure continued response to selection and maintain high-frequency alleles in conserved populations. However, larger populations, potentially comprising several thousand individuals, are necessary to retain multiple copies of low-frequency alleles.

Effective conservation planning benefits from thoroughly understanding the species' genetic architecture, which can be estimated using molecular markers or provenance-progeny tests. Since many tree species exhibit substantial genetic variation among populations, it is essential to strategically protect and sample multiple populations across different geographical locations to ensure comprehensive coverage of gene diversity.

Forest management practices, including harvesting, can generally be conducted without significantly impacting genetic diversity, provided adequate regeneration measures are in place. Studies have shown that most shelterwood and seed tree cuts do not greatly reduce genetic variation in new stands compared to the parental stand, although some rare alleles may be absent. It is crucial to retain high-quality phenotypes to provide seeds for regeneration and avoid dysgenic selection. When clear-cuts are replanted with seedlings, these new plantations may exhibit higher levels of genetic diversity than the preceding forests, particularly if the seeds are sourced from many mother trees across several stands.

In conclusion, genetic conservation strategies, whether in situ or ex-situ, play a vital role in preserving forest tree species' adaptive potential and long-term viability, ensuring their ability to thrive due to environmental changes and ongoing evolutionary processes.

9.2.1 DESCRIPTION OF EX-SITU METHODS:

Ex situ genetic resources for forest trees include various methods for preserving genetic material outside their natural habitats, such as seed and pollen banks, breeding populations, genetic tests, field conservation banks, and breeding arboreta. Trees in urban arboreta and botanical gardens also serve as ex-situ collections, though they often have a narrow genetic base. Less commonly used methods for forest trees include tissue culture and DNA storage in genomic libraries, which could become more prevalent with technological advancements. Among these methods, seed storage is particularly widespread, used both for conserving endangered populations and for species involved in breeding programs. These seed collections are typically organized into a core set, representing the genetic diversity of the species, and a reserve set for reforestation and breeding purposes. Seed storage is cost-effective and can maintain viability for many years under appropriate conditions, such as subfreezing temperatures. However, some species, particularly those with recalcitrant seeds like certain oaks and eucalypts, cannot be stored for long periods. Additionally, seed banks require periodic replenishment to maintain viability and do not evolve with natural selection, posing some limitations to their use.

Pollen storage is less common in forest tree conservation but may be used in breeding programs for controlled pollination. Pollen is generally not as useful as seed for conservation because its genetic diversity is only realized

when fertilizing seeds. For species with recalcitrant seeds, pollen storage might be the only feasible method to preserve a wide range of genetic material. Genetic tests and breeding arboreta, including progeny and provenance tests, are valuable ex-situ resources. These tests help assess adaptive genetic variation and guide the establishment of breeding zones and seed transfer guidelines. Progeny tests, where the genetic identities of offspring are known, are especially useful for conserving genetic material, as they are typically planted in multiple locations, exposing them to various environmental conditions over many years. This helps researchers identify traits that may be beneficial in future breeding efforts, especially in response to new ecological challenges or diseases.

A good example of ex-situ gene conservation using genetic tests is CAMCORE (the organization presented later in this chapter), which has collected seeds from over 35 tropical tree species and established provenance-progeny tests across multiple locations. For species like *Pinus tecunumanii*, these tests involve planting thousands of trees representing a broad spectrum of the species' natural genetic diversity. Although genetic tests provide valuable insights into the adaptability and resilience of tree species, they have a finite lifespan. They must be part of a broader strategy for long-term conservation, including developing advanced-generation strategies for continual genetic testing and breeding. For detailed information about specific conservation programs, see Eriksson et al. (2008); White et al. (2007), and Zobel & Talbert (1984).

9.3 ORGANIZED GENE CONSERVATION WORLDWIDE:

The increasing world population and subsequent pressure to utilize remaining natural forests to sustain this population have heightened the awareness and need to conserve the genetic resources of forest trees. Tree improvement programs have been established for commercially important species, and several countries have foreign aid agencies that have helped in exploratory gene conservation efforts. An excellent example of a formal gene conservation organization is the **Central American and Mexican Coniferous Resources Cooperative (CAMCORE),** founded in 1980 (Dvorak et al., 1996).

1. The European Forest Genetic Resources Programme (EUFORGEN)

EUFORGEN (https://www.euforgen.org/) is a pan-European program that promotes the conservation and sustainable use of forest genetic resources. It was established in 1994 under the auspices of the United Nations and has been active in coordinating the efforts of European countries to conserve the genetic diversity of forest trees. The program focuses on in-situ and ex-situ conservation methods and has developed a network of gene conservation units across Europe.

Target Species: A wide range of forest tree species native to Europe, including Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), and European beech (*Fagus sylvatica*). The program is divided into five operating networks that focus on conifers, Mediterranean oaks, *Populus nigra*, temperate oaks and beech, and other hardwoods.

2. The North American Forest Genetic Resources Program (NAFGRP)

NAFGRP is a collaborative initiative between Canada, the United States, and Mexico to conserve forest genetic resources in North America. It seeks to promote the conservation of genetic diversity in forest trees through research, gene conservation strategies, and information sharing. The program works closely with national and regional conservation initiatives and focuses on in-situ and ex-situ conservation.

Target Species: Includes species like Douglas-fir (*Pseudotsuga menziesii*), eastern white pine (*Pinus strobus*), and sugar maple (*Acer saccharum*).

3. The Australian Tree Seed Centre (ATSC)

The ATSC, operated by CSIRO (Commonwealth Scientific and Industrial Research Organisation), collects, conserves, and supplies seeds from a wide range of Australian native trees and shrubs. The center focuses on ex-situ conservation through seed banking and providing material for research and restoration projects. It also engages in international collaboration for the conservation of globally significant species.

Target Species: Native Australian species like Eucalyptus, Acacia, and Casuarina, as well as species from other regions important for forestry and restoration.

4. The Center for Forest Conservation Genetics at the University of British Columbia

This center focuses on conserving genetic diversity in forest trees in Canada, particularly in the province of British Columbia. The center researches the genetic diversity of tree species, develops conservation strategies, and provides guidance on forest management practices that support genetic conservation. The center collaborates with government agencies, conservation organizations, and the forestry industry.

Target Species: Includes species like western red cedar (Thuja plicata), Sitka spruce (Picea sitchensis), and lodgepole pine (Pinus contorta).

5. The Tropical Tree Seed Bank at the Royal Botanic Gardens, Kew

This famous institution (https://www.kew.org/) has a global outreach focusing on tropical regions. As part of the Millennium Seed Bank Partnership, the Tropical Tree Seed Bank at Kew Gardens aims to conserve the genetic diversity of tropical tree species through ex-situ seed banking. This initiative is critical for preserving tropical forests, which are highly biodiverse but increasingly threatened by deforestation and climate change. The program also works on capacity building and training in tropical countries to support local conservation efforts.

Target Species: A wide range of tropical tree species from various regions, including Africa, Southeast Asia, and South America.

6. FAO

The Food and Agriculture Organization (FAO) (https://www.fao.org/home/en) provides countries with comprehensive technical and scientific support on forest gene conservation. This support encompasses planning and implementing strategies to preserve genetic diversity in forest species, which is crucial for maintaining resilient and sustainable forest ecosystems. Its Panel of Experts guides the FAO's initiatives in forest gene conservation. It involves collaboration with various international organizations, including the Danida Forest Seed Centre (DFSC), the International Genetic Resources Institute (IGPRI), the International Union of Forest Research Organizations (IUFRO), and the World Conservation Union (IUCN).

FAO's activities in this domain include promoting, exploring, conserving, and testing forest species and their provenances across national boundaries. This ensures that genetic resources are preserved and utilized sustainably, regardless of geographical and political boundaries. Current programs managed by FAO include coordinating the evaluation of ex-situ (off-site) conservation stands in 12 developing countries and in situ (on-site) conservation efforts in three countries. FAO aims to maintain genetic diversity within and among tree species through these programs, essential for adapting to environmental changes and sustaining forest health and productivity.

Additionally, FAO produces a variety of documents and guidelines that provide strategic frameworks and technical guidance for gene conservation efforts. These publications serve as valuable resources for policymakers, researchers, and practitioners involved in the worldwide conservation and sustainable use of forest genetic resources.

7. CAMCORE

CAMCORE (https://camcore.cnr.ncsu.edu/) is the International Tree Conservation and Domestication, established in 1980 and based at North Carolina State University (NCSU). It is one of the oldest and most successful forest gene conservation programs. This cooperative is supported by government and industry members from 14 countries. Its mission is "to conserve, test, and improve forest species in the tropics and subtopics for the benefit of humankind." To accomplish this, CAMCORE collects genetic material from natural populations, establishes the material in provenance and progeny trials, and evaluates the tests for breeding objectives. A strategy is then to link breeding and conservation objectives. Conservation is ex-situ and involves designing and maintaining genetic tests as long-term conservation plantations outside of the species' native range. Seed collected from the ex-situ genetic tests is subsequently being reintroduced into the original donor countries. The cooperative has established programs to conserve the genetic resources of 36 tree species (White et al., 2007).

These programs are examples of how different regions and organizations are working to conserve the genetic diversity of forest trees, each focusing on the specific needs and challenges of their target species and ecosystems. In summary, EUFORGEN, CAMCORE, and the Center for Forest Conservation Genetics focus primarily on in situ strategies, ATSC and Kew Gardens on ex-situ strategies, while NAFGRP incorporates both approaches.

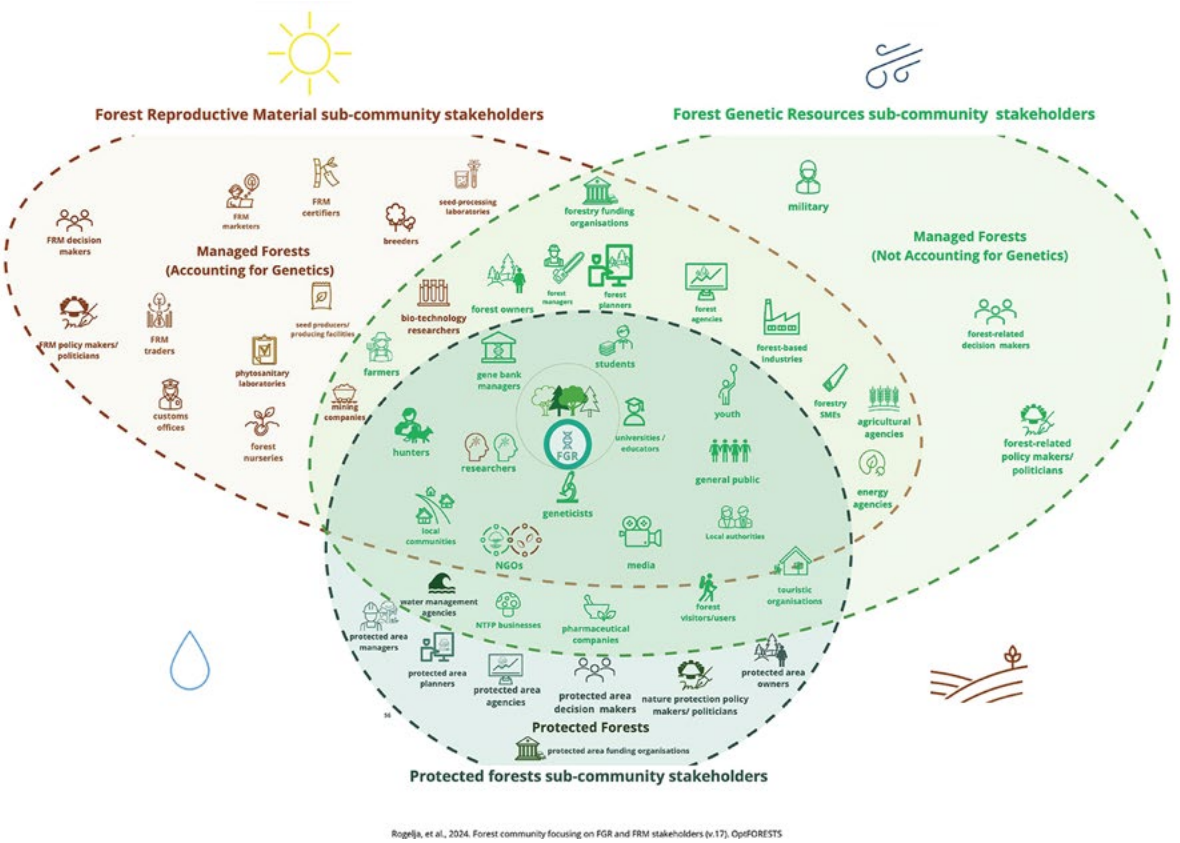
9.4 REFERENCES

Dvorak, W. S., Donahue, J. K., & Vasquez, J. A. (1996). Provenance and progeny results for the tropical white pine, Pinus chiapensis, at five and eight years of age. *New Forests*, *12*(2), 125–140. https://doi.org/10.1007/BF00036625

Eriksson, G., Ekberg, I., & Clapham, D. (2008). *An Introduction to Forest Genetics*. https://res.slu.se/id/publ/22529

White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI.

Zobel, B., & Talbert, J. (1984). *Applied forest tree improvement*. Wiley.



Rogelja, et al., 2024. Forest community focusing on FGR and FRM stakeholders (v.17). OptFORESTS

Name: INTRODUCTION TO FOREST GENETICS – For master students of forestry engineering
Author:doc. Ing. Jan Stejskal, Ph.D.
Publisher: Czech University of Life Sciences Prague
Approved by the Editorial Board
Number of pages: 150
Year of publication: 2026
ISBN: 978-80-213-3417-5
Publisher: Czech University of Life Sciences Prague, Faculty of Forestry and Wood Sciences

INTRODUCTION TO FOREST GENETICS

doc. Ing. Jan Stejskal, Ph.D.

2026

9 788021 334175