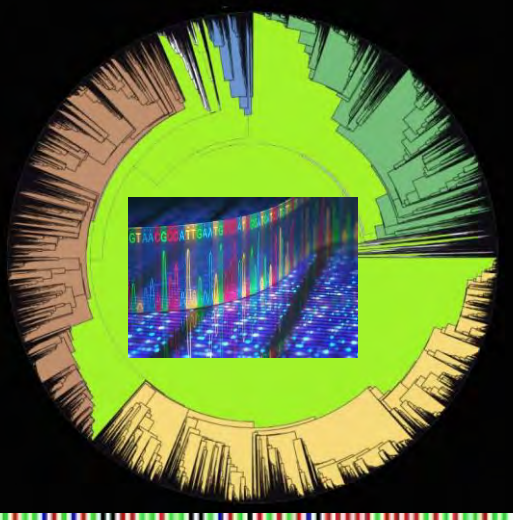
ADVANCE EXPERIMENTAL TAXONOMY



Dr. M. Ajmal Ali, PhD, FEHT
Professor
Department of Botany and Microbiology
College of Science, King Saud University
P Box 2455,Riyadh-11451, Saudi Arabia
Email: alimohammad@ksu.edu.sa

Course Specification

(Postgraduate Programs)

Course Title:	*Advanced Experimental Taxonomy*
Course Code:	*BOT621*
Program:	*Ph.D. in Botany*
Department:	*Botany and Microbiology*
College:	*Science*
Institution:	*King Saud University*
Version:	4th
Last Revision Date:	2/2/2026

Table of Contents

A. General information about the course:

1. Course Identification:

1. Credit hours: 2 (2+0)

2. Course type

A. ☐University ☐College ☒ Department ☐Track

B. ☐ Required ☒ Elective

3. Level/year at which this course is offered: (1st level/ 1st year)

4. Course General Description:

The course will detail the theory and practices of classic and modern plant systematics.

5. Pre-requirements for this course (if any):

None

6. Pre-requirements for this course (if any):

None

7. Course Main Objective(s):

To introduce the students to the explanation of advances in tools and techniques, and relevance of molecular plant taxonomy to understand the principles of Plant taxonomy

2. Teaching Mode: (mark all that apply)

No	Mode of Instruction	Contact Hours	Percentage
1	Traditional classroom	**30**	**100%**
2	E-learning		
3	Hybrid • Traditional classroom • E-learning	**0**	**0**
4	Distance learning	**0**	**0**

3. Contact Hours: (based on the academic semester)

No	Activity	Contact Hours
1.	**Lectures**	30
2.	**Laboratory/Studio**	

3.	Field	
4.	Tutorial	
5.	Others (specify).....	
	Total	30

Code	Course Learning Outcomes	Code of CLOs aligned with program	Teaching Strategies	Assessment Methods
1.0	Knowledge and understanding. At the end of the program, the graduate will be able to			
1.1	Analyze taxonomic evidence derived from morphological, anatomical, chemical, and molecular data to resolve complex taxonomic problems.	K2, K3	Advanced lectures, guided reading, case studies, and group discussions	Written reports, assignments, and a final exam
1.2	Evaluate contemporary theories, methodologies, and technologies in experimental and molecular plant taxonomy.	K3, K4	Seminars, literature review sessions, student-led discussions	Literature review report, oral presentation
2.0	Skills			
2.1	Synthesize scientific literature in plant taxonomy and molecular systematics to develop research-based arguments.	S3, S4	Journal club, guided reading, group discussions, research writing workshops	Literature review paper, research proposal mini-project, written assignment rubric
2.2	Communicate scientific findings effectively through oral presentations, written reports, and scholarly discussions.	S4, S6	Student-led presentations, peer-review sessions, academic writing workshops, and seminar discussions	Oral presentation rubric, written report rubric, participation and discussion assessment
3.0	Values, autonomy, and responsibility			
3.1	Apply ethical principles and research integrity standards in taxonomic and molecular research.	V1	Ethics case studies, research integrity workshops, guided discussions, policy brief reviews	Ethics reflection report, plagiarism check, rubric-based assessment, participation evaluation
3.2	Collaborate effectively to plan and execute advanced research tasks,	V2	Group research projects, collaborative learning	Peer evaluation rubric, group project report, participation

	demonstrating accountability and professional conduct.		activities, mentoring sessions, peer-learning workshops	assessment, teamwork performance rubric

B. Course Learning Outcomes (CLOs), Teaching Strategies and Assessment Methods:

C. Course Content:

No	List of Topics	Contact Hours
1.	Principles and practices of Plant taxonomy	2
2.	Taxonomic circumscription and the need for taxonomic evidences	4
3.	Structural information as a source of taxonomic evidences	4
4.	Non-structural information as a source of taxonomic evidences	6
5.	Molecular Systematics (tools and techniques of sampling to sequencing), Choosing a molecular marker for phylogenetic analyses	6
6.	Phylogenetic analyses: DNA sequences as taxonomic evidence, Interpretation of molecular phylogenetic trees	4
7.	Assessment of genetic diversity using molecular data, DNA barcoding	4
	Total	30

D. Students Assessment Activities:

No	Assessment Activities *	Assessment timing (in week no)	Percentage of Total Assessment Score
1.	1st presentation	4th	20%
2.	2ND presentation	8th	20%
3.	3ed presentation	12th	20%
4.	Final exam	14th -15th	40%

*Assessment Activities (i.e., Written test, oral test, oral presentation, group project, essay, etc.)

E. Learning Resources and Facilities:

1. References and Learning Resources:

Essential References	Michael G. Simpson (2010) Plant Systematics. Elsevier Science Publishing Co Inc, San Diego, United States (ISBN10 012374380X).
Supportive References	
Electronic Materials	▪ American Journal of Botany (http://www.amjbot.org)

	▪ Botanical Journal of the Linnaean Society (http://www.blackwellpublishing.com/jnl_default.asp) ▪ Molecular Biology & Evolution (http://mbe.oupjournals.org) ▪ Molecular Phylogenetics & Evolution (http://www.elsevier.com) ▪ Systematic Botany (http://www.sysbot.org/) ▪ Taxon (http://www.botanik.univie.ac.at/iapt/s_taxon.php) ▪ http://www.plantsystematics.org/index.html ▪ The plant list (http://www.theplantlist.org/) ▪ NCBI (https://www.ncbi.nlm.nih.gov/) ▪ Phylogeny programs (http://evolution.genetics.washington.edu/phylip/software.html) ▪ Flora of Saudi Arabia (http://plantdiversityofsaudiarabia.info/Biodiversity-Saudi-Arabia/Flora/Flora.htm) ▪ https://www.tropicos.org/home ▪ Online Resources for Taxonomic Research (https://stories.rbge.org.uk/archives/1002) ▪ Eflora (http://www.efloras.org/flora_page.aspx?flora_id=2) ▪ Royal Botanic Gardens, Kew (https://www.kew.org/)
Other Learning Materials	**Molecular phylogenetic analysis software (BioEdit, ClustalX, MEGA) is available with the instructor.**

2. Educational and Research Facilities and Equipment Required:

Items	Resources
facilities (Classrooms, laboratories, exhibition rooms, simulation rooms, etc.)	**Classrooms, laboratories**
Technology equipment (Projector, smart board, software)	**Projector & smart board**
Other equipment (Depending on the nature of the specialty)	• **A computer with a digital camera and suitable programs** **-New laboratory equipment for plant surface**

F. Assessment of Course Quality:

Assessment Areas/Issues	Assessor	Assessment Methods
Effectiveness of teaching	The students	Indirect
Effectiveness of students assessment	Faculty	Direct
Quality of learning resources	Faculty & students	Indirect
The extent to which CLOs have been achieved	Program Leaders	Direct
Other		

Assessor (Students, Faculty, Program Leaders, Peer Reviewer, Others (specify)
Assessment Methods (Direct, Indirect)

G. Specification Approval Data:

COUNCIL/ COMMITTEE	ACADEMIC ACCREDITATION COMMITTEE OF THE DEPARTMENT OF BOTANY AND MICROBIOLOGY
REFERENCE NO.	
DATE	2 FEBRUARY 2026

Course Assignment Details – BOT 621

Semester: 1447 AH – 2nd Semester

Assignment Components

Assignment Title	Mode of Submission	Marks	Notes
Research project on a topic related to **Experimental Taxonomy**	Written / Oral presentation	**30**	—
• Report writing	Written	**15**	—
• Oral presentation	Oral	**10**	—
• Academic integrity	< %15	**5**	—
Submission of **15 identified herbarium specimens**	Practical	**30**	Group cooperation in the collection and identification; each student submits individually
Final Exam	Written	**40**	—
Total		**100**	

Research Topics and Student Names

Research Topic	Student Name
Taxonomic Keys	
Scientific Nomenclature	
Numerical Taxonomy	
Anatomical Taxonomy	
Chemical Taxonomy	

A BRIEF NOTE

1. Introduction

Taxonomy is the fundamental biological discipline concerned with the identification, nomenclature, and classification of organisms. In plant sciences, taxonomy plays a pivotal role in understanding biodiversity, evolutionary relationships, and the sustainable utilization of plant resources. Traditionally, plant taxonomy relied heavily on morphological and anatomical characters. However, with the rapid advancement of experimental techniques and molecular biology, taxonomy has transformed into a highly interdisciplinary science. The integration of classical taxonomy with experimental and molecular tools has given rise to what is now termed **advanced experimental taxonomy**.

Advanced experimental taxonomy emphasizes the use of controlled experiments, cytological analyses, biochemical markers, and molecular data to resolve complex taxonomic problems. It moves beyond superficial morphological resemblance and seeks to uncover evolutionary and genetic relationships among taxa. This modern approach is particularly valuable in groups exhibiting high phenotypic plasticity, cryptic species, hybridization, and polyploidy, where traditional methods often fail to provide clear species boundaries.

The current era of taxonomy is characterized by the application of DNA sequencing, molecular phylogenetics, and next-generation sequencing (NGS) technologies. These tools have revolutionized species delimitation, phylogenetic reconstruction, and biodiversity assessment. As a result, experimental taxonomy today is not only descriptive but also predictive, analytical, and integrative, forming the backbone of modern systematics, conservation biology, and evolutionary studies.

2. Classical Plant Taxonomy

2.1 Concept and Scope

Classical plant taxonomy, also known as traditional or alpha taxonomy, is based primarily on observable morphological characters. These include features such as habit, leaf shape, floral structure, inflorescence type, fruit and seed morphology, and anatomical traits. Classical taxonomy laid the foundation for plant classification systems developed by renowned taxonomists such as Linnaeus, Bentham and Hooker, and Engler and Prantl.

The major strength of classical taxonomy lies in its simplicity, cost-effectiveness, and field applicability. It enables botanists to identify plants using dichotomous keys, floras, herbarium specimens, and monographs. Even today, classical taxonomy remains indispensable, especially for field surveys, biodiversity inventories, and ecological studies.

2.2 Limitations of Classical Taxonomy

Despite its importance, classical taxonomy has several limitations. Morphological characters can be influenced by environmental conditions, leading to phenotypic plasticity. Closely related species may exhibit very similar morphology, while unrelated species may converge morphologically due to similar ecological pressures. Such situations often result in misidentification, synonymy, or artificial groupings. Moreover, classical taxonomy struggles to resolve taxonomic problems in cases of hybridization, polyploidy, cryptic species complexes, and asexual reproduction. In these scenarios, experimental and molecular approaches become essential to provide objective and reproducible evidence.

3. Taxonomic Evidences in Experimental Taxonomy

Advanced experimental taxonomy relies on multiple lines of evidence to establish natural relationships among taxa. These evidences can be broadly categorized into morphological, cytological, biochemical, and molecular data.

3.1 Morphological and Anatomical Evidence

Morphological evidence remains the primary layer of taxonomic study. However, experimental taxonomy emphasizes quantitative and statistical analysis of morphological traits rather than subjective descriptions. Microscopic studies of epidermal features, stomatal types, trichomes, vascular anatomy, and embryological characters provide deeper insights into species relationships.

3.2 Cytological Evidence

Cytotaxonomy involves the study of chromosome number, structure, behavior, and genome organization. Chromosome counts, karyotype analysis, and meiotic behavior are valuable tools for understanding evolutionary trends and species differentiation. Polyploidy, aneuploidy, and chromosomal rearrangements often play a crucial role in plant speciation.

Cytological data are especially important in resolving species complexes and identifying hybrid species. Differences in chromosome number or pairing behavior can indicate reproductive isolation and evolutionary divergence.

3.3 Biochemical and Chemotaxonomic Evidence

Chemotaxonomy is based on the distribution of chemical compounds such as alkaloids, flavonoids, terpenoids, phenolics, and secondary metabolites. These compounds are genetically controlled and relatively stable, making them reliable taxonomic markers.

For example, the presence or absence of specific alkaloids has been used to classify families such as Solanaceae and Papaveraceae. Protein profiling and enzyme polymorphism also contribute to understanding genetic variation and species boundaries.

4. DNA Sequencing and Its Role in Taxonomy

4.1 Concept of DNA Sequencing

DNA sequencing refers to the determination of the precise order of nucleotides in a DNA molecule. In taxonomy, DNA sequences serve as molecular signatures that reflect evolutionary history and genetic divergence among organisms.

The development of Sanger sequencing in the late 20th century marked the beginning of molecular taxonomy. It enabled researchers to compare DNA regions across species and construct phylogenetic trees based on sequence similarity.

4.2 Molecular Phylogenetics

Molecular phylogenetics uses DNA sequences to infer evolutionary relationships. Commonly used genetic markers in plants include nuclear genes (ITS), chloroplast genes (rbcL, matK), and mitochondrial genes. These markers provide varying levels of resolution depending on the taxonomic level.

Phylogenetic trees generated from DNA data often reveal natural groupings that differ from traditional classifications. Many plant families and genera have been reclassified based on molecular evidence, leading to more natural and evolutionarily meaningful systems.

5. Molecular Taxonomy and DNA Barcoding

5.1 Molecular Taxonomy: An Overview

Molecular taxonomy is the application of molecular data, particularly DNA sequences, to taxonomic classification and identification. It offers several advantages over classical methods, including objectivity, reproducibility, and the ability to detect cryptic diversity.

Molecular taxonomy has become indispensable in modern systematics, especially for groups with limited morphological variation, such as algae, fungi, bryophytes, and microorganisms.

5.2 DNA Barcoding

DNA barcoding is a standardized method for species identification using short, conserved DNA sequences. In plants, the most widely accepted barcode regions are rbcL and matK, often supplemented by ITS or trnH-psbA.

The principle of DNA barcoding is based on the concept that genetic variation within a species is smaller than variation between species. Thus, a barcode sequence can serve as a unique identifier, similar to a supermarket barcode.

5.3 Applications of DNA Barcoding

DNA barcoding has numerous applications in taxonomy and beyond:

Rapid species identification in biodiversity surveys

Detection of cryptic and invasive species

Authentication of medicinal plants and herbal products

Conservation and monitoring of endangered species

Forensic analysis and environmental DNA studies

Barcoding has significantly accelerated the pace of taxonomic research and reduced dependence on expert morphological identification.

6. Next-Generation Sequencing (NGS) and Modern Taxonomy

6.1 Concept of Next-Generation Sequencing

Next-generation sequencing refers to high-throughput technologies capable of sequencing millions of DNA fragments simultaneously. Unlike traditional Sanger sequencing, NGS produces massive amounts of data in a short time and at lower cost.

Popular NGS platforms include Illumina, Ion Torrent, PacBio, and Oxford Nanopore. These technologies have transformed taxonomy into a data-rich, genome-scale science.

6.2 Applications of NGS in Taxonomy

NGS has opened new dimensions in experimental taxonomy:

6.2.1 Genome Sequencing

Whole-genome sequencing provides comprehensive genetic information, allowing precise species delimitation and evolutionary analysis.

6.2.2 Transcriptomics

RNA sequencing helps in understanding gene expression patterns and adaptive traits related to ecological specialization.

6.2.3 Phylogenomics

Phylogenomics uses hundreds or thousands of genes to reconstruct deep evolutionary relationships with high accuracy.

6.2.4 Environmental DNA (eDNA)

NGS enables detection of species from environmental samples such as soil, water, and air, revolutionizing biodiversity assessment and monitoring.

6.3 Advantages of NGS in Experimental Taxonomy

High resolution at species and population levels
Ability to detect hybridization and introgression
Discovery of new and cryptic species
Integration with bioinformatics and AI tools
Support for large-scale biodiversity projects

NGS has effectively shifted taxonomy from a descriptive science to a predictive and computational discipline.

7. Integration of Classical and Molecular Approaches

The future of taxonomy lies in integrative taxonomy, which combines morphological, cytological, biochemical, and molecular data. No single method is sufficient to resolve all taxonomic problems. Instead, multiple lines of evidence provide a more robust and reliable classification framework.

Classical taxonomy provides ecological and morphological context, while molecular tools offer genetic and evolutionary insights. Together, they enable accurate species delimitation, natural classification, and meaningful evolutionary interpretation.

8. Conclusion

Advanced experimental taxonomy represents the culmination of centuries of taxonomic research enriched by modern experimental and molecular techniques. While classical plant taxonomy remains the foundation, its integration with cytology, biochemistry, DNA sequencing, and next-generation technologies has revolutionized the field.

Molecular taxonomy and DNA barcoding have made species identification faster, more accurate, and more accessible. Next-generation sequencing has further expanded the scope of taxonomy to genome-level analysis, enabling unprecedented insights into plant evolution, diversity, and adaptation.

In the era of biodiversity loss, climate change, and emerging biological challenges, advanced experimental taxonomy plays a critical role in conservation, sustainable utilization of plant resources, and understanding life on Earth. It stands as a dynamic, interdisciplinary science essential for both fundamental research and applied biological studies.

SHORT LECTURE

Objective : The general course objective of the course advance experimental taxonomy is to introduce the students about the explanation of advances in tools and techniques, and relevance of molecular plant taxonomy in order to understand the principles of Plant taxonomy.

Lecture – Topics 10 and examinations

1. Principles and practices of Plant taxonomy
2. Taxonomic circumscription and need of taxonomic evidences
3. Structural information as a source of taxonomic evidences
4. Non-structural information as a source of taxonomic evidences
5. Molecular Systematics (tools and techniques of sampling to sequencing)
6. Choosing molecular marker for phylogenetic analyses
7. Phylogenetic analyses: DNA sequences as taxonomic evidences
8. Interpretation of molecular phylogenetic trees
9. Assessment of genetic diversity using molecular data
10. DNA barcoding

Criteria for evaluation of the students

- First midterm exam: 20 marks
- Second midterm exam: 20 marks
- Reports, presentation and assignments: : 20 marks
- Final Exam: : 40 marks

Essential References Materials

- American Journal of Botany (http://www.amjbot.org)
- Botanical Journal of the Linnaean Society (http://www.blackwellpublishing.com/jnl_default.asp)
- Molecular Biology & Evolution (http://mbe.oupjournals.org)
- Molecular Phylogenetics & Evolution (http://www.elsevier.com)
- Systematic Botany (http://www.sysbot.org/)
- Taxon (http://www.botanik.univie.ac.at/iapt/s_taxon.php)

- Website: http://www.plantsystematics.org/index.htmlSoftware: Molecular phylogenetic analysis software (BioEdit, ClustalX, MEGA4)

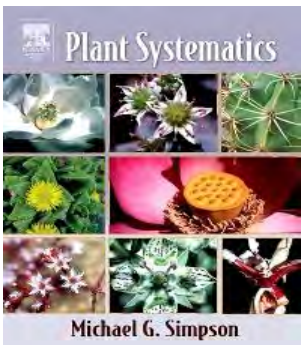
Course learning outcome

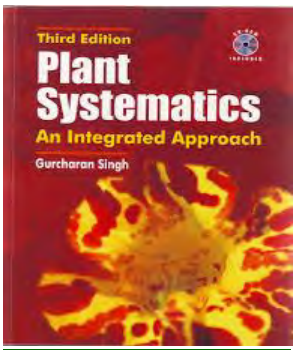
Knowledge:

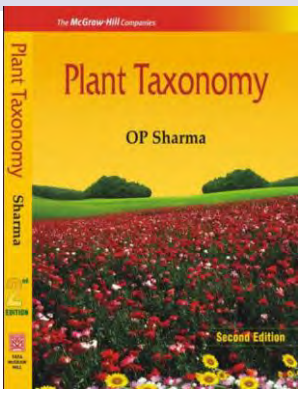
- The student will be able to relate the origin of the taxonomic questions.
- The student will be able to relate the tools and techniques used to gather taxonomic evidences to resolve the taxonomic questions.

Cognitive:

- The student will be able to relate the relevance of plant molecular taxonomy in order to understand the principles of Plant taxonomy by the practice of DNA sequence data analysis







- Michael G. Simpson (2010) Plant Systematics. Elsevier Science Publishing Co Inc, San Diego, United States (ISBN10 012374380X).
- Gurcharan Singh (2010) Plant Systematics: An Integrated Approach, Third Edition, CRC Press (ISBN 9781578086689).
- O.P. Sharma (2009)
- Plant Taxonomy, second edition, Tata McGraw-Hill Education Pvt. Ltd.,
- ISBN 10: 0070141592

Introduction about Plant Taxonomy / Systematics

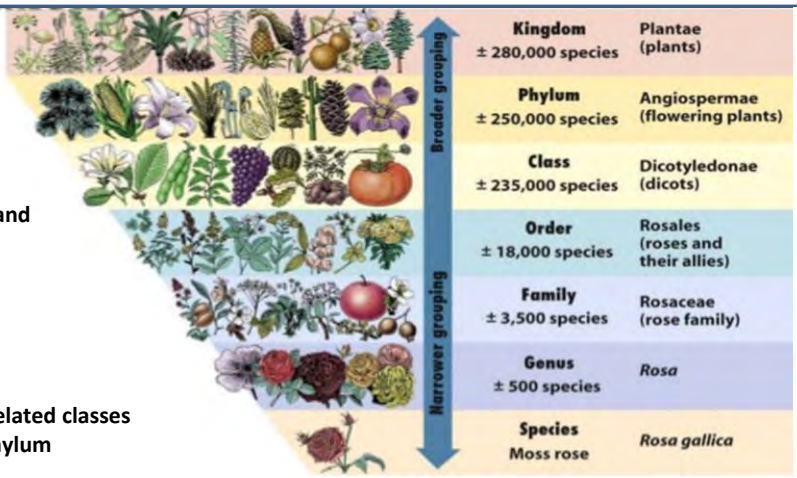
- Every Human being is a Taxonomist

Plant Biodiversity

We study plants because:

TAXONOMIC HIERARCHY

- Carrolus Linnaeus first adopted the hierarchic system of taxonomy classification in 1753.
- The succession groups are as follow:
- **Species**: Organisms sharing a set of biological traits and reproducing only their exact kind.
- The lowest major group, representing plants and animals referred to as Species.
- **Species is the fundamental unit in taxonomy**
- **Genus**: Genus are the closely related species
- **Family** : Family is the closely related genera
- **Order** : Order is the closely related families
- **Class** : Class are the closely related order
- **Division / Phylum**: Division or Phylum is the related classes
- **Kingdom**: Kingdom is the related Division / Phylum



Objective / Goals / Aims of Plant Taxonomy

- ❑ To provide an inventory of plant taxa for local, regional or continental needs.
- ❑ To establish suitable method for identification, nomenclature and description of plant taxa.
- ❑ Classification of organism into classes, Order, Families, Genera, and species
- ❑ To provide significantly valuable information concerning wild and medicinal species, endangered species, unique plants, genetic and ecological diversity

Scope of Taxonomy

- ❖ Taxonomy is one of the oldest sciences.
- ❖ It provides thorough knowledge of living species and their various forms.
- ❖ All the branches of biology are dependent on taxonomy for proper identification the species.
- ❖ It has been proceeded further incorporating data from phytochemistry, cyto-genetics supported by proper computation.

Basic components (Principles) of Plant Taxonomy / Plant Systematics

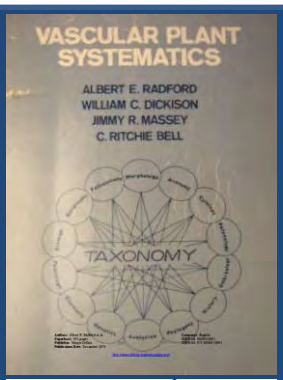
- Plant collection, Preservation and Documentation
- Plant Structure (Taxonomic Terminology, Taxonomic description of external and internal morphology)
- Taxonomic Identification
- Scientific Nomenclature / Botanical nomenclate : Nomenclature deals with the application of a correct name to a plant or a taxonomic group. Scientific names are necessary because the same common name is used for different plants in different areas of the world.
- Taxonomic Classification (History and Systems of Plant Classification)
- Taxonomic evidences / Source of data (Morphology, Anatomy, Embryology, palynology, Micromorphology, Chemistry, DNA etc.) in plant taxonomy

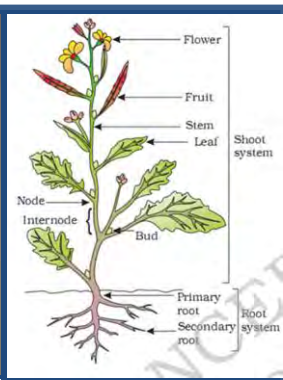
Kingdom: Plantae
Class: Angiosperms
Order: Arecales
Family: Arecaceae
Genus: *Phoenix*
Species: *P. dactylifera*

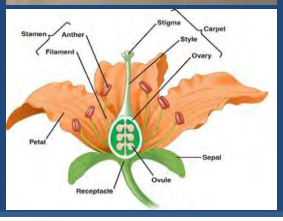
Types of Taxonomy / Taxonomic Studies / Plant Taxonomic Classification

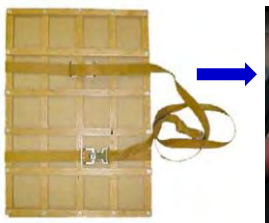
Alpha (α) Taxonomy / classical taxonomy:- It involves description and naming of organisms. It is the parent of other types of taxonomy.
Omega (Ω) Taxonomy: - It is the modern experimental taxonomy in which the taxonomic activities have been enriched with data from ecology, phyto-chemistry, phyto-geography, cyto-genetics and physiology coupled with adequate computation.
Herbarium: Plant collecting, Preservation and Documentation

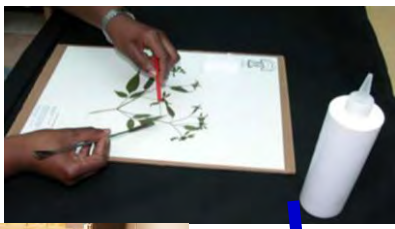
- **To make a herbarium specimen, the plant is collected, and notes are made about it. The plant is then pressed until dry between blotters that absorb moisture and mounted onto a herbarium sheet with a suitable label, and stored in steel cabinet arranged into some system of classification.**
- **Herbarium techniques involve : (i) Collection, (ii) Drying, (iii) Poisoning, (iv) Mounting, (v) Stitching, (vi) Labelling, and (vii) Deposition.**
- **Flora = it is the documentation of plants occurring in a particular region.**
- The FLORA is the main Resources of Taxonomic Information
- **A HERBARIUM is a collection of dried plants systematically named and arranged for ready reference and study.**



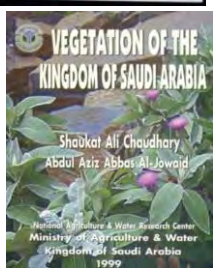


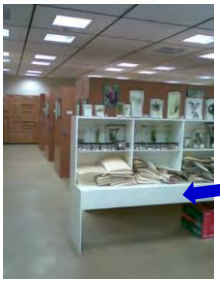




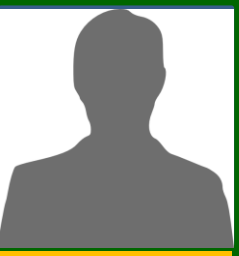


Phoenix dactylifera Linnaeus, Sp. Pl. 2: 1188. 1753.
Stems solitary or clustered and then with few shoots, to 30 m tall, to 50 cm in diam., rough with persistent, diamond-shaped leaf bases. Leaves 3-5 m; sheath and petiole to 1 m; rachis 1-2 m; acanthophylls many per side of rachis; pinnae to 200 per side of rachis, linear, irregularly arranged and spreading in different planes; middle pinnae to 40 × 2 cm. Male inflorescences erect, to 1 m, with many rachillae, these ca. 30 cm; female inflorescences erect, becoming pendulous, to 2 m, with to 150 rachillae, these to 40 cm. Fruits variable in shape, usually oblong, to 7 × 3 cm, brown or black; endosperm homogeneous.

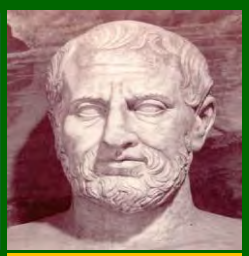




SYSTEM OF PLANT CLASSIFICATION

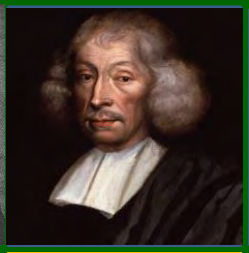


Preliterate Mankind / Folk taxonomies:



Theophrastus (372 BC to 287 BC):

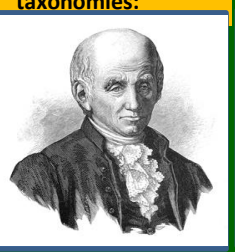
Andrea Cesalpino (1519-1603)



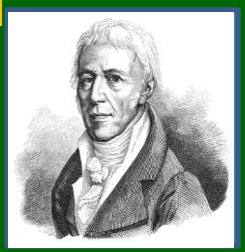
John Ray (1627-1705)

J. P. de Tournefort (1656-1708)

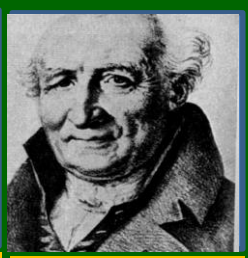
Carolus Linneaus (1753)



Michel Adanson (1727-1806)

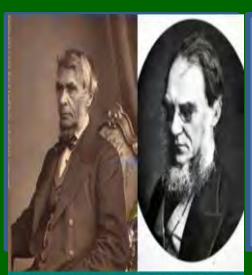


Jean B.P. Lamarck (1744-1829)



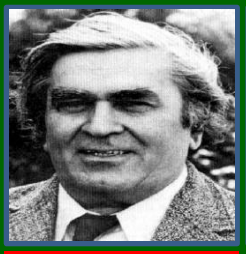
Antoine Laurent de Jussieu (1748-1836)

de Candolle (1778–1841)



Charles E. Bessey (1845-1915)

Auther Cronquist 1968



Armen Takhtajan 1969

Rolf Dahlgren (1932-87)

John Hutchinson (1884-1972)

APG Angiosperm Phylogeny Group (1998)

Bentham and Hooker System of Plant Classification

- Bentham and Hooker, two English botanists, represented the most well developed natural system of plant classification. The classification was published in a three-volume work Genera
- Hooker supervised the publication of Index Kewensis (2 volumes, 1893), listing the names of all known species and their synonyms.
- Many important herbaria of the world have specimens arranged according to Bentham and Hooker system of plant classification.
- Bentham and Hooker recognized three class:

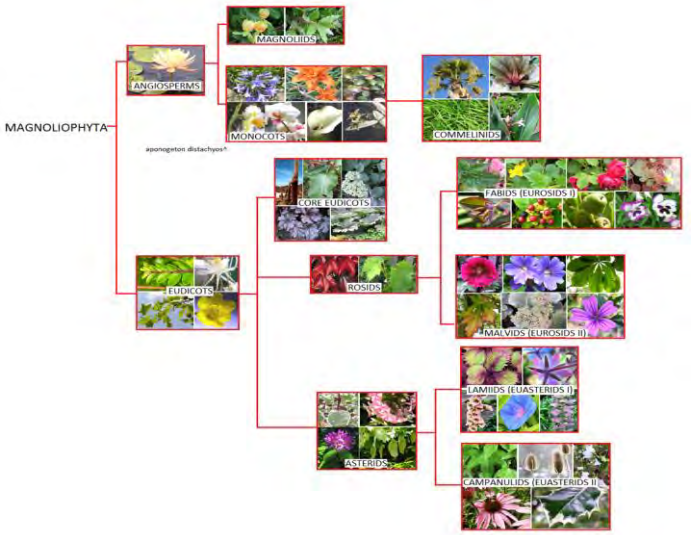
Class Dicotyledones:

Subclass POLYPETALE with three series Series 1. THALAMIFLORÆ, Series 2. DISCIFLORÆ, Series 3. CALYCIFLORÆ;
Subclass DICOTYLEDONES (GAMOPETALÆ) with three series that is Series 1. INFERÆ, Series 2. HETEROMERÆ, Series 3. BICARPELLATÆ, and
Subclass DICOTYLEDONES MONOCHLAMIDEÆ.

Class Gymnospermeæ (Gymnosperms are placed between Dicotyledons and Monocotyledons)

Class Monocotyledones

Angiosperm Phylogeny Group (APG)



- The APG system of flowering plant classification is the modern, mostly molecular-based, system of plant taxonomy for flowering plants (angiosperms) being developed by the Angiosperm Phylogeny Group (APG).
- The APG was first published in 2008.
- Currently the APG IV system recognizes a total of 64 angiosperm orders and 416 families.
- The families in APG classification have been grouped into 40 putative monophyletic orders under a small number of informal monophyletic higher groups: monocots, commelinoids, eudicots, core eudicots, rosids, eurosids I, eurosids II, asterids, euasterids I and euasterids II

SCIENTIFIC NOMENCLATURE / BOTANICAL NOMENCLATE :

Nomenclature deals with the application of a correct name to a plant or a taxonomic group.

- ❖ **We have millions of species distributed in different geographical regions of the world.**
- ❖ **The Scientific names (Botanical name and Zoological name)of the living organism (Plants and Animals) are necessary because the same common name is used for different plants / Animals in different areas of the world.**
- **Swedish Botanist Carolus Linnaeus introduced Binomial Nomenclature.**
- **The Binomial nomenclature uses two Latin words to indicate the genus and the species. The first word is the genus and the second word is the species. Example- the botanical name of Dates is *Phoenix dactylifera***

❖ **Binomial Nomenclature and Carolus Linneaus System of Plant Classification**

❖ **Taxonomic Systems of Classification: Ideally our systems of classification should allow us to place similar species of plants together in the same category.**

❖ **There are two types of Classification Schemes:**

❑ **Artificial taxonomy was a system of grouping unrelated plant species by a common criteria (i.e. a flowers sexual organs)**

❑ **Natural classification reflects relationships among taxon**

➢ **Carolus Linneaus was a Swedish botanist.**

➢ **Carolus Linneaus traveled to Lapland (Blue Lake, CA) and collected large number of plants.**

➢ **Carolus Linneaus introduced Binomial Nomenclature.**

Binomial nomenclature = Uses two Latin words to indicate the genus and the species. The first word is the genus and the second word is the species. Example- the botanical name of dates is *Phoenix dactylifera*

➢ **Carolus Linneaus published the book 'Species Plantarum' in 1753.**

➢ **Carolus Linneaus classified the plants based on the plant's method of reproduction and structure of reproductive parts.**

➢ **Produced his sexual system of classification (Artificial classification)**

➢ **Carolus Linneaus divided plants into 24 classes. The Classes in the Linneaus is based largely on the amount, union and length of stamens**

Species Concept

- o **Species is the basic unit of classification**
- o **Plants in the same species consistently produce plants of the same types**
- o **The name of the plants must should be written in italics. For example *Phoenix dactylifera***

TAXONOMIC RANKS OF LAND PLANTS	ENDING	EXAMPLE TAXON
Kingdom	(various)	**Plantae**
Phylum [Division]	**-phyta**	**Magnoliophyta**
Subphylum [Subdivision]	-phytina	Magnoliophytina
Class	**-opsida**	**Asteropsida**
Subclass	-idae	Asteridae
Order	**-ales**	**Asterales**
Suborder	-ineae	Asterineae
Family	**-aceae**	**Asteraceae**
Subfamily	-oideae	Asteroideae
Tribe	-eae	Heliantheae
Subtribe	-inae	Helianthinae
Genus	(various)	***Helianthus***
Subgenus	(various)	*Helianthus*
Section	(various)	*Helianthus*
Series	(various)	*Helianthus*
Species [abbr. sp. (sing.), spp. (pl.)]	(various)	***Helianthus annuus***
Subspecies [abbr. subsp. or ssp. (sing.), subspp. or sspp. (pl.)]	(various)	*Helianthus annuus* ssp. *annuus*
Variety [abbr. var. (sing.), vars. (pl.)]	(various)	*Helianthus annuus* var. *annuus*
Form [abbr. f.]	(various)	*Helianthus annuus* f. *annuus*

Classes

1. Monandria- stamen one
2. Diandria- stamens two
3. Triandria- stamens three
4. Tetrandria- stamens four
5. Pentandria- stamens five
6. Hexandria- stamens six
7. Heptandria- stamens seven
8. Octandria- stamens eight
9. Ennandria- stamens nine
10. Decandria- stamens ten
11. Dodecandria- stamens 11-19
12. Icosandria- stamens 20 or more, on the calyx
13. Polyandria- stamens 20 or more, on the receptacle
14. Didynamia- stamens didynamous; 2 short, 2 long
15. Tetradynamia- stamens tetradynamous; 4 long, 2 short
16. Monadelphia- stamens monadelphous; united in 1 group
17. Diadephia- stamens diadelphous; united in 2 groups
18. Polyadelphia- stamens polyadelphous; united in 3 or more groups
19. Syngenesia- stamens syngenesious; united by anthers only
20. Gynandria- stamens united with the gynoecium
21. Monoecia- plants monoecious
22. Dioecia- plants dioecious
23. Polygamia- plants polygamous
24. Cryptogamia- flowerless plants

International Code of Botanical Nomenclature (ICBN)

The current activity of botanical nomenclature is governed by the International Code of Botanical Nomenclature (ICBN) published by the International Association of Plant Taxonomy (IAPT).

The Code is divided into 3 divisions:

I. Principles

II. Rules and recommendations

III. Provisions for the governance of the Code

Principles of ICBN

- **Botanical Nomenclature is independent of Zoological Nomenclature. The Code applies equally to the names of taxonomic groups treated as plants whether or not these groups were originally so treated.**
- **The application of names of taxonomic groups is determined by means of nomenclatural types / TYPIFICATION.**
- **Nomenclature of a taxonomic group is based upon Priority Of Publication.**
- **Each taxonomic group with a particular circumscription, position and rank can bear Only One Correct Name, the earliest that is in accordance with the rules.**
- **Scientific names of taxonomic groups are treated as LATIN, regardless of derivation.**
- **The rules of nomenclature are Retroactive, unless expressly limited.**

- **Generic Name: The Generic name is usually a noun and singular, which is spelled or written with a capital letter.**
- **Specific Epithet: The specific epithet is often an adjective and it is written with a small initial letter.**
- **In the hand written manner, both the generic names and specific epithet should be underlined, while if printed it should be in italics.**

Synonyms and related terminology

- **Synonyms: A name rejected due to misuse or difference in taxonomic judgement.**
- **Basionym:**
  - **The basionym is the first name ever given to a taxon. Further studies and revisions may reject the basionym as the most correct one, but it still is useful as a nomenclatural reference for that species.**
  - **Also, according to the priority rules of the ICBN, after a taxonomic revision that results in a species being reclassified in another genus, the specific epithet must remain the same as the one in the Basionym.**
  - **A short example: Linnaeus classified the Tea Plant as *Thea sinensis*. Some decades later, Sweet noticed that the genus *Thea* was not really different from the genus *Camellia*, and renamed all the *Theas* as *Camellias*. *Thea sinensis* became *Camellia sinensis*, because he had to keep the specific epithet the same as the original name (Basionym) for that species, given by Linnaeus.**
- **Homonym: A case in which two or more identical names are baed on different type , of which only one can be a legitimate name , is called as homonym.**
- **Tautonym: A case in which name of genus and the name of the species is the same.**

Names of Taxa

Rank	*Ending*	*Example*
Kingdom	-bionta	Chlorobionta
Division	-phyta	Magnoliophyta
	-mycota (Fungi)	Eumycota
Subdivision	-phytina	Pterophytina
	-mycotina (Fungi)	Eumycotina
Class	-opsida	Magnoliopsida
	-phyceae (Algae)	Chlorophyceae
	-mycetes (Fungi)	Basidiomycetes
Subclass	-opsidae	Pteropsidae
	-idae (Seed plants)	Rosidae
	-physidae (Algae)	Cyanophysidae
	-mycetidae (Fungi)	Basidiomycetidae
Order	-ales	Rosales
Suborder	-ineae	Rosineae
Family	-aceae	Rosaceae
Subfamily	-oideae	Rosoideae
Tribe	-eae	Roseae
Subtribe	-inae	Rosinae
Genus	-us, -um, -is, -a, -on	*Pyrus, Allium, Arabis, Rosa, Polypogon*
Subgenus		*Cuscuta* subgenus *Eucuscuta*
Section		*Scrophularia* section *Anastomosanthes*
Subsection		*Scrophularia* subsection *Vernales*
Series		*Scrophularia* series *Laterifloraе*
Species		*Rosa canina*
Subspecies		*Crepis sancta* subsp. *bifida*
Varietas		*Lantana camara* var. *varia*
Forma		*Tectona grandis* f. *punctata*

Typification: Type Specimen is the one representative of the taxon.

- **Holotype: A specimen designated by the author in the original publication (nomenclatural type).**
- **Isotype: A duplicate specimen of the holotype collected at the same time and place (may be in other herbarium).**
- **Lectotype: A specimen chosen from the author's original material when no holotype has been designated.**
- **Neotype: A specimen selected when all original specimens have been destroyed**

Author Citation, Effective Publication and Principle of Priority

Author Citation

- **For a name to be complete, it should be accompanied by the name of the author or authors who first published the name validly. The names of the authors are commonly abbreviated, Example L. for Carolus Linnaeus**
- ***Aizoon canariense* L.**
- ***Tribulus macropterus* var. *arabicus* (Hosni) Al-Hemaid & J. Thomas**

Basic structure of a taxonomic Research papers / Recent publication of a new species in taxonomic journal

Ann. Bot. Fennici 53: 37–39 ISSN 0003-3847 (print) ISSN 1797-2442 (online)
Helsinki 4 January 2016 © Finnish Zoological and Botanical Publishing Board 2016

Silene langshanensis (Caryophyllaceae), a new species from Inner Mongolia, China

Li-Qing Zhao[1,*], Zhi-Ming Xin[2] & Yi-Zhi Zhao[1]

[1] College of Life Science, Inner Mongolia University, Hohhot 010021, China (*corresponding author's e-mail: zhaoliening@126.com)
[2] Experimental Center for Desert Forestry, Chinese Academy of Forestry, Dengkou, Inner Mongolia 015200, China

Received 22 Apr. 2015, final version received 9 Oct. 2015, accepted 9 Oct. 2015

Zhao L.Q., Xin Z.M. & Zhao Y.Z. 2016: *Silene langshanensis* (Caryophyllaceae), a new species from Inner Mongolia, China. — *Ann. Bot. Fennici* 53: 37–39.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin *sp. nova* (Caryophyllaceae), is described and illustrated from Inner Mongolia, China. It appears to be most closely related to *S. scabrifolia* of *Silene* sect. *Holopetalae*. *Silene langshanensis* can be distinguished by the basally pubescent carpophore, petals with obtuse auricles, stems and leaves with dense, short hairs, and by the glabrous calyx.

In total, there are about 600 species of *Silene s. lato* (Caryophyllaceae) (Zhou *et al.* 2001). They are distributed mainly in the northern temperate regions, but occur also in Africa and South America (Zhou *et al.* 2001). Among these species, 110 are known from China, of which 67 are endemic. Twenty of the endemics (nine species of *Silene s. stricto*, nine of *Melandrium*, one of *Cucubalus* and one of *Lychnis*) are found in Inner Mongolia.

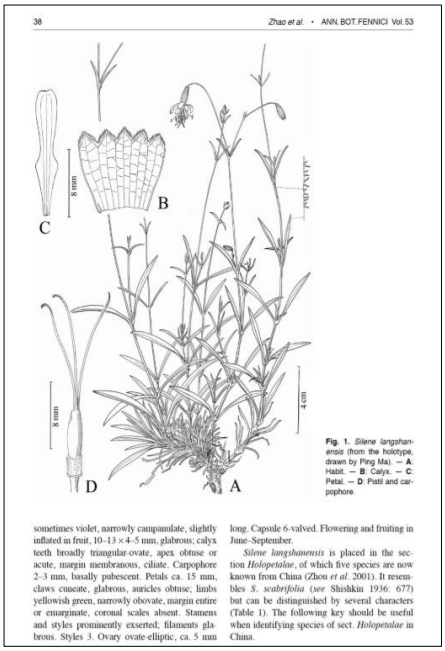
In September 2008 and later, in 2014, the authors Zhao and Xin collected specimens of *Silene* from Langshan in Bayannaoer (Inner Mongolia) from desert steppe communities on mountain slopes at 1150–1400 m a.s.l. After careful study, we concluded that the specimens represented an undescribed species of *Silene*.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin, *sp. nova* (Fig. 1)

HOLOTYPE: China. Inner Mongolia, Bayannaoer, Dengkou, Mt. Langshan, 40°43'58.4"N, 106°22'28.5"E, on stony mountain slopes, 1371 m a.s.l., 9 June 2014 *Li-Qing Zhao, Zhi-Ming Xin, Shuai Qin & Long Chen N14-001* (HIMC). — PARATYPES (all HIMC?): Same location as holotype, 9 June 2014 *Li Qing Zhao, Zhi-Ming Xin, Shuai Qin & Long Chen N14-002, N14-003, N14-004*; Mt. Langshan, 40°39'27.7"N, 106°25'14.1"E, on stony mountain slopes, alt. 1185 m a.s.l., 10 September 2008 *Li-Qing Zhao & Zhi-Ming Xin N08-001*.

Herbs perennial, 20–50 cm tall. Roots robust, lignified. Plants cespitose with erect stems, multi-branched or sparsely branched, pubescent, upper parts glabrescent. Basal leaves oblanceolate, 20–60 × 2–6 mm, both surfaces pubescent, margins ciliate, base attenuate into a long petiole, apex acute; cauline leaves 3–7 pairs, linear-oblanceolate or linear-lanceolate, with short, axillary sterile branches or sometimes elongated flowering axillary branches. Flowers in a racemiform-like thyrse; cymes alternating (resulting from suppression of opposite cyme at same node) or opposite, 1-flowered (rarely 2), peduncles nearly equal or shorter than pedicels. Pedicel 2–6 cm, glabrescent; bracts ovate-lanceolate, ciliate, base connate, apex acuminate. Calyx tubular, green, sometimes violet, narrowly campanulate, slightly inflated in fruit, 10–13 × 4–5 mm, glabrous; calyx teeth broadly triangular-ovate, apex obtuse or acute, margin membranous, ciliate. Carpophore 2–3 mm, basally pubescent. Petals ca. 15 mm, claws cuneate, glabrous, auricles obtuse; limbs yellowish green, narrowly obovate, margin entire or emarginate, coronal scales absent. Stamens and styles prominently exserted; filaments glabrous. Styles 3. Ovary ovate-elliptic, ca. 5 mm long. Capsule 6-valved. Flowering and fruiting in June–September.

Silene langshanensis is placed in the section *Holopetalae*, of which five species are now known from China (Zhou *et al.* 2001). It resembles *S. scabrifolia* (*see* Shishkin 1936: 677) but can be distinguished by several characters (Table 1). The following key should be useful when identifying species of sect. *Holopetalae* in China.



38 Zhao et al. • ANN. BOT. FENNICI Vol. 53

Fig. 1. *Silene langshanensis* (from the holotype, drawn by Ping Ma). — **A**: Habit. — **B**: Calyx. — **C**: Petal. — **D**: Pistil and carpophore.

ANN. BOT. FENNICI Vol. 53 • *Silene langshanensis, a new species from Inner Mongolia, China* 39

Table 1. Main morphological differences between *Silene langshanensis* and *S. scabrifolia*.

Character	*S. langshanensis*	*S. scabrifolia* (= *S. komarovii*)
Stem	densely pubescent, upper part glabrescent when flowering	pubescent in lower part, glabrous and viscid above
Basal leaves	oblanceolate, 20–60 × 2–6 mm	spatulate or lanceolate, 60–80 × 5–10 mm
Cyme	1-flowered (rarely 2)	multiflowered
Pedicel	20–60 mm long, glabrescent	5–10 mm long, sparsely pubescent
Calyx	narrowly campanulate, 10–13 × 4–5 mm, glabrous	tubular-clavate, 8–12 × 2–3 mm, glabrous or sparsely villous
Carpophore	shortly pubescent	glabrous
Petal	with obtuse auricles	without distinct auricles
Limbs	yellowish green	yellowish white

1. Leaves ovate-lanceolate, 15–30 mm wide ..... *S. langyeana*
1. Leaves lanceolate or linear, 1.5–10 mm wide 2
2. Leaves linear, 10–30 × 1.5–3 mm ..... *S. holopetala*
2. Leaves oblanceolate or lanceolate, 30–80 mm long, usually more than 4 mm wide 3
3. Stems usually not branched; calyx 6–9 mm; petals pinkish abaxially ..... *S. pseudotenuis*
3. Stems branched, calyx 8–13 mm; petals yellowish green or yellowish white 4
4. Stem pubescent in lower part, glabrous and viscid above; cymes multiflowered; petals yellowish white, without obvious auricles; carpophore glabrous ..... *S. scabrifolia*
4. Stem with dense short hairs, upper part glabrescent when flowering; cymes 1-flowered (rarely 2); petals yellowish green, with obtuse auricles; carpophore basally pubescent ..... *S. langshanensis*

Acknowledgements

We are grateful to Ping Ma for the drawing. This study was financially supported by Natural Science Foundation of Inner Mongolia Autonomous Region (2014ZD02), National Key Basic Research Program of China (2014CB138802) and the Central Public-interest Scientific Institution Basal Research Fund (CAFYBB2014MA016).

References

Shishkin B.K. [Шишкин Б.К.] (ed.) 1936: [*Flora USSR*], vol. 6. — Izdatel'stvo Akademii Nauk USSR, Leningrad. [In Russian]

Zhou L.H., Wu Z.Y., Lidén M. & Oxelman B. 2001: *Silene*. — In: Wu Z.Y. & Raven P.H. (eds.), *Flora of China*, vol. 6: 66–100. Science Press, Beijing & Missouri Botanical Garden Press, Saint Louis.

This article is also available at http://www.annbot.net and http://www.bioone.org/loi/anbt

Ann. Bot. Fennici 53: 37–39 ISSN 0003-3847 (print) ISSN 1797-2442 (online)
Helsinki 4 January 2016 © Finnish Zoological and Botanical Publishing Board 2016

Silene langshanensis (Caryophyllaceae), a new species from Inner Mongolia, China

Li-Qing Zhao[1,*], Zhi-Ming Xin[2] & Yi-Zhi Zhao[1]

[1)] College of Life Science, Inner Mongolia University, Hohhot 010021, China (*corresponding author's e-mail: zhaotieniu@126.com)
[2)] Experimental Center for Desert Forestry, Chinese Academy of Forestry, Dengkou, Inner Mongolia 015200, China

Received 22 Apr. 2015, final version received 9 Oct. 2015, accepted 9 Oct. 2015

Zhao L.Q., Xin Z.M. & Zhao Y.Z. 2016: *Silene langshanensis* (Caryophyllaceae), a new species from Inner Mongolia, China. — *Ann. Bot. Fennici* 53: 37–39.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin *sp. nova* (Caryophyllaceae) is described and illustrated from Inner Mongolia, China. It appears to be most closely related to *S. scabrifolia* of *Silene* sect. *Holopetalae*. *Silene langshanensis* can be distinguished by the basally pubescent carpophore, petals with obtuse auricles, stems and leaves with dense, short hairs, and by the glabrous calyx.

In total, there are about 600 species of *Silene s. lato* (Caryophyllaceae) (Zhou *et al.* 2001). They are distributed mainly in the northern temperate regions, but occur also in Africa and South America (Zhou *et al.* 2001). Among these species, 110 are known from China, of which 67 are endemic. Twenty of the endemics (nine species of *Silene s. stricto*, nine of *Melandrium*, one of *Cucubalus* and one of *Lychnis*) are found in Inner Mongolia.

In September 2008 and later, in 2014, the authors Zhao and Xin collected specimens of *Silene* from Langshan in Bayannaoer (Inner Mongolia) from desert steppe communities on mountain slopes at 1150–1400 m a.s.l. After careful study, we concluded that the specimens represented an undescribed species of *Silene*.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin, *sp. nova* (Fig. 1)

HOLOTYPE: China. Inner Mongolia, Bayannaoer, Dengkou, Mt. Langshan, 40°43′58.4″N, 106°22′28.5″E, on stony mountain slopes, 1371 m a.s.l., 9 June 2014 *Li-Qing Zhao, Zhi-Ming Xin, Shuai Qin & Long Chen N14-001* (HIMC). — PARATYPES (all HIMC!): Same location as holotype, 9 June 2014 *Li-Qing Zhao, Zhi-Ming Xin, Shuai Qin & Long Chen N14-002, N14-003, N14-004*; Mt. Langshan, 40°39′27.7″N, 106°23′14.1″E, on stony mountain slopes, alt. 1185 m a.s.l., 10 September 2008 *Li-Qing Zhao & Zhi-Ming Xin N08-001*.

Herbs perennial, 20–50 cm tall. Roots robust, lignified. Plants cespitose with erect stems, multi-branched or sparsely branched, pubescent, upper parts glabrescent. Basal leaves oblanceolate, 20–60 × 2–6 mm, both surfaces pubescent, margins ciliate, base attenuate into a long petiole, apex acute; cauline leaves 3–7 pairs, linear-oblanceolate or linear-lanceolate, with short, axillary sterile branches or sometimes elongated flowering axillary branches. Flowers in a racemiform-like thyrse; cymes alternating (resulting from suppression of opposite cyme at same node) or opposite, 1-flowered (rarely 2), peduncles nearly equal or shorter than pedicels. Pedicel 2–6 cm, glabrescent; bracts ovate-lanceolate, ciliate, base connate, apex acuminate. Calyx tubular, green,

Effective publication in the journal, available to Botanist

Date of valid publication (principles of priority): If the same species will be published by some one else after this date then the publication will be not valid. (/Principles of Priority).

Abstract / Summary / Synopsis.

Previously it was required to write in Latin.

Specimens examined

Botanical name in Latin

Rank indicated

Type Specimen indicated

Taxonomic Description

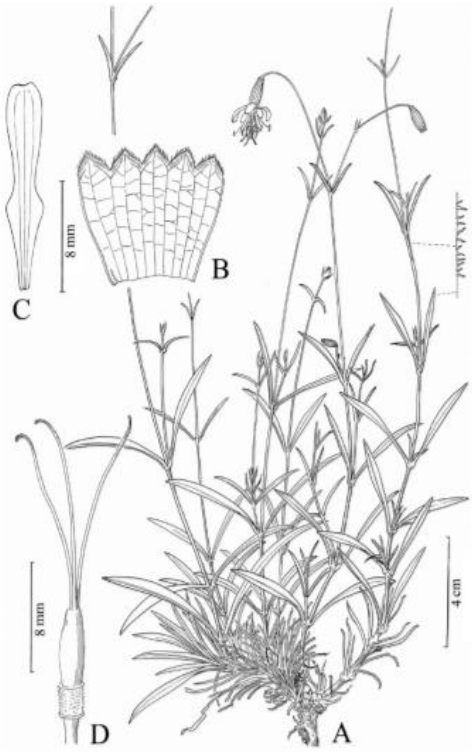


Fig. 1. *Silene langshanensis* (from the holotype, drawn by Ping Ma). — **A**: Habit. — **B**: Calyx. — **C**: Petal. — **D**: Pistil and carpophore.

sometimes violet, narrowly campanulate, slightly inflated in fruit, 10–13 × 4–5 mm, glabrous; calyx teeth broadly triangular-ovate, apex obtuse or acute, margin membranous, ciliate. Carpophore 2–3 mm, basally pubescent. Petals ca. 15 mm, claws cuneate, glabrous, auricles obtuse; limbs yellowish green, narrowly obovate, margin entire or emarginate, coronal scales absent. Stamens and styles prominently exserted; filaments glabrous. Styles 3. Ovary ovate-elliptic, ca. 5 mm long. Capsule 6-valved. Flowering and fruiting in June–September.

Silene langshanensis is placed in the section *Holopetalae*, of which five species are now known from China (Zhou *et al.* 2001). It resembles *S. scabrifolia* (*see* Shishkin 1936: 677) but can be distinguished by several characters (Table 1). The following key should be useful when identifying species of sect. *Holopetalae* in China.

Table 1. Main morphological differences between *Silene langshanensis* and *S. scabrifolia*.

Character	*S. langshanensis*	*S. scabrifolia* (= *S. komarovii*)
Stem	densely pubescent, upper part glabrescent when flowering	pubescent in lower part, glabrous and viscid above
Basal leaves	oblanceolate, 20–60 × 2–6 mm	spatulate or lanceolate, 60–80 × 5–10 mm
Cyme	1-flowered (rarely 2)	multiflowered
Pedicel	20–60 mm long, glabrescent	5–10 mm long, sparsely pubescent
Calyx	narrowly campanulate, 10–13 × 4–5 mm, glabrous	tubular-clavate, 8–12 × 2–3 mm, glabrous or sparsely villous
Carpophore	shortly pubescent	glabrous
Petal	with obtuse auricles	without distinct auricles
Limbs	yellowish green	yellowish white

1. Leaves ovate-lanceolate, 15–30 mm wide ..... *S. kungessana*
1. Leaves lanceolate or linear, 1.5–10 mm wide 2
2. Leaves linear, 10–30 × 1.5–3 mm ..... *S. holopetala*
2. Leaves oblanceolate or lanceolate, 30–80 mm long, usually more than 4 mm wide 3
3. Stems usually not branched; calyx 6–9 mm; petals pinkish abaxially .....
3. Stems branched; calyx ... or yellow
4. Stem ... cyme ... obvi...
4. Ste...

Acknowledgements

We are grateful to Ping Ma for the drawing. This study was financially supported by Natural Science Foundation of Inner Mongolia Autonomous Region (2014ZD02), National Key Basic Research Program of China (2014CB138802) and the Central Public-interest Scientific Institution Basal Research Fund (CAFYBB2014MA016).

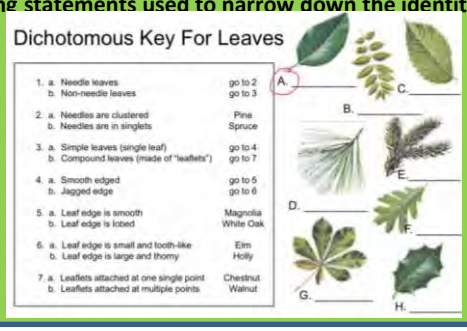
References

...akin B.K. [Шишкин Б.К.] (ed.) 1936: [*Flora USSR*], vol. ... — Izdatel'stvo Akademii Nauk USSR, Leningrad. [In ...ussian].
...L.H., Wu Z.Y., Lidén M. & Oxelman B. 2001: *Silene*. ... In: Wu Z.Y. & Raven P.H. (eds.), *Flora of China*, vol. ...: 66–100. Science Press, Beijing & Missouri Botanical Garden Press, Saint Louis.

Line drawing

Taxonomic Key for Identification

Taxonomic Key: An identification device, consisting of contrasting statements used to narrow down the identity of a taxon



Taxonomic (Order / Family / Genus / Species) Circumscription and Need of taxonomic evidences

Identifying Plant Families

CARYOPHYLLACEAE

- Herbs
- Leaves in opposite pairs, unlobed, untoothed
- Flowers usually have 5 petals
- Flowers usually have 5 sepals
- Flowers in cymes (group of flowers, terminal flower opens first)
- Single capsule fruit

APIACEAE

- Herbs
- Leaves usually alternate with sheathing, inflated leaf-stalk bases
- Flowers have 5 separate petals
- Flowers small
- Umbels type of inflorescence

ASTERACEAE / COMPOSITAE

- Largest family of flowering plants worldwide
- Herbs
- Leaves without stipules
- Flowers small in dense heads
- Petals always joined into a corolla-tube (petals fused together below forming a tube)

ASCLEPIADACEAE

- Perennial herbs, vines, and shrubs with milky sap, some cactus-like
- Leaves opposite or whorled, simple, entire
- Flowers bisexual, actinomorphic, with elaborate corona containing hoods and horns
- Highly specialized pollination mechanism
- Pollen contained in waxy pollinia connected in pairs to glands
- Stamens and carpels united into gynostegium
- Fruit a follicle, seeds with tuft of silky hairs

BRASSICACEAE

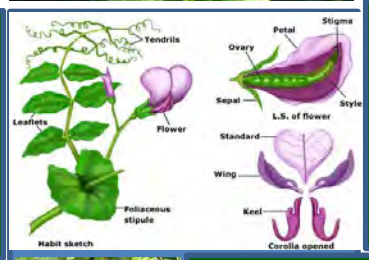
- Herbs
- Alternate leaves
- No stipules
- Flowers have 4 petals in a cross
- Flowers have 4 sepals
- Many cultivated vegetables

CUCURBITACEAE

- Herbaceous vines , Tendrils present , Plants usually monecious , Flowers 5-merous
- Ovary inferior, Fruit usually a pepo

LAMIACEAE / LABIATAE

Herbs
Square stems
Leaves opposite
Leaves often toothed
No stipules
Tubular flowers
Flowers usually have hood and prominent lower lip



FABACEAE / LEGUMINOSAE

- Five-petalled flowers
- Leaves usually trifoliate or pinnate
- Wide standard petal at top
- 5 sepals forming calyx-tube (lower parts of sepals fused)
- Fruit an elongated pod

EUPHORBIACEAE

- Habit: herbs, shrubs, stem succulents, trees; often with milky sap
- Leaves: alternate, opposite, whorled; simple (rarely palmately compound); stipulate
- Plants: monoecious or dioecious
- Inflorescence: cymose, racemes, cyathium
- Perianth: 0 (4-6); distinct or basally connate, free or adnate at base to stamens
- Stamens: 1-many, distinct or variously connate
- Ovary: 3 carpels; connate; superior; 3 (1-4) locules with 1 or 2 apical-axile ovules per locule; styles 3 (1-4), often forked
- Fruit: schizocarpic capsule (drupe, berry, pod, samara)

Taxonomic (Order / Family / Genus / Species) Circumscription and Need of taxonomic evidences

Circumscription is the definition of a taxon, that is, a group of organisms.

Haloxylon persicum
A small tree of deep sand, terminal shoot often pendulous, flowering branches thin, 1-1.5 mm across, stigma 5

Haloxylon salicornicum
Shrub up to about 1 m tall, branches as a rule not dropping, flowering branches thicker, 5-7 mm across, stigma 2

Fagonia indica
Leaves with single leaflet, branches terete, fruiting pedicle longer then fruit, sepal persistent in fruit

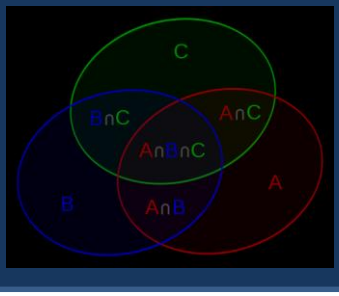
Fagonia ovalifolia
Branches terete, fruiting pedicles longer then fruit, Sepal deciduous, plant covered with sessile or stipulate glands, leaves narrowly oblanceolate or obovate, petiole 2-6 mm long

Species complex:

A species complex is a group of closely related species that are very similar in appearance to the point that the boundaries between them are often unclear.

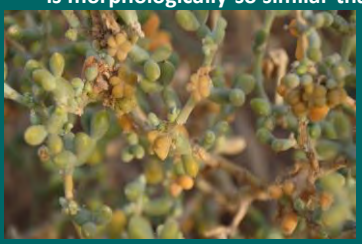
Tribulus macropterus* var. *arabicus
Flower 2-4 cm wide

Tribulus macropterus* var. *macropterus
Flower 0.8-1.5 cm wide



Tetraena (Family Zygophyllaceae)

- Currently, ten species of *Tetraena* are known from Saudi Arabia, Atl east four species are morphologically looking very similar.
- the genus *Zygophyllum is looking morphologically similar to Tetraena.*
- Based on the combined analyses of morphological and molecular data, Beier et al. (2003) transferred 35 species from Zygophyllum to *Tetraena* as new combinations.
- Therea re three spices of *Tetraena alba that is* *T. alba* var. alba., *T. alba* var. *arabica* and *T. alba* var. *amblyocarpa*. These all three verities is morphologically so similar that it like all similar, and very difficult to differentiate at the varietal level.



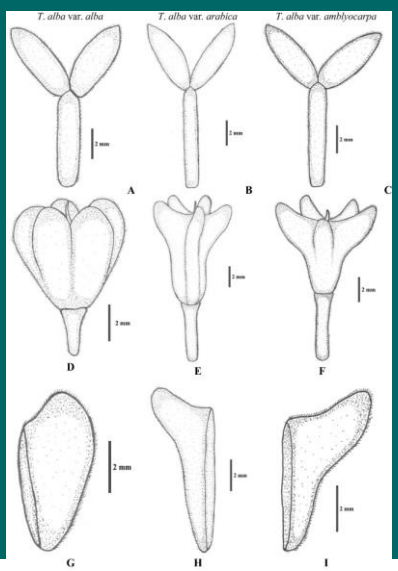
Tetraena alba

Tetraena coccinea

Tetraena qatarensis

Tetraena propinqua

Figure (A–D): (A) Leaves, flowers, and fruits of *T. alba* var. *alba*, (B) leaves and fruits of *T. alba* var. *alba*, (C) leaves, flowers, and fruits of *T. alba* var. *arabica*, (D) leaves, flowers, and fruits of *T. alba* var.



Key to the varieties of *Tetraena alba*

(1) Leaflets petiole up to 15 mm long; flowers 4–4.5 × 3–4.5 mm; capsules obconical star-shaped, with thick broad lobes 8–10 × 7–10 mm; pedicel up to 3 mm long..... var. *alba*

(2) Leaflets petiole up to 18 mm long; flowers 5.5 × 5 mm; capsules oblong obconical star-shaped, with slightly narrow lobes 11–13 × 8–10 mm; pedicel up to 6 mm long.....var. *arabica*

(3) Leaflets petiole up to 10 mm long; flowers 4 × 4 mm; capsules obconical-acute, with keeled lobes 9–13 × 8–12 mm; pedicel up to 6 mm long.....var. *amblyocarpa*

Figure (A–I): (A) Leaf of *T. alba* var. *alba*, (B) leaf of *T. alba* var. *arabica*, (C) leaf of *T. alba*var. *amblyocarpa*, (D) fruit of *T. alba* var. *alba*, (E) fruit of *T. alba* var. *arabica*, (F) fruit of *T. alba*var. *amblyocarpa*, (G) capsule lobe of *T. alba* var. *alba*, (H) capsule lobe of *T. alba* var. *arabica*, (I) capsule lobe

Taxonomic Evidences

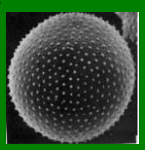
- **Specific diversity in millions, the issues of species complex and nomenclature brings many taxonomic question, and to solve it taxonomic evidences required which is could be morphology to molecules developed with the development of tools and technology of biological sciences**

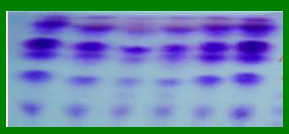
Taxonomic evidence for the establishment of classifications and phylogenies is gathered from a variety of sources

Morphology to Molecules

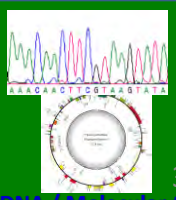

Morphology


Anatomy


Pollen

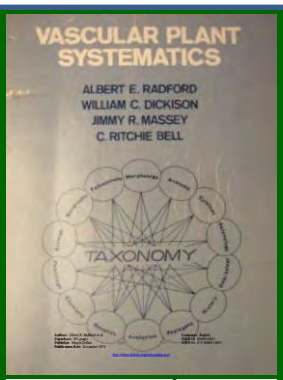

Chemistry

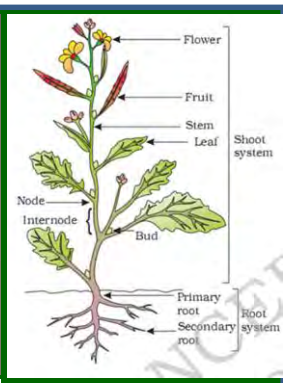

Chromosomes

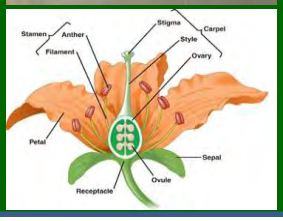

DNA / Molecular taxonomy

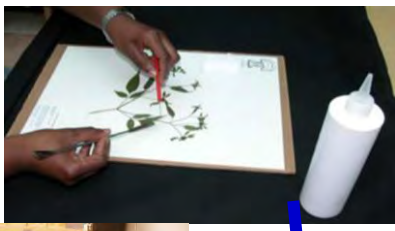
Source of Taxonomic Evidences: Vegetative and Floral Morphology

- Since there is huge diversity in the vegetative (external plant characteristics) and floral morphology among flowering plants, the vegetative and floral morphological characters is the first step in the plant identification and classification of angiospermic plants.
- Plant Morphology: Study of external structure of a plant
- Plant Anatomy: Study of Internal structure of a plant
- Flowering plants possess three kinds of vegetative (non-reproductive) organs: Roots, Stems, and Leaves
- The flower is the reproductive organ of the Angiosperms / Flowering plants.
- Alpha taxonomy
- Herbarium techniques

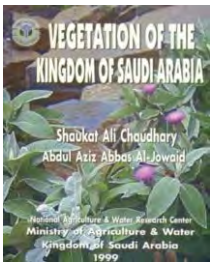


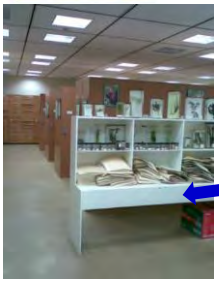






Phoenix dactylifera Linnaeus, Sp. Pl. 2: 1188. 1753.
Stems solitary or clustered and then with few shoots, to 30 m tall, to 50 cm in diam., rough with persistent, diamond-shaped leaf bases. Leaves 3-5 m; sheath and petiole to 1 m; rachis 1-2 m; acanthophylls many per side of rachis; pinnae to 200 per side of rachis, linear, irregularly arranged and spreading in different planes; middle pinnae to 40 × 2 cm. Male inflorescences erect, to 1 m, with many rachillae, these ca. 30 cm; female inflorescences erect, becoming pendulous, to 2 m, with to 150 rachillae, these to 40 cm. Fruits variable in shape, usually oblong, to 7 × 3 cm, brown or black; endosperm homogeneous.

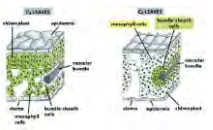




Source of Taxonomic Evidences: Plant Anatomy - (Internal Characteristics) and Physiology in Relation to Plant Taxonomy

- The Anatomical features is the most useful taxonomic characters in classification of the higher taxonomic categories .
- Anatomical features (plant cell & tissue types) (vs. morphological features) are somewhat more conservative characters that are not easily modified by growing conditions.
- Anatomical features of vegetative structures (roots, stems, leaves) are used to distinguish gymnosperms from angiosperms and monocots from dicots

- Physiological Evidence - C3 vs. C4 vs. CAM plants (in terms of their strategies for photosynthesizing.
- C4 photosynthesis occurs in about 10 unrelated families of monocots and dicots .... and is associated with plants that are adapted to arid environments.
- Cyperaceae
- Poaceae / Gramineae
- Aizoaceae
- Asteraceae
- Capparidaceae
- Euphorbiaceae
- Nyctaginaceae
- Portulacaceae
- Zygophyllaceae
- Hydrocharitaceae
- Acanthaceae
- Amaranthaceae
- Boraginaceae
- Caryophyllaceae
- Molluginaceae
- Polygonaceae
- Scrophulariaceae



Experimental Techniques:

- Cutting of thin slices / section (Transverse section or Longitudinal section) of plant organs
- Preparation of temporary slides or permanent slides
- Observation under light compound microscope using tissue stain like safranin, fast green

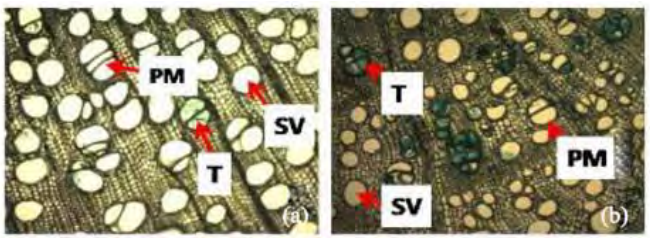
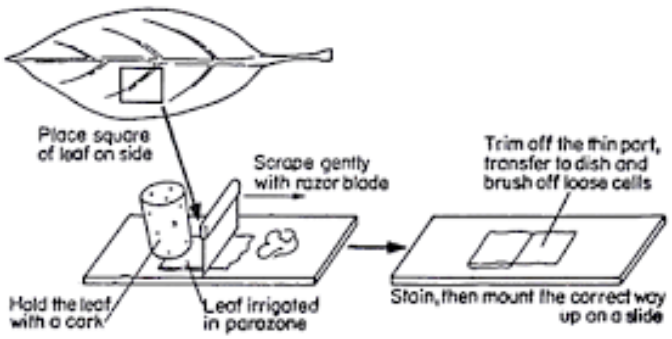
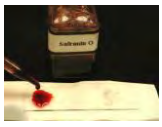


Figure: Transverse Sections of stem *Artocarpus atilis* (a) and *Artocarpus* communis (b).

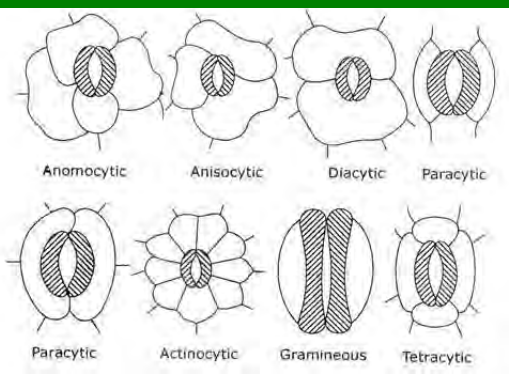
- PM: Pore multiple
- T=Tylose (Tyloses are outgrowths on parenchyma cells of xylem vessels of secondary heartwood
- SV: Solitary vessel





Source of Taxonomic Evidences: Systematic significance of Stomata

Stomata types produced by characteristic arrangements of guard cells and subsidiary cells can be of taxonomic use at the family or higher level.

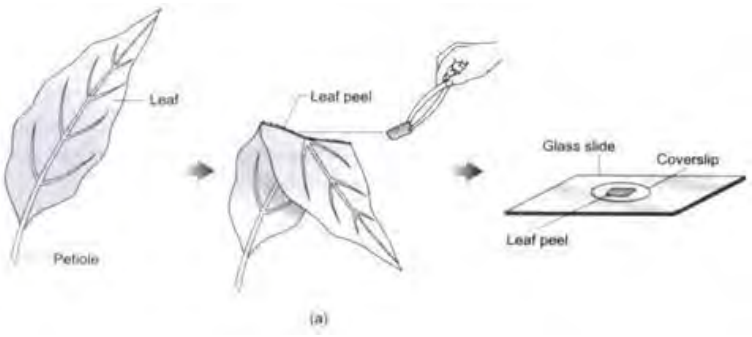


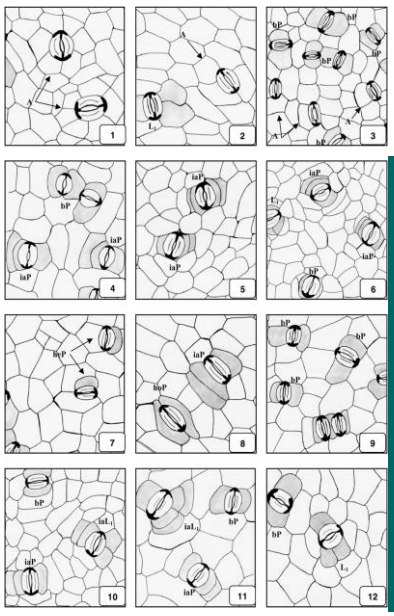
Different stomatal apparatus in Angiosperms

- **Anomocytic type: with epidermal cells around stomata not differentiated**
- **Paracytic type: with two or more cells parallel to the guard cells differentiated as subsidiary cells**
- **Diacytic type: with two subsidiary cells at right angles to the guards cells**
- **Anisocytic type: with three subsidiary cells of unequal size**
- **Actinocytic type: with stomata surrounded by a circle of radiating cells**
- **Tetracytic type: with four subsidiary cells**
- **Cyclocytic type: with concentric rings of subsidiary cells**
- **Graminaceous type: with dumb-bell shaped guard cells with two small subsidiary cells parallel to the guard cells.**

Experiment Techniques:

- **Usually peeling of leaves and observation under llight compound microscope (using tissue stain like safranin, fast green or without stain)**



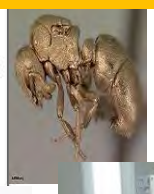


- **FARROKH et al., studies 32 *Salix species of Salix* Species (Salicaceae) in order to find the systematic significance of trichomes in Angiosperms**

Source of Taxonomic Evidences: Systematic Significance of Micromorphological Character of Leaf Surface / Trichomes / Electron Microscopy in Relation to Taxonomy

Experiment Techniques:

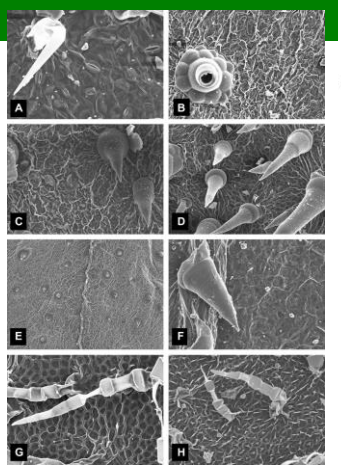
- Today's strongest compound microscopes have magnifying powers of 1,000 to 2,000X.
- SEM (Scanning electron Microscope) or TEM (transmission electron Microscope) is required to study ultra structure.
- SEM and TEM is costly microscope (price in Million or Million plus Riyal).
- Magnification about 500,000 times.
- Material to be studies kept on aluminum stub, and then placed under vacuum condition (gold coating machine) for gold coating.
- Gold coated biological sample placed in SEM chamber.
- Specimen passed thru electron beam
- Images can be only observed at computer monitor.

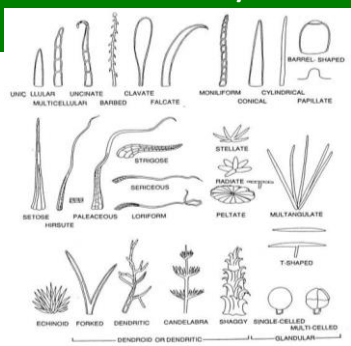


Palynology is the study of plant pollen and spores. There are two pollen types: monosulcate and tricolpate Monosulcate pollen are boat shaped with one long furrow and one germinal aperture (associated with primitive docots and the majority of monocots, the cycads and ferns). Triculpate pollen are found and typically have 3 apertures and is characteristic of the more advanced dicots.

Erdtman (1963) used the pollen charactersin solving the taxonomic problem of 105 family

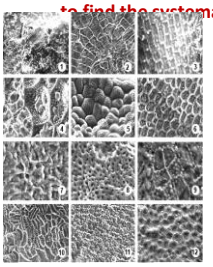
SEM of pollen grains. A: Nonaperturate pollen grain of *Persea americana; B: Monosulcate* pollen grain of *Magnolia grandiflora; C: Monoporate pollen grain of Siphonoglossa; D:* Tricolporate pollen grain of *Scaevola glabra; E: Polyporate spinose pollen grain of Ipomoaea wolcottiana; F: Tricolpate pollen grain of Disanthus cercidifolius.*





Trichomes morphology in Cucurbitaceae: (A) *Benincasa hispida* x300, (B) *Citrullus lanatus* x300, (C) *Cucumis melo* var. *agrestis* x300, (D) *C. sativus* x300, (E) *Diplocyclos palmatus* x50, (F) *Edgaria dargeelingensis* x300, (G) *Gynostemma burmanicum* x300, and (H) *G. pentaphyllum* x300.

- Trichomes meaning "hair", are fine outgrowths or appendages on plants.
- Ali and Al-Hemaid (2011) studies trichomes of 23 species of the member of the family Cucurbitaceae using Electron Microscope in order to find the systematic significance of micromorphological characters of

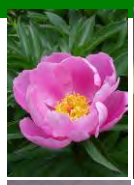


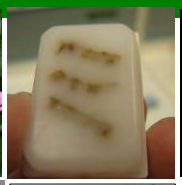
Scanning electron micrograph of the seed surface in Cucurbitaceae: 1. *Benincasa hispida ×400* (rugulate); 2. *Citrullus colocynthis ×400 (reticulate); 3. Cucumis melo var. agrestis ×400 (reticulate); 4. Diplocyclos palmatus ×1000 (reticulate); 5. Gynostemma laxiflorum ×600 (colliculate); 6. Hemsleya longivillosa ×400 (reticulate); 7. Luffa echinata ×1000 (reticulate); 8. Momordica charantia ×700* (reticulate); 9. *Momordica cymbalaria ×1000 (reticulate); 10. Schizopepon bryoniifolius ×400* (reticulate); 11. *Sicyos angulatus ×300 (rugulate); 12. Trichosanthes cucumerina ×320 (reticulate).*

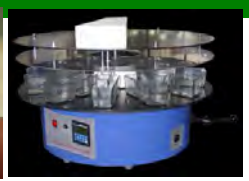
- Spermoderm refers to the pattern present on the seed coat of mature seeds.
- Seed characteristic, particularly exomorphic features as revealed by scanning electron microscopy, have been used by many workers in resolving taxonomic problems (Koul et al., 2000; Pandey and Ali, 2006) and evolutionary relationships (Kumar et al., 1999; Segarra and Mateu, 2001).
- Ali et al. (2003) studied the sppermoderm pattern of the members of the family cucurbitaceae using Electron Microscope in order to find the systematic significance of micromorphological characters seed surface

Source of Taxonomic Evidences: Systematic Significance of Embryology / Embryology in Relation to Taxonomy

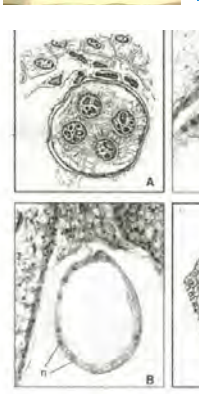
- Embryology is the branch of biology that studies the prenatal development of gametes (sex cells), fertilization, and development of embryos and seed coats.
- The major embryological character that separates the monocots from the dicots is the number of embryonic cotyledon leaves.
- Embryological features are normally constant at the family level and below.
- The genus *Paeonia* was earlier included under the family Ranunculaceae. But *Paeonia* differs from Ranunculaceae in chromosome number, vascular anatomy, floral anatomy.
- Worsdell (1908) suggested its removal to a distinct family, Paeoniaceae.
- The separation is supported by the embryological features: (i) centrifugal stamens (not centripetal); (ii) pollen with reticulately-pitted exine with a large generative cell (not granular, papillate and smooth, small generative cell); (iii) unique embryogeny in which early divisions are free nuclear forming a coenocytic stage, later only the peripheral part becomes cellular (not onagrad or solanad type); and (iv) seed arillate.







Station	Time (min)	Solution
1		Water
2	45	70% alcohol
3	45	80% alcohol
4	45	90% alcohol
5	45	100% alcohol
6	60	100% alcohol
7	60	100% alcohol
8	60	Clearing reagent (xylene or substitute)
9	60	Clearing reagent (xylene or substitute)
10	60	Paraffin 1
11	60	Paraffin 2
12	60	Paraffin 3



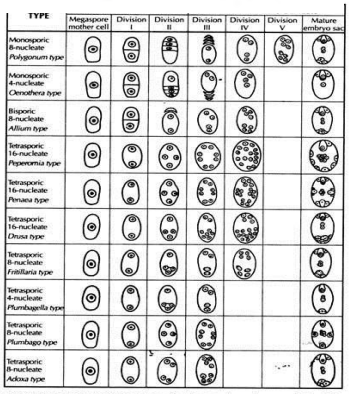


Fig. 3.8 : Development of different types of embryo sac in angiosperms (*after* Maheshwa [Micropyle above in all illustrations]

Fig. 12.14 Early stages of embryog onia sp. **A,B.** Coenocytic embryo. **C.** Cellularization. **D.** Formation of embryos in the coenocytic-cellular stage. *n*, nuclei; *c*, cells (*from* Czapik and Izmailow, 2001)

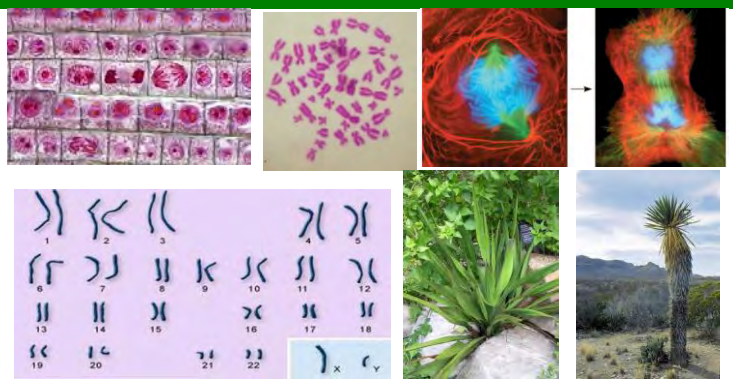
Source of Taxonomic Evidences: / Cytology in Relation to Taxonomy

- Cytology is the study of the cell.
- Chromosome is a thread-like structure of nucleic acids and protein found in the nucleus of the living cells, carrying genetic information in the form of gene.
- Number of chromosome are fixed for a species.

Chromosome Set:

- Number of chromosome can be counted in the metaphase stage of cell division.
- One copy of each of the different chromosomes in the nucleus containing one copy of each different gene.
- Haploid Number (n): The number of chromosomes comprising one set.
- Diploid Number (2n): The number of chromosomes in a cell containing two sets.
- Human Haploid (n)= 23, Diploid (2n)=46
- Dates Haploid (n)= 14, Diploid (2n)=28

- In plants, only information about chromosome number, shape or pairing at meiosis is used for classification purposes.
- The term karyotype is used for the phenotypic appearance for the somatic chromosomes.
- The diagrammatic representation of the karyotype is termed as idiogram.
- The characteristic of chromosome having taxonomic values are: chromosome number, chromosome size, chromosome morphology, and chromosome behavior during meiosis.
- The genus *Yucca* had long been treated as a member if Liliaceae because of the superior ovary. Hutchinson shifted *Yucca* to the family Agavaceae because the genus *Yucca* possess 25 small and 5 large chromosome which is similar to the member of family Agavaceae



Yucca carnerosana

Experiment Techniques:

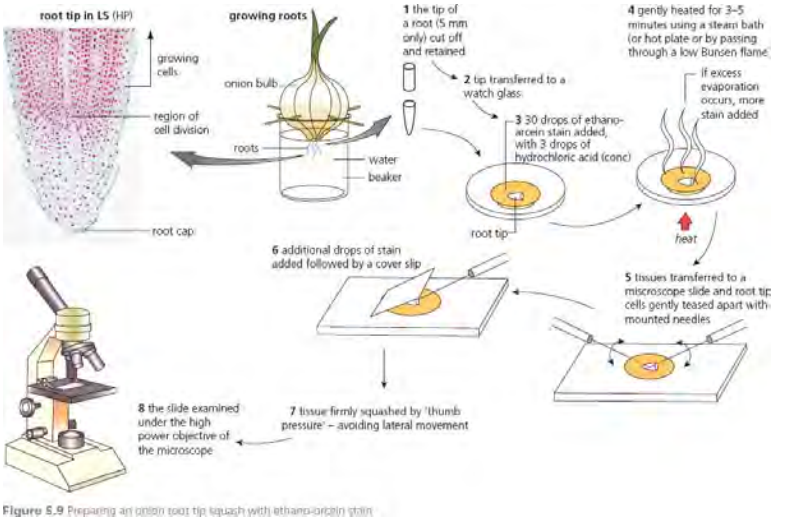
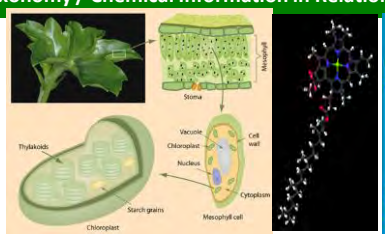
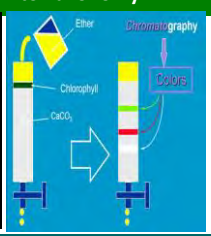


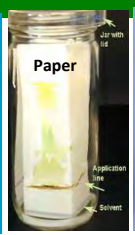
Figure 5.9 Preparing an onion root tip squash with ethano-arcein stain

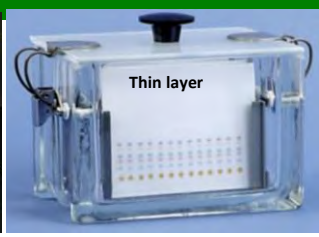
Source of Taxonomic Evidences: / Chemotaxonomy / Chemical Information in Relation to Taxonomy

- Application of chemistry to taxonomy is called chemical taxonomy / chemotaxonomy.
- Some of the major classes of the chemical evidence include Anthocyanin, Flavonoids, Alkaloids, Glycosides, Terpenes, Amino acid, Fatty acids, Aromatic compounds, Polysaccharides, Carotenoids
- Caryophyllales produces Betalin and not anthocyanin
- Polygonales produce anthocyanin and not Betalin
- Highly aromatic compound are found in Lamiaceae









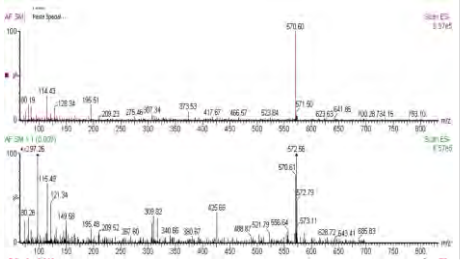
Chromatography is used to separate mixtures of substances into their components. All forms of chromatography work on the same principle. They all have a stationary phase (a solid, or a liquid supported on a solid) and a mobile phase (a liquid or a gas).

Mass spectrometry

Mass spectrometry (MS) is an analytical technique that ionizes chemicals and measures the masses within a sample.

Preparative HPLC

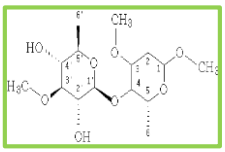
LCMS

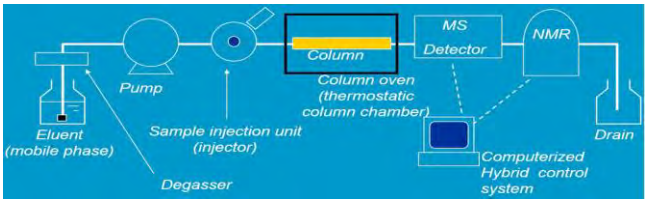


GCMS

LCMS NMR

Al-Allah et al., 2018 identified as methyl β-lilacinobioside islated from *Caralluma retrospiciens*





Source of Taxonomic Evidences: / Ecology in Relation to Taxonomy

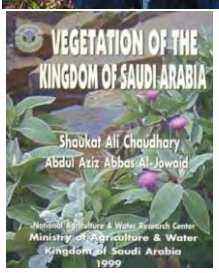
- The ecological criteria are of comparatively little direct importance in taxonomy.
- Ecological Evidence provides information about variation within plant taxa associated with plant adaptations and the distribution of plants.
- Plant ecologists frequently examine edaphic (soil) specializations, pollinating mechanisms (co-evolution), effect of habitat on hybridization, plant-herbivore interactions (co-evolution), seed-dispersal mechanisms, reproductive isolating mechanisms.
- Information from plant ecology has implications for classification below the level of genus.
- **Ecotypes:**
- Ecotypes is a distinct form or race of a plant species occupying a particular habitat.

Erect form of *Euphorbia hira*

Prostrate form of *Euphorbia hira*

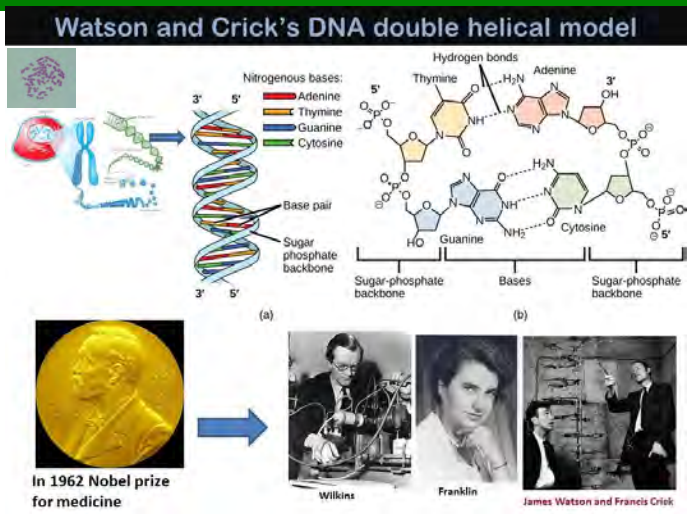
Methods

Filed studies to observe diversity and spices richness, Physiochemical properties, morphological variation, Habitat etc.



Source of Taxonomic Evidences: Molecular Data / DNA / Molecular Taxonomy

DNA (Deoxyribo Nucleic Acid)



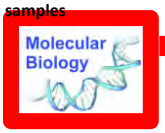
- The Cell is the basic structural, functional and biological unit of all known living organisms. The Nucleus is enclosed in an envelope which is a double membrane structure. The nucleus of each eukaryotic cell contains Deoxyribonucleic Acid (DNA). The Nucleus contains DNA in the form of loose threads called chromatin / Chromosomes.
- The chromosomes are the thread-like structure of nucleic acids and protein found in the nucleus of the living cells, carrying genetic information in the form of gene.
- The DNA is tightly packed into structures called chromosomes, which consist of long chains of DNA and associated proteins. Chromosome is the physical basis of heredity while DNA is the chemical basis of hereditary material.
- The DNA located in the nucleus of the cell called as nuclear DNA, but a small amount of DNA can also be found in the mitochondria (mitochondrial DNA or mtDNA) or chloroplast (chloroplast DNA or plastid DNA e.g. cpDNA).
- Genes carries genetic information from one generation to another generation. Genes lies on Chromosomes. Genes are made up of DNA. There are large number of genes occurs in each cell on each chromosomes.
- The model of DNA was given by James Watson and Francis Crick in 1962.
- An important property of DNA is that it can replicate, or make copies of itself. Each strand of DNA in the double helix can serve as a pattern for duplicating the sequence of bases.
- DNA is the molecule that carries the genetic information in all cellular forms of life. It belongs to a class of molecules called the nucleic acids, which are polynucleotides - that is, long chains of nucleotides. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). The order, or sequence, of these bases determines the information available for building and maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences. DNA bases pair up with each other, A with T and C with G, to form units called base pairs. Each base is also attached to a sugar molecule and a phosphate molecule. Together, a base, sugar, and phosphate are called a nucleotide. Nucleotides are arranged in two long strands that form a spiral called a double helix. The structure of the double helix is somewhat like a ladder, with the base pairs forming the ladder's and the sugar and phosphate molecules forming the vertical sidepieces of the ladder.
- Gene expression is the process of converting information from gene to cellular product. Protein synthesis is the main function of the gene. DNA transcribed in to RNA (called as Transcription), and then RNA translated into Amino Acids (called as Translation). There are 20 different types of amino acids. Several amino acids in a fixed sequenced forms protein. Several protein in a fixed sequenced forms Enzymes. The enzymes participate in the biochemical reaction of the cell. There are many biochemical simultaneously occurring in a cell. Proper biochemical reactions of cell ensure the life.

Molecular systematics

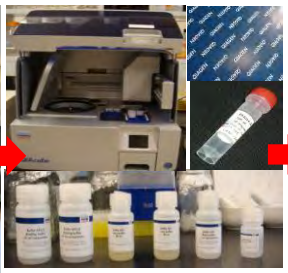
the utilization of nucleic acid data. As DNA sequence of a gene is constant in a species, hence advantage over morphological data for taxonomic studies.

- Taxonomist use molecular data from three different locations within a plant cell: chloroplast, mitochondrion and the nucleus.
- Molecular systematics involves following steps: (1) Sample collection, (2) DNA extraction, (3) Amplification using PCR –Polymerase chain Reaction, (4) DNA / Gene Sequencing, (5) Analysis of Sequence data.
- DNA barcoding can speed up identification of species. DNA barcoding helps in Wild plant identification / Medicinal plant authentication
- A DNA barcode is a short gene sequence taken from standardized portions of the genome, used to identify species

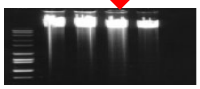
samples



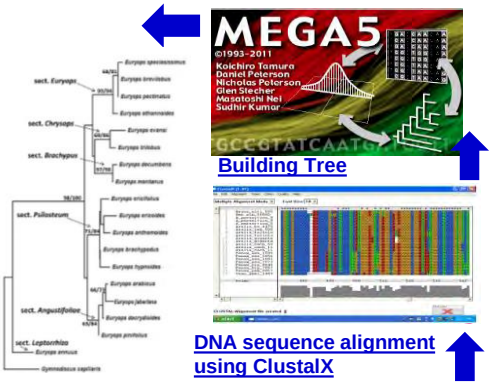
A view of molecular biology laboratory



Gel electrophoresis

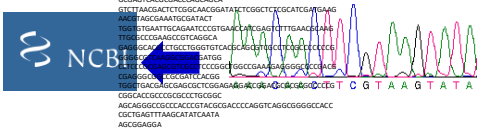


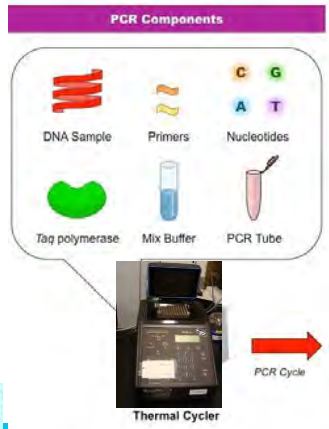
Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

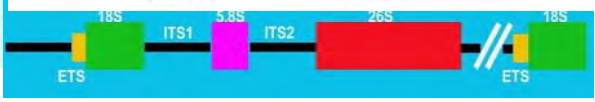


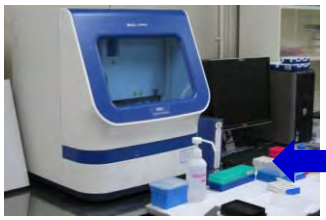
Building Tree

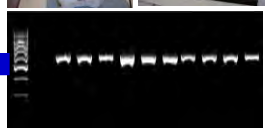
DNA sequence alignment using ClustalX



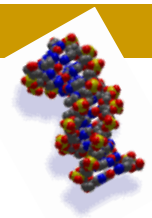






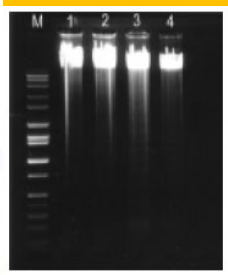


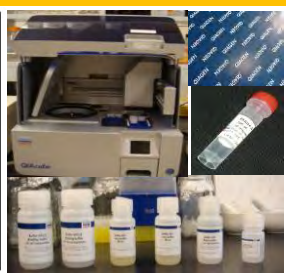
Sampling of leaf material for the molecular taxonomic study and DNA extraction

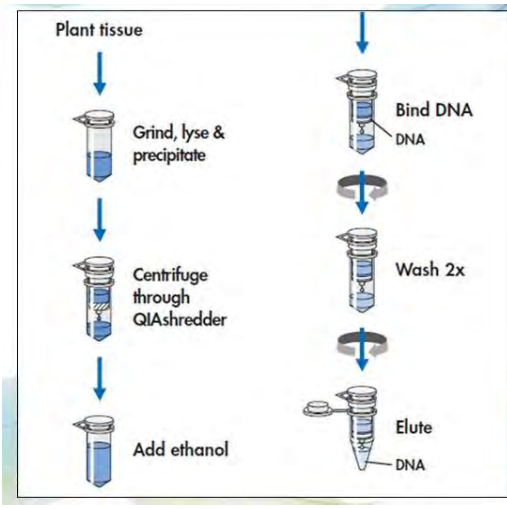


- **Doyle and Doyle (1990) is widely used protocol for DNA Extraction from plant tissue. But it involves preparation of several buffer manually. It takes long times. This method atleast take more than one day preparation and about whole day in DNA extraction. It also involves several times centrifugation. This method requires large amount of fresh leaves (10 gram or even more).**
- **Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15**
- **In contrast to manual method, there are several DNA extraction kit and automated DNA extraction machine is available like Qiagen automated DNA extraction machine, and Qiagen DNA extraction Kit.**
- **In Qiagen DNA extraction all the buffer are provided and ready to use. DNA can be extracted from small amount of 20 mg l dried leaf tissue or from very small piece of leaf collected from even old herbarium specimens. By using Qiagen DNA can be extracted in 3 hours. It do not required centrifugation manually.**

QIAGEN automated DNA extraction method







Choosing molecular marker, and application of PCR in plant molecular taxonomy / DNA taxonomy

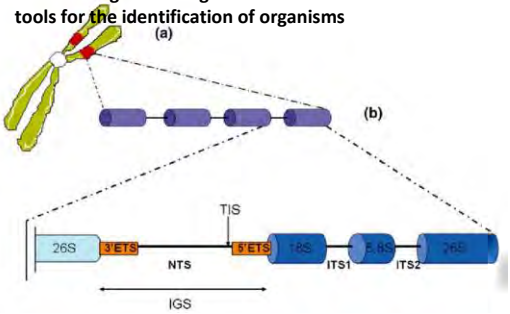
- In DNA sequencing method based practice of plant molecular taxonomy required DNA sequences.
- To obtain DNA sequence of a taxon required extraction of whole genomic DNA first. And then amplification of gene of interest. The amplification using gene interest is achieved by the polymerase chain reaction (PCR). The PCR results into billions of copies of gene of interest which can be observed in a gel under UV light. The amplified DNA later used for the purpose of DNA sequencing. So, for the cloning of the gene of interest using PCR requires primer. The primers are also called as molecular markers. To begin plant molecular taxonomy, selection of molecular marker is very critical and important.
- The most commonly used molecular marker in molecular taxonomy are ITS, rbcL, matK, psb, ndhF, trn gene.
- The molecular marker gene could be coding gene or non coding gene.

- **Properties of ideal marker genes**
  - A single-copy gene may be more useful than multiple-copy gene
  - The substitution rate (without altering protein) should be optimum so as to provide enough informative sites and alignment should be easy.
  - Primers should be available to selectively amplify the marker gene

- The nuclear ribosomal locus coding for the large subunit is represented in tandem arrays in the plant genome.
- ITS is located between the 18 and 26S rRNA genes.
- The 5.8S region on the other hand is only about 160 bp long and highly conserved within major organism groups.
- The ITS region consists of three parts: the ITS1 and ITS2 and the highly conserved 5.8S rDNA exon located in between. The total length of this region varies between 500 and 750 bp in angiosperms while in other seed plants it can be much longer, up to 1,500–3,500 bp.
- Spacer DNA is a region of non-coding DNA between genes.
- In contrast to the coding regions, spacers evolve more quickly, like the internal transcribed spacer (ITS) region, which is extensively used as a marker for phylogenetic reconstruction at different levels.
- The ITS is present in virtually all organisms. The advantages of this region are: (1) easy PCR amplification, with several universal primers available for a various kind of organisms; (2) multicopy structure; (3) moderate size allowing easy sequencing; and (4) it has a high degree of variation even between closely related species.,
- variability is due to frequently occurring nucleotide polymorphisms or to common insertions/deletions in the sequence.
- As DNA of ITS regions is removed and it is not part of the mature RNA molecule, they are considered noncoding regions of the genome

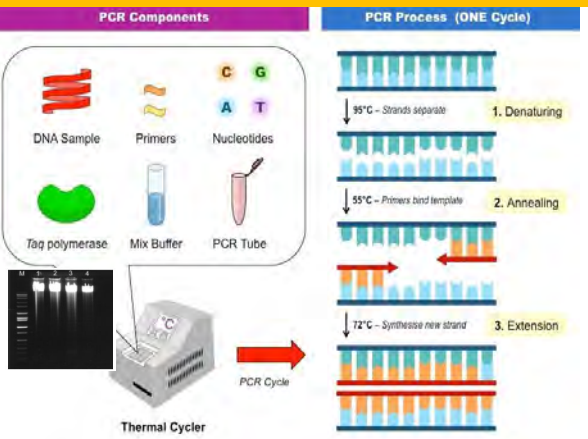
A fascinating feature of biological life is the common use of the DNA genetic code and its subsequent processing into functional units of protein through the intermediate RNA molecule.
The transcription of DNA into RNA and translation of RNA into protein are both highly
regulated and compartmentalized in all living organisms.

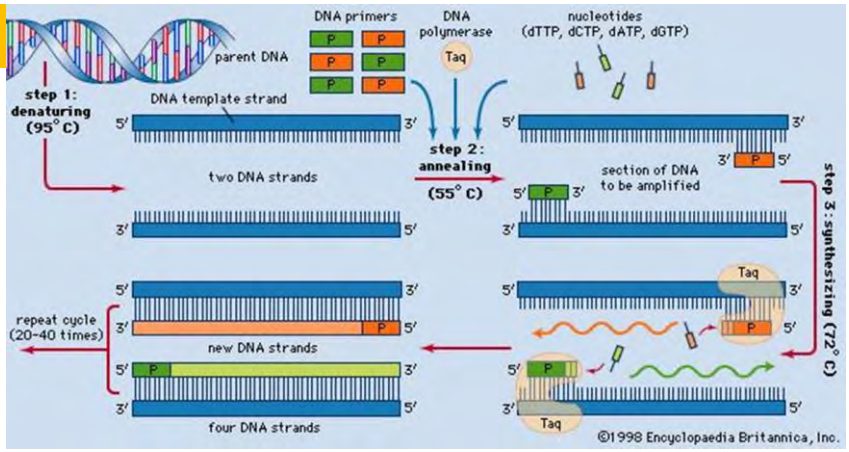
The cellular factory responsible for the production of protein is the ribosome. As the essential functions of ribosomes are critical for survival, their physical parameters have been conserved in all forms of life. Some components within the ribosomal factories have, however, changed sometimes. These similarities, as well as the changes within genetic material can be used as tools for the identification of organisms



MARKER	SEQUENCE	REFERENCE
ITS1 F	TCCGTAGGTGAACCTGCGG	White et al. (1990)
ITS4 R	TCCTCCGCTTATTGATATGC	White et al. (1990)
rbcLa F	ATGTCACCACAAACAGAGACTAAAGC	Levin (2003)
rbcLa R	GTAAAATCAAGTCCACCRCG	Kress and Erickson (2007)
MatK 390 F	CGATCTATTCATTCAATATTTC	Cuenoud et al. (2002)
MatK 1326 R	TCTAGCACACGAAAGTCGAAGT	Cuenoud et al. (2002)
psbA-trnH F	GTTATGCATGAACGTAATGCTC	Sang et al. (1997)
psbA-trnH R	CGCGCATGGTGGATTCACAATCC	Tate and Simpson (2003)
trn L–F R	GGTTCAAGTCCCTCTATCCC	Taberlet et al. (1991)
trn L–F F	ATTTGAACTGGTGACACGAG	Taberlet et al. (1991)

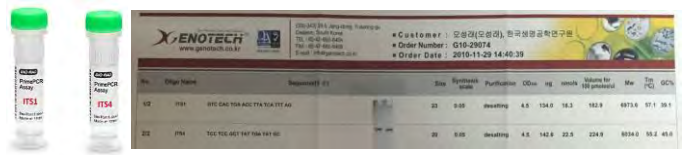
PCR (Polymerase Chain Reaction)





Contents of HF PCR premix Reaction size (20 µl reaction): 1. DNA polymerase 1µl, 2. Each dNTP (dATP, dCTP, dGTP, dTTP) 250 µM, 3. 10X reaction buffer Stabilizer and tracking dye2µl

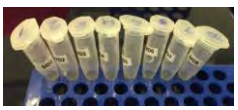
Template DNA (1µl ~ 100 ng), Primer (1µl each of F and R, 5 ~20 pmole)

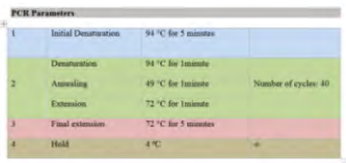


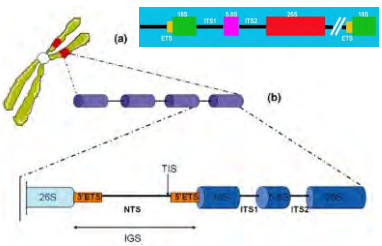
MARKER	SEQUENCE	REFERENCE
ITS1 F	TCCGTAGGTGAACCTGCGG	White et al. (1990)
ITS4 R	TCCTCCGCTTATTGATATGC	White et al. (1990)
rbcLa F	ATGTCACCACAAACAGAGACTAAAGC	Levin (2003)
rbcLa R	GTAAAATCAAGTCCACCRCG	Kress and Erickson (2007)
MatK 390 F	CGATCTATTCATTCAATATTTC	Cuenoud et al. (2002)
MatK 1326 R	TCTAGCACACGAAAGTCGAAGT	Cuenoud et al. (2002)
psbA-trnH F	GTTATGCATGAACGTAATGCTC	Sang et al. (1997)
psbA-trnH R	CGCGCATGGTGGATTCACAATCC	Tate and Simpson (2003)
trn L-F R	GGTTCAAGTCCCTCTATCCC	Taberlet et al. (1991)
trn L-F F	ATTTGAACTGGTGACACGAG	Taberlet et al. (1991)

1/10th genomic DNA dilution: Add 10 µl total genomic DNA in 90 µl molecular grade distilled water.

Dilution of primer for stock solution (100 pmoles/ µl): nmols X 10 Distilled water (ddH_2O) = 100 pmoles/ µl (Stock)



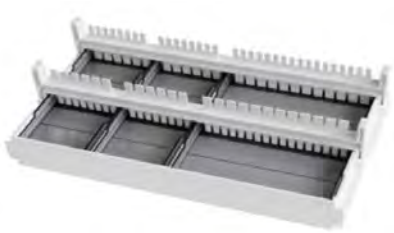


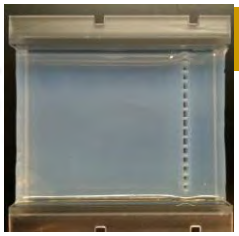


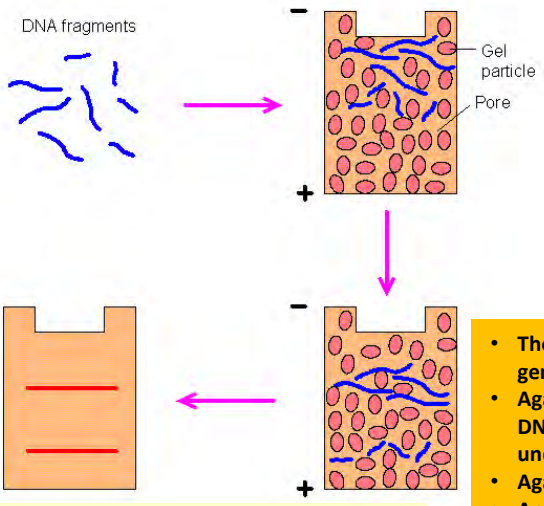
Agarose Gel Electrophoresis

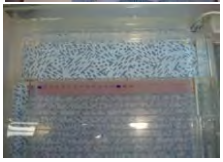

Agarose

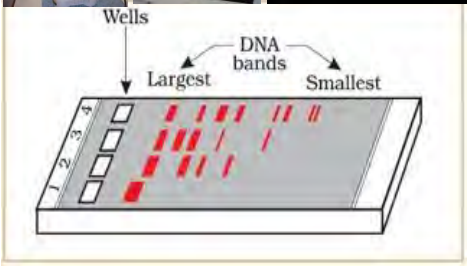

Buffer Solution

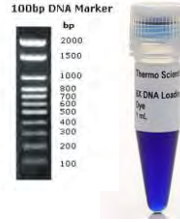


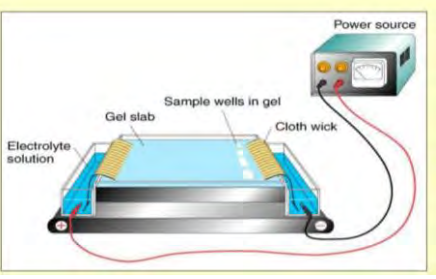






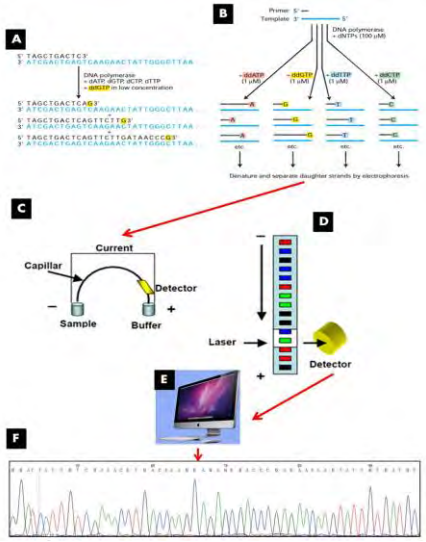






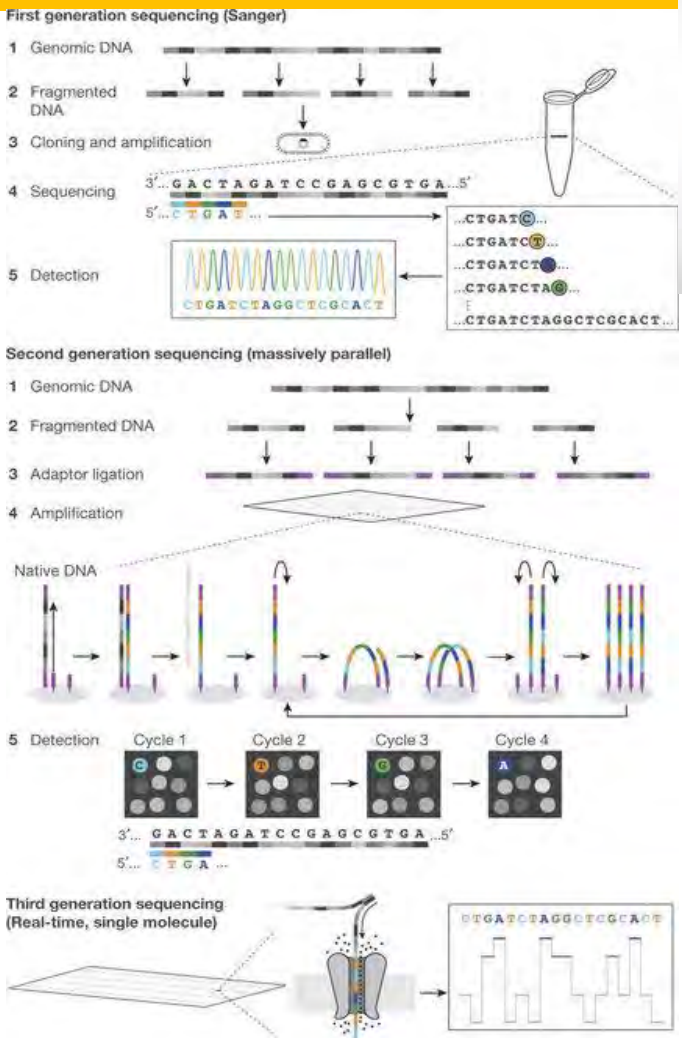
- **The main purpose of agarose gel electrophoresis is to determine the presence or absence of genomic DNA or PCR products and quantify the size (length of the DNA molecule).**
- **Agarose gel electrophoresis is a widely used technique for the preparation and analysis of DNA. Electrophoresis is a method of separating DNA based on the rate of movement while under the influence of an electric field.**
- **Agarose is a polysaccharide purified from seaweed.**
- **An agarose gel is created by suspending dry agarose in a buffer solution, boiling until the solution becomes clear, and then pouring it into a casting tray and allowing it to cool. During electrophoresis, the gel is submersed in a chamber containing a buffer solution and a positive and negative electrode.**
- **The DNA to be analyzed is forced through the pores of the gel by the electrical current.**
- **Under an electrical field, DNA moves to the positive electrode (red) and away from the negative electrode (black).**
- **DNA itself is not visible within an agarose gel.**
- **The DNA visualized by the use of dye that binds to DNA.**

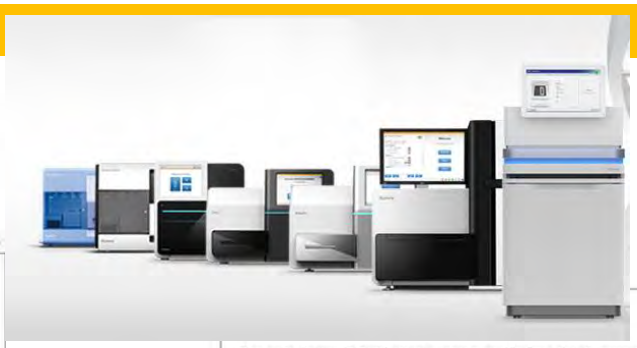
DNA sequencing



- DNA sequencing is the process of determining the sequence of nucleotides (A, T, C, and G) in a piece of DNA.
- In Sanger sequencing, the target DNA is copied many times, making fragments of different lengths. Fluorescent "chain terminator" nucleotides mark the ends of the fragments and allow the sequence to be determined.
- Next-generation sequencing techniques are new, large-scale approaches that increase the speed and reduce the cost of DNA sequencing.
- Sanger sequencing: The chain termination method
- Regions of DNA up to about 900 base pairs in length are routinely sequenced using a method called Sanger sequencing or the chain termination method.
- Ingredients for Sanger sequencing
- Sanger sequencing involves making many copies of a target DNA region. Its ingredients are similar to those needed for DNA replication in an organism, or for polymerase chain reaction (PCR), which copies DNA *in vitro*. They include:
- A DNA polymerase enzyme
- A primer, which is a short piece of single-stranded DNA that binds to the template DNA and acts as a "starter" for the polymerase
- The four DNA nucleotides (dATP, dTTP, dCTP, dGTP)
- The template DNA to be sequenced
- However, a Sanger sequencing reaction also contains a unique ingredient:
- Dideoxy, or chain-terminating, versions of all four nucleotides (ddATP, ddTTP, ddCTP, ddGTP), each labeled with a different color of dye
- Dideoxy nucleotides are similar to regular, or deoxy, nucleotides, but with one key difference: they lack a hydroxyl group on the 3' carbon of the sugar ring. In a regular nucleotide, the 3' hydroxyl group acts as a "hook," allowing a new nucleotide to be added to an existing chain.
- Once a dideoxy nucleotide has been added to the chain, there is no hydroxyl available and no further nucleotides can be added. The chain ends with the dideoxy nucleotide, which is marked with a particular color of dye depending on the base (A, T, C or G) that it carries.
- The DNA sample to be sequenced is combined in a tube with primer, DNA polymerase, and DNA nucleotides (dATP, dTTP, dGTP, and dCTP). The four dye-labeled, chain-terminating dideoxy nucleotides are added as well, but in much smaller amounts than the ordinary nucleotides.
- The mixture is first heated to denature the template DNA (separate the strands), then cooled so that the primer can bind to the single-stranded template. Once the primer has bound, the temperature is raised again, allowing DNA polymerase to synthesize new DNA starting from the primer. DNA polymerase will continue adding nucleotides to the chain until it happens to add a dideoxy nucleotide instead of a normal one. At that point, no further nucleotides can be added, so the strand will end with the dideoxy nucleotide.
- This process is repeated in a number of cycles. By the time the cycling is complete, it's virtually guaranteed that a dideoxy nucleotide will have been incorporated at every single position of the target DNA in at least one reaction. That is, the tube will contain fragments of different lengths, ending at each of the nucleotide positions in the original DNA (see figure below). The ends of the fragments will be labeled with dyes that indicate their final nucleotide.
- After the reaction is done, the fragments are run through a long, thin tube containing a gel matrix in a process called capillary gel electrophoresis. Short fragments move quickly through the pores of the gel, while long fragments move more slowly. As each fragment crosses the "finish line" at the end of the tube, it's illuminated by a laser, allowing the attached dye to be detected.
- The smallest fragment (ending just one nucleotide after the primer) crosses the finish line first, followed by the next-smallest fragment (ending two nucleotides after the primer), and so forth. Thus, from the colors of dyes registered one after another on the detector, the sequence of the original piece of DNA can be built up one nucleotide at a time. The data recorded by the detector consist of a series of peaks in fluorescence intensity, as shown in the chromatogramabove. The DNA sequence is read from the peaks in the chromatogram.

Advances in next generation DNA sequencing





The milestones listed below correspond to key developments in the evolution of sequencing technologies. This is a large topic, and we apologize for any omissions.

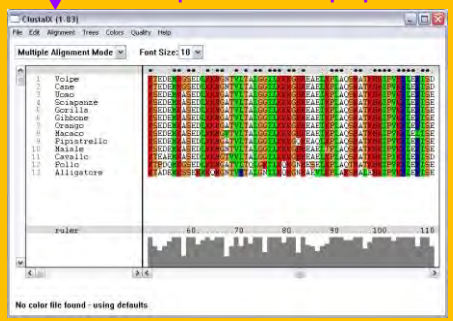
Technical milestones

1953: Sequencing of insulin protein[2]
1965: Sequencing of alanine tRNA[4]
1968: Sequencing of cohesive ends of phage lambda DNA[6]
1977: Maxam–Gilbert sequencing[9]
1977: Sanger sequencing[8]
1981: Messing's M13 phage vector[12]
1986–1987: Fluorescent detection in electrophoretic sequencing[14,15,17]
1987: Sequenase[18]
1988: Early example of sequencing by stepwise dNTP incorporation[139]
1990: Paired-end sequencing[22]
1992: Bodipy dyes[140]
1993: *In vitro* RNA colonies[37]
1996: Pyrosequencing[44]
1999: *In vitro* DNA colonies in gels[38]
2000: Massively parallel signature sequencing by ligation[47]
2003: Emulsion PCR to generate *in vitro* DNA colonies on beads[42]
2003: Single-molecule massively parallel sequencing-by-synthesis[33,34]
2003: Zero-mode waveguides for single-molecule analysis[57]
2003: Sequencing by synthesis of *in vitro* DNA colonies in gels[49]
2005: Four-colour reversible terminators[51–53]
2005: Sequencing by ligation of *in vitro* DNA colonies on beads[41]
2007: Large-scale targeted sequence capture[93–96]
2010: Direct detection of DNA methylation during single-molecule sequencing[66]
2010: Single-base resolution electron tunnelling through a solid-state detector[141]
2011: Semiconductor sequencing by proton detection[142]
2012: Reduction to practice of nanopore sequencing[143,144]
2012: Single-stranded library preparation method for ancient DNA[145]

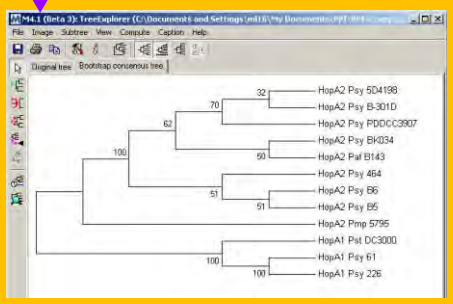
Major steps in Phylogenetic analyses

GTCGAAACCTGCATAGCAGAACGACCCGCGAACACGTTACACTACCAGGTGAGGGACGAGGGGTGCGCAA
GCTCCCCAAGTTTCAAACCCATGGTCGGGGACCACCCTTGGGTGGCCTCGTCCGAACAACGACCCCCCGG
CGCGGAATGCGCCAAGGAAATCAAACTGAACTGCACGCGTCCCCCCCGTTTGCGGGCGGCGGAAGCGTCT
TTCTAAAACACAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCA
TTAGGCCGAGGGCACGTCTGCCTGGGCGTCACACATCGCGTCGCCCCCCAACCCATCACTCCCTTGCGGG
AGTTGAGGCGGAGGGGCGGATAATGGCCTCCCGTGTCTCACCGCGCGGTTGGCCCAAATGCGAGTCCTTG
GCGATGGACGTCACGACAAGTGGTGGTTGTAAAAAGCCCTCTTCTCATGTCGTGCGGTGACCCGTCGCCA
GCAAAATCTCTCATGACCCTGTTGCGCCGAGCCTCGACGCGCGCTCCGACCGCGACCCC

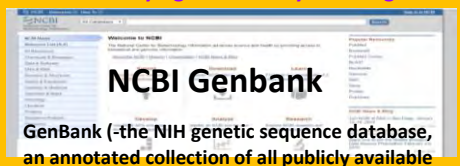
DNA sequence dataset preparation



DNA sequence alignment using ClustalX



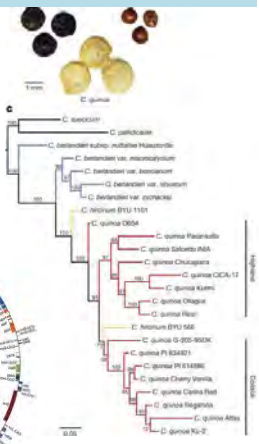
Molecular Phylogenetic analyses using MEGA

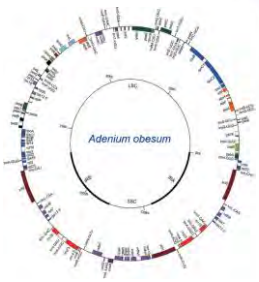


GenBank (-the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences) has a very important role in molecular phylogeny and DNA barcoding.

Nanopore Portable DNA sequencer

Advances in next generation DNA sequencing and its application in systematics





❖ **Whole Chloroplast Genome Sequencing of *Adenium obesum***

- ❖ **Chloroplast (cp) is a special subcellular organelle which contains the entire enzymatic machinery for photosynthesis.**
- ❖ **Chloroplast contains its own small genome of 120–217 kb in size and 110-130 genes, consists of a circular double-stranded DNA.**
- ❖ **The cp genome can be used to investigate molecular evolution and phylogenies.**
- ❖ **The cp genomes are maternally inherited, which is beneficial in genetic engineering.**

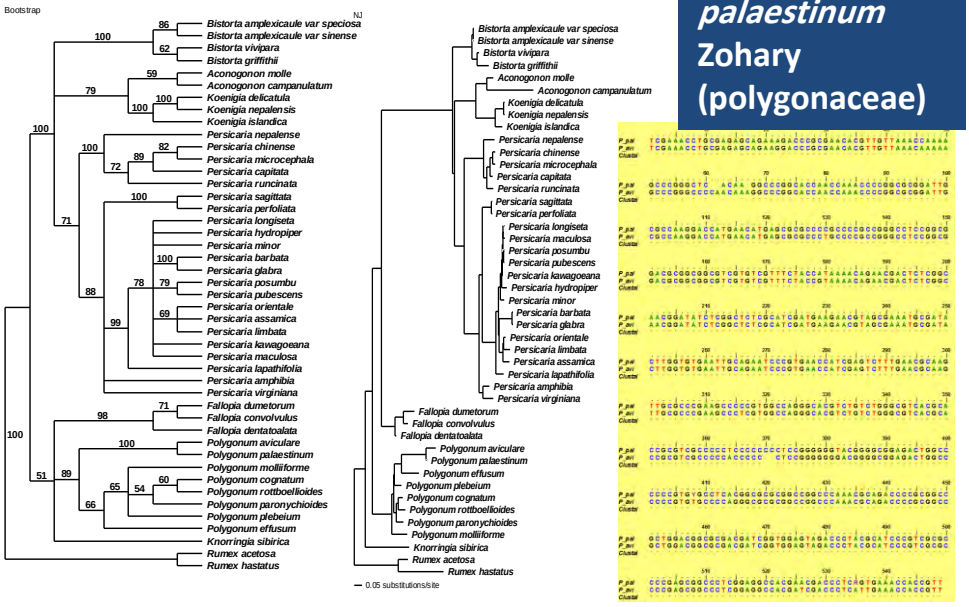
Whole genome sequencing

- ❖ ***Chenopodium quinoa* (quinoa) is a highly nutritious grain identified as an important crop** Jarvis 2017 Nature. ; 542(7641):307-312. **The genome of Chenopodium quinoa.**

Molecular systematic studies on *Polygonum palaestinum* Zohary (polygonaceae) from Saudi Arabia using ITS sequences of nuclear ribosomal DNA

- **The taxonomy of the genus Polygonum is highly controversial because of diverse variation within species among the species has resulted into lack of consensus on taxonomic circumscription. Therefore, there is disagreement among the taxonomists that to which species should be retain within the genus Polygonum and to which species should be elevated to their own genus.**

  - **The genus Polygonum in Saudi Arabia includes P. argyrocoleum Steud. ex Kunze, P. aviculare L. and P. palaestinum Zohary. Two out of these Polygonums of Saudi Arabia i.e. P. argyrocoleum and P. aviculare are common weed distributed throughout. The distribution of P. palaestinum is restricted to Harratal Harra area of Saudi Arabia.**
  - **Decraene and Akeroyd (1988) have segregated Polygonum in the broad sense into two separate tribes, Polygoneae and Persicarieae.**
  - **The systematic status of P. palaestinum is unresolved**



Polygonum palaestinum Zohary (polygonaceae)

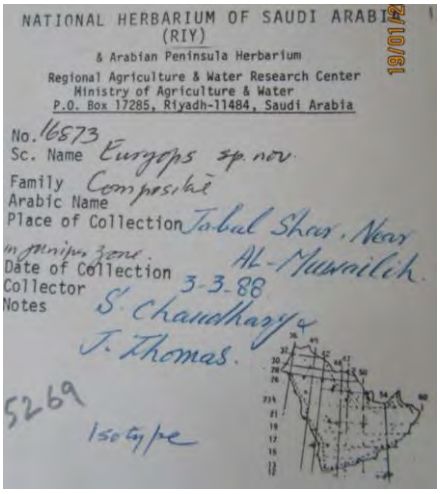
Position in sequence alignment	*Polygonum aviculare*	*Polygonum palaestinum*
22	G	A
46	A	C
59	C	T
61	C	-
62	A	-
67	A	-
130	T	C
179	G	A
316	T	C
364	A	T
370	-	C
371	-	C
382	G	T
408	T	Y C T
413	C	T
415	G	C
526	T	A
535	A	R A G
536	T	G

Phylogenetic Implication of Molecular Genotyping of *Euryops jaberiana* Abedin & Chaudhary (Asteraceae)

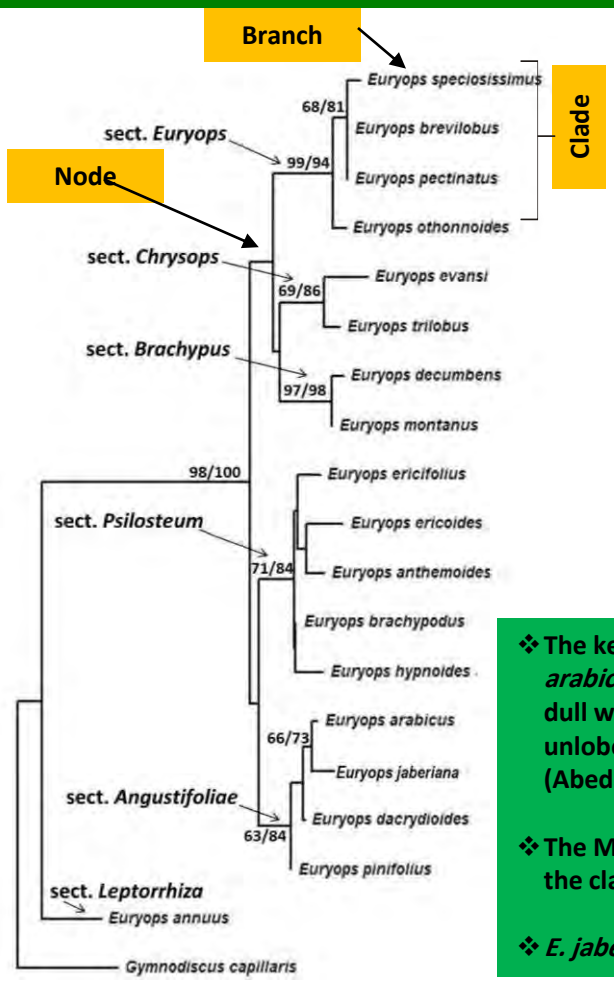
E. arabicus

- In Saudi Arabia, the genus *Euryops* (family Asteraceae) is represented by two species, *viz. E. arabicus* Steud. *ex* Jaub. & Spach, and *E. jaberiana* Abedin & Chaudhary.
- *E. arabicus* is is endemic to Arabian Peninsula, while *E. jaberiana* is endemic to northern Saudi Arabia.
- Morphologically *E. jaberiana* very closely resembles with *E. arabicus* / very narrow differences in m morphological characters (Abedin and Chaudhary, 2000).

- The taxonomic status of *Euryops jaberiana* Abedin & Chaudhary (tribe Senecioneae, was evaluated (Ali et al., 2016) based on molecular phylogenetic analyses of internal transcribed spacer sequence (ITS) of nuclear ribosomal DNA (nrDNA) in order to ascertain its position within the genus.



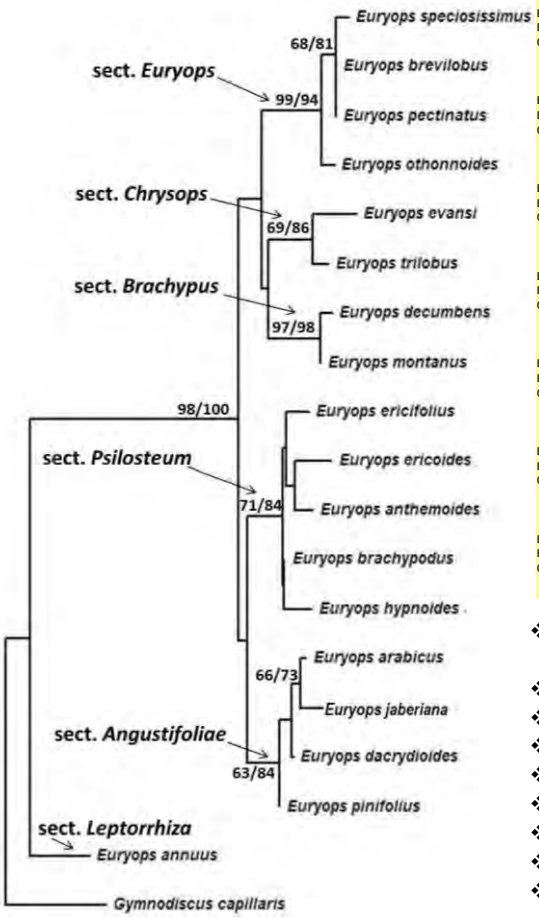
Phylogenetic Implication of Molecular Genotyping of *Euryops jaberiana* Abedin & Chaudhary (Asteraceae) Contd……

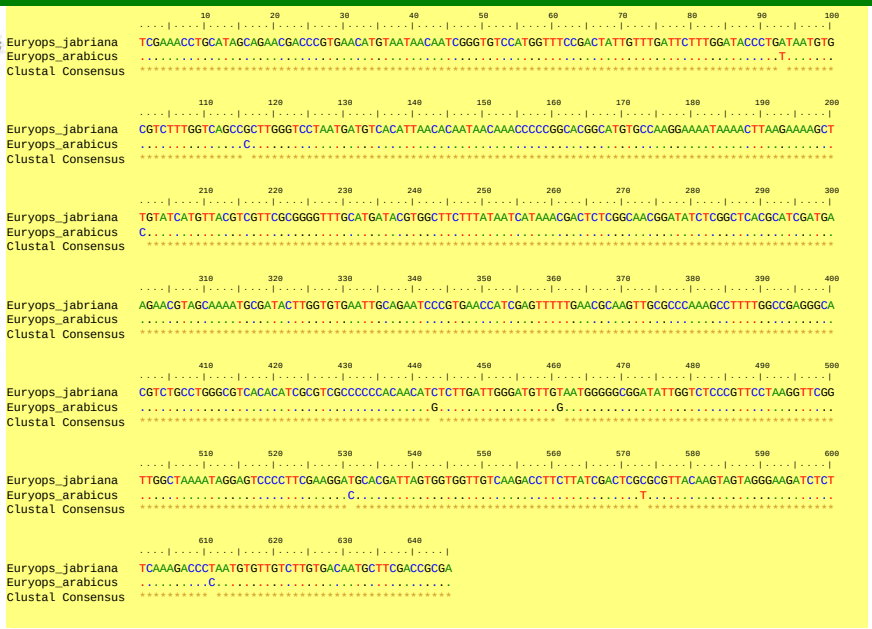


- In molecular taxonomic studies, the most convenient way of presenting taxonomic relationships among a group of organisms is the phylogenetic tree.
- Node: a branch point in a tree
- Branch: defines the relationship between the taxon
- Topology: the branching patterns of the tree
- Branch length: represents the number of changes that have occurred in the branch
- Clade: a group of two or more taxa closed together based on DNA sequences data analsysis
- Maximum parsimony is an optimality criterion under which the phylogenetic tree that minimizes the total number of character-state changes is to be preferred.
- Bootstrap: Bootstrapping is a procedure where DNA sequence data run for the phylogenetic analysis, and the reported value is the percentage of bootstrap replicates, for examples 100 means that the node is well-supported, it showed in all tress.

- The key morphological features which differentiate *E. jaberiana* from *E. arabicus* are: leaves 3-lobed at the tips, pappus hairs transparent or rarely dull white, and achenes glabrescent, while in *E. arabicus*, the leaves are unlobed, pappus hairs are dull white and achene densely lanate hairy (Abedin and Chaudhary, 2000).
- The Maximum Parsimony analyses reveals that *E. jaberiana* nested within the clade of the section *Angustifoliae.*
- *E. jaberiana* shows proximity with *E. arabicus* (66% bootstrap support.

Phylogenetic Implication of Molecular Genotyping of *Euryops jaberiana* Abedin & Chaudhary (Asteraceae) Contd……





- A total of eight specific nucleotide differences were detected between *E. jaberiana* and *E. Arabicus i.e.* at the alignment position:
- 93 (A →T)
- 116 (G → C)
- 201 (T → C)
- 443 (C → G)
- 461 (T → G)
- 531 (T → C)
- 573 (C → T)
- 611 (T → C)

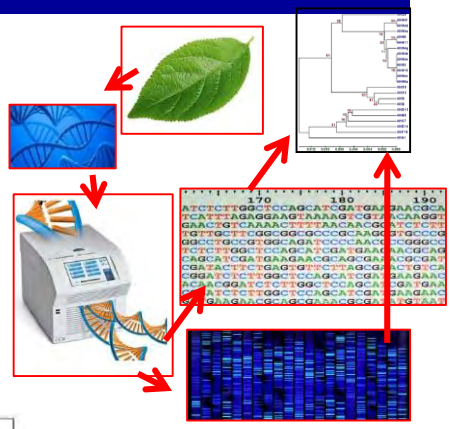
Thus on the basis of phylogenetic relationships of *E. jaberiana* within the genus and nucleotide differences, Ali et al. (2016) recognized *E. jaberiana* as a distinct species and different from *E. arabicus.*

Genetic diversity is the total number of genetic characteristics in the genetic makeup of a species.

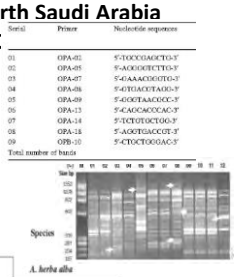
- **Molecular analyses comprise a large variety of DNA molecular markers, which can be employed for analysis of variation.**

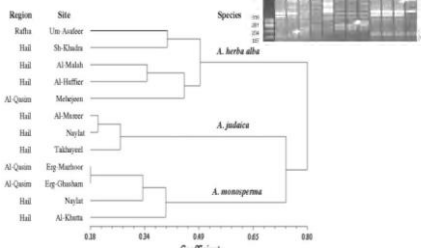
AFLP	Amplified Fragment Length Polymorphism
AP-PCR	Arbitrarily primed PCR
ARMS	Amplification Refractory Mutation System
ASAP	Arbitrary Signatures from Amplification
ASH	Allele-Specific Hybridization
ASLP	Amplified Sequence Length Polymorphism
ASO	Allele Specific Oligonucleotide
CAPS	Cleaved Amplification Polymorphic Sequence
CAS	Coupled Amplification and Sequencing
DAF	DNA Amplification Fingerprint
DGGE	Denaturing Gradient Gel Electrophoresis
GBA	Genetic Bit Analysis
IRAP	Inter-Retrotrasposon Amplified Polymorphism
ISSR	Inter-Simple Sequence Repeats
ISTR	Inverse Sequence-Tagged Repeats
MP-PCR	Microsatellite-Primed PCR
OLA	Oligonucleotide Ligation Assay
RAHM	Randomly Amplified Hybridizing Microsatellites
RAMPs	Randomly Amplified Microsatellite Polymorphisms
RAPD	Randomly Amplified Polymorphic DNA
RBIP	Retrotrasposon-Based Insertion Polymorphism
REF	Restriction Endonuclease Fingerprinting
REMAP	Retrotrasposon-Microsatellite Amplified Polymorphism
RFLP	Restriction Fragment Length Polymorphism
SAMPL	Selective Amplification of Polymorphic Loci
SCAR	Sequence Characterised Amplification Regions
SNP	Single Nucleotide Polymorphism
SPAR	Single Primer Amplification Reaction
SPLAT	Single Polymorphic Amplification Test
S-SAP	Sequence-Specific Amplification Polymorphisms
SSCP	Single Strand Conformation Polymorphism
SSLP	Single Sequence Length Polymorphism
SSR	Simple Sequence Repeats
STMS	Sequence-Tagged Microsatellite Site
STS	Sequence-Tagged-Site
TGGE	Thermal Gradient Gel Electrophoresis
VNTR	Variable Number Tandem Repeats
RAMS	Randomly Amplified Microsatellites

GENETIC DIVERSITY



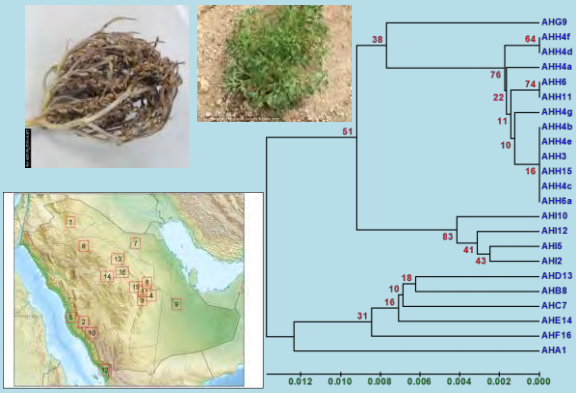
- **Genetic diversity of *Artemisia* in central and north Saudi Arabia based on RAPD**





Badr, A., El-Shazly, H.H., Helail, N.S. et al. Genetic diversity of *Artemisia* populations in central and north Saudi Arabia based on morphological variation and RAPD polymorphism. Plant Syst Evol (2012) 298: 871)

Assessment of genetic diversity of *Anastatica hierochuntica* (kaff maryam) from Saudi Arabia based on Internal Transcribed Spacer sequences of nuclear ribosomal DNA gene



- ***Anastatica hierochuntica*** **(Rose of Jericho) is among the common medicinal plants widely used in Hijaz, Najd, and Al Rub'Al Khali. The plant is prescribed in folk medicine for difficult labor, uterine hemorrhage and to facilitate the expulsion of dead fetuses. A total number of 23 population of Anastatica hierochuntica from Saudi Arabia were sequenced.**
- **The resulted UPGMA tree reveals that the populations of different geographic location sampled in the present study grouped into three major group.**
- **Group I consists of population from Hanifa valley, Summan, Rumah, Hair area, Riyadh, Khurma, and Khoris;**
- **Group II consists of population from Al-Baha, Jeedah, Ranyah and Zazan; and**
- **Group III consists of population from Hail, Darb Al Hafer, Qasim Buraydah, Afif, and Marat), and the groups were according to their geographic locations;**
- **however it was interesting to note that population collected from the geographic location of Haradh and Buseita (Tabarjal) and were nested within the group I and II respectively, which might be due to evolution under reproductive isolation and different environmental conditions, and this may be most probably due to long distance distribution, and possibility of genetic exchange among the populations of *Anastatica hierochuntica* distributed in Saudi Arabia.**

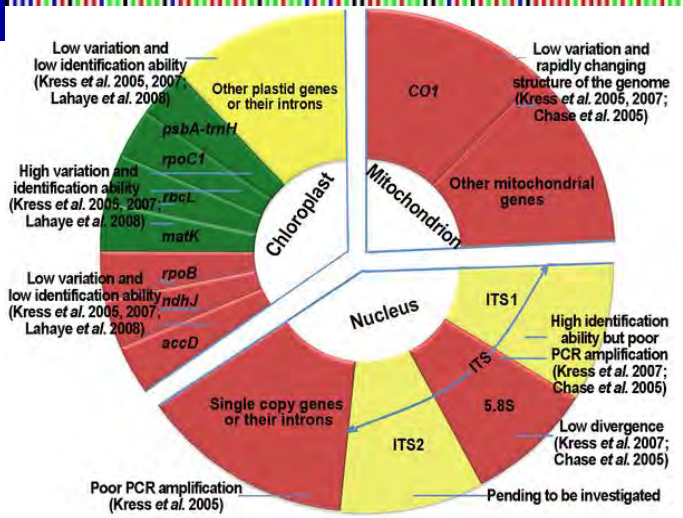
DNA barcoding

- DNA barcoding is a system for fast and accurate species identification that makes ecological system more accessible by using short DNA sequence instead of whole genome and is used for eukaryotes. The short DNA sequence is generated from standard region of genome known as marker. This marker is different for various species like CO1 cytochrome c oxidase 1 for animals, matK for plants and Internal Transcribed Spacer (ITS) for fungus. DNA barcoding has many applications in various fields like preserving natural resources, protecting endangered species, controlling agriculture pests, identifying disease vectors, monitoring water quality, authentication of natural health products and identification of medicinal plants.

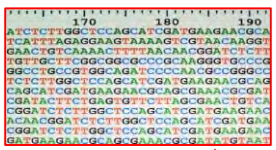
❖ **DNA barcoding can speed up identification of species.**

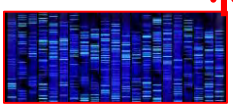
❖ **DNA barcoding can provide an avenue to encourage new participants into taxonomy.**

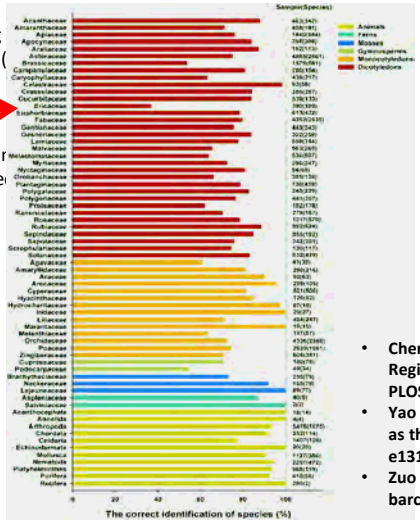
❖ **Raw drug authentication / Medicinal plant identification or authentication**



- In DNA barcoding, complete data set can be obtained from a single specimen irrespective to morphological or life stage characters.
- The core idea of DNA barcoding is based on the fact that the highly conserved stretches of DNA, either coding or
- non coding regions, vary at very minor degree during the evolution within the species.
- Sequences suggested to be useful in DNA barcoding
- NA (e.g. cox1) and chloroplast DNA (
- ai Phylogeny Programs identifi
- eal DNA barc plified ar
- blages of taxa and provides a high level of spe







- Chen S, Yao H, Han J, Liu C, Song J, et al. (2010) Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PLOS ONE 5(1): e8613.
- Yao H, Song J, Liu C, Luo K, Han J, Li Y, et al. (2010) Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. PLoS ONE 5(10): e13102.
- Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, Zhou S. (2011) DNA barcoding of Panax species. Planta Med. 2011 Jan;77(2):182-7.

THANKS

DETAILED LECTURE

جامعة الملك سعود
King Saud University
1957

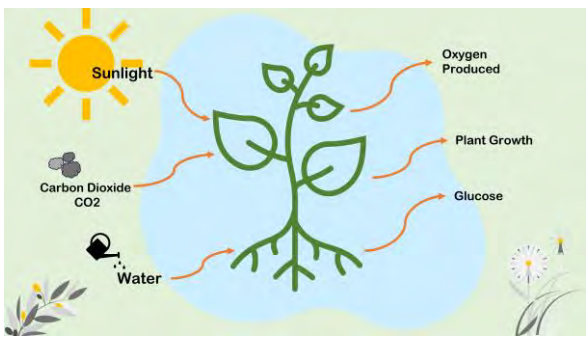
Advance Experimental Taxonomy

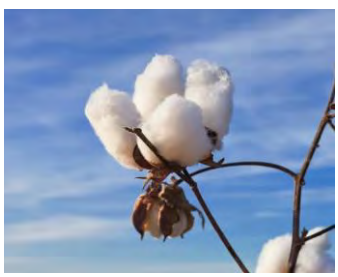
Chapter 1- Quick Note Principles of Plant taxonomy

Chapter 1- Quick Note Principles of Plant taxonomy

We study plants because:

- **Plants produce oxygen. We breathe oxygen. We cannot live without oxygen.**
- **Plants convert Carbon dioxide gas into sugars through the process of photosynthesis.**
- **Every things we eat comes directly or indirectly from plants.**
- **Plants provide fibres for paper or fabric.**
- **Many chemicals produced by the plants used as medicine.**
- **Study of plants science helps to conserve endangered plants.**
- **Plants can be a source of biofuels. Sugars, starches and cellulose can be fermented into ethanol. Ethanol is used as fuel.**
- **Study of plants science helps to learn more about the natural world**





- **Plant Biodiversity:** diversity among and within plant and animal species in an environment
- We have millions of different kind of plants, animals and microorganism. There are about 15000 species of Mossess, 13000 species of fern plants, 900 species of Gymnosperms, 250000 species of angiosperms, 10000 species of algae, 5.1 million of fungi species. We need to scientifically identify, name and classify all the living organism.
- Taxonomy / Systematics is the branch of science deals with classification of organism.








Plant Biodiversity

Q: Why we keep the stuffs of our home at the fixed place or arrange into some kinds of system?

- **Every Human being is a Taxonomist**

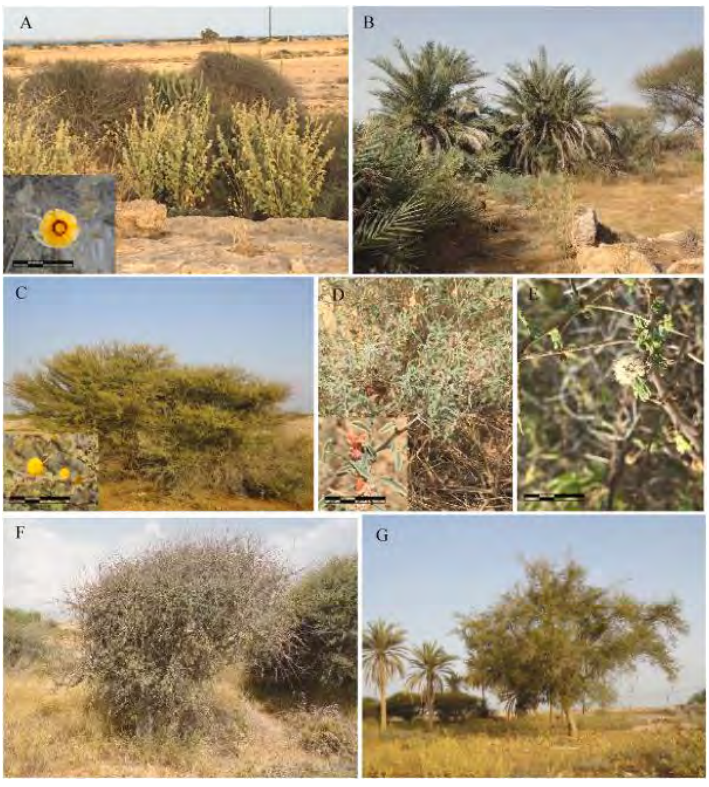
The idea that "**Every human being is a taxonomist**" suggests that humans naturally categorize and organize the world around them. Just like in taxonomy, where organisms are classified into different groups (families, orders, etc.) based on shared characteristics, we, too, group and categorize our belongings. For example, we may separate kitchen items from cleaning supplies, or books from personal items, based on their function or type. This helps us not only in maintaining a clean environment but also in managing our space effectively.

Objective / Goals / Aims of Plant Taxonomy

- To provide an inventory of plant taxa for local, regional or continental needs.
- To establish suitable method for identification, nomenclature and description of plant taxa.
- Classification of organism into classes, Order, Families, Genera, and species
- To provide significantly valuable information concerning wild and medicinal species, endangered species, unique plants, genetic and ecological diversity

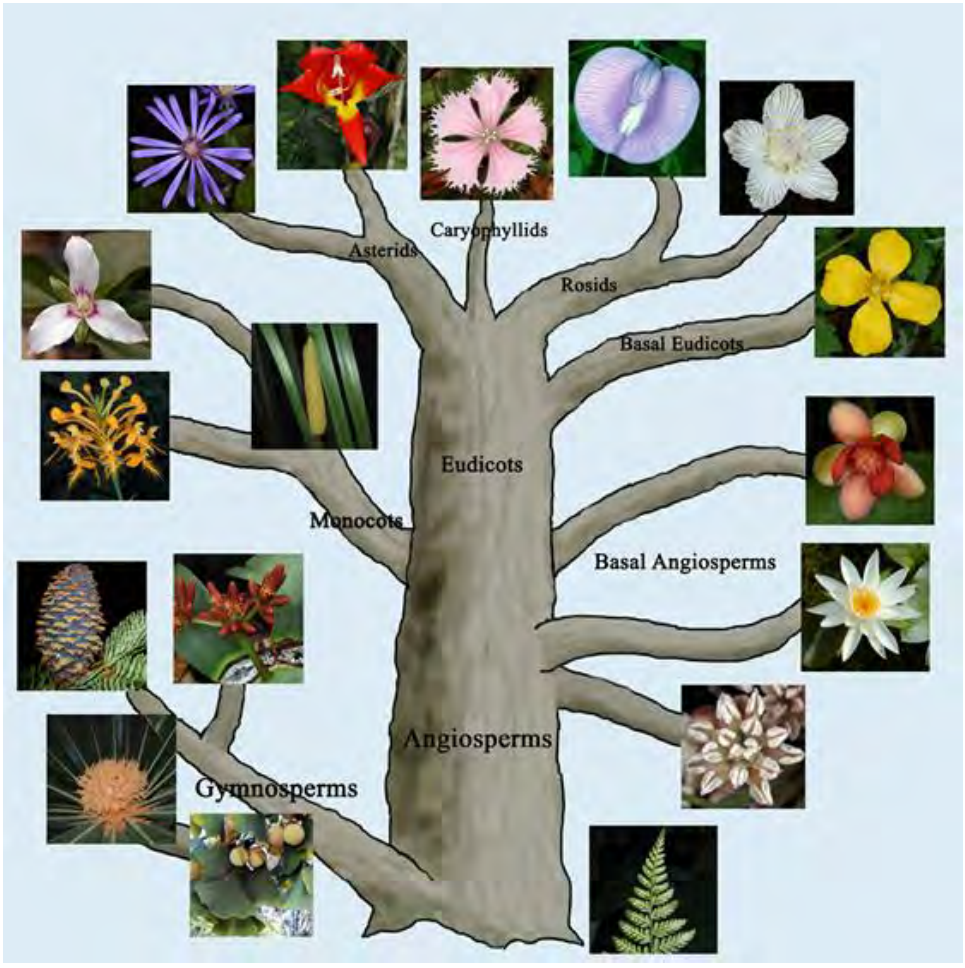
Scope of Taxonomy

- Taxonomy is one of the oldest sciences.
- It provides thorough knowledge of living species and their various forms.
- All the branches of biology are dependent on taxonomy for proper identification the species.
- It has been proceeded further incorporating data from phytochemistry, cyto-genetics supported by proper computation.



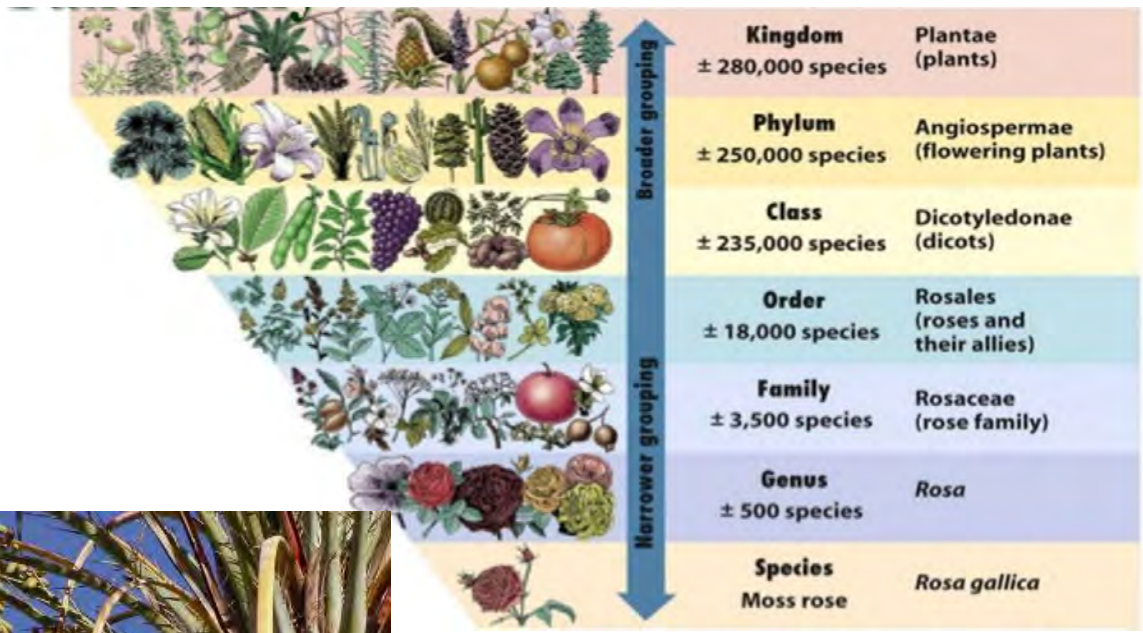
Basic components (Principles) of Plant Taxonomy / Plant Systematics

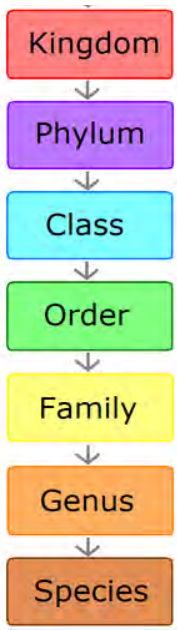
- Plant collection, Preservation and Documentation
- Plant Structure (Taxonomic Terminology, Taxonomic description of external and internal morphology)
- Taxonomic Identification
- Scientific Nomenclature / Botanical Nomenclature : Nomenclature deals with the application of a correct name to a plant or a taxonomic group. Scientific names are necessary because the same common name is used for different plants in different areas of the world.
- Taxonomic Classification (History and Systems of Plant Classification)
- Taxonomic evidences / Source of data (Morphology, Anatomy, Embryology, palynology, Micromorphology, Chemistry, DNA etc.) in plant taxonomy



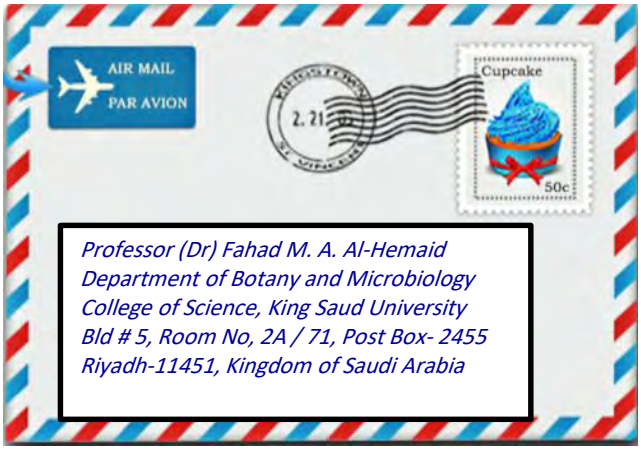
TAXONOMIC HIERARCHY

- Carrolus Linnaeus first adopted the hierarchic system of taxonomy classification in 1753.
- The succession groups are as follow:
- **Species**: Organisms sharing a set of biological traits and reproducing only their exact kind.
- The lowest major group, representing plants and animals referred to as Species.
- **Species is the fundamental unit in taxonomy**
- **Genus**: Genus are the closely related species
- **Family** : Family is the closely related genera
- **Order** : Order is the closely related families
- **Class** : Class are the closely related order
- **Division / Phylum**: Division or Phylum is the related classes
- **Kingdom**: Kingdom is the related Division / Phylum





Kingdom: Plantae
Class: Angiosperms
Order: Arecales
Family: Arecaceae
Genus: *Phoenix*
Species: *P. dactylifera*



Professor (Dr) Fahad M. A. Al-Hemaid
Department of Botany and Microbiology
College of Science, King Saud University
Bld # 5, Room No, 2A / 71, Post Box- 2455
Riyadh-11451, Kingdom of Saudi Arabia

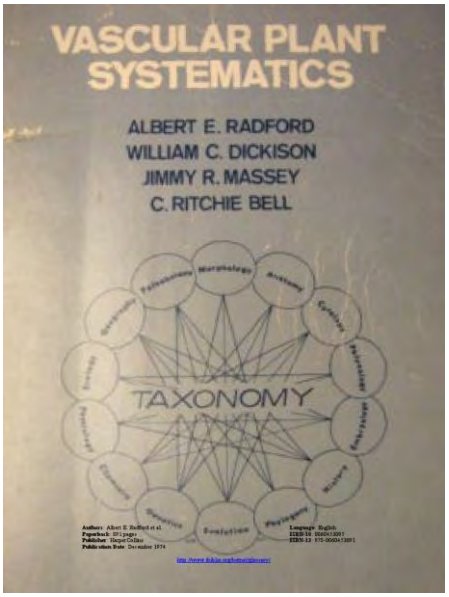
Types of Taxonomy / Taxonomic Studies / Plant Taxonomic Classification

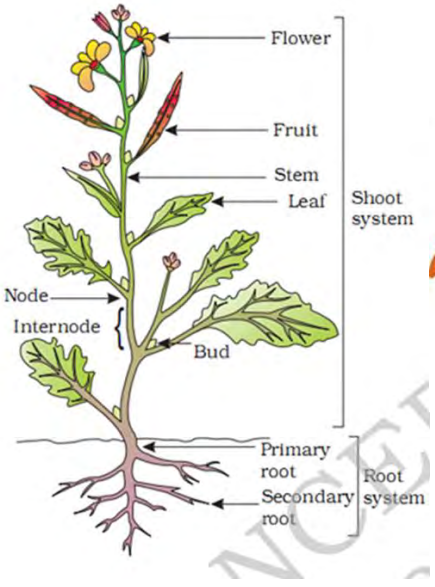
Alpha (α) Taxonomy / classical taxonomy:- It involves description and naming of organisms. It is the parent of other types of taxonomy.

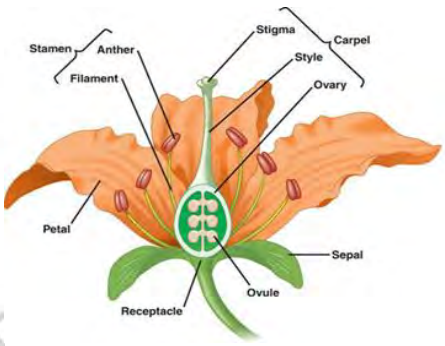
Omega (Ω) Taxonomy: - It is the modern experimental taxonomy in which the taxonomic activities have been enriched with data from ecology, phyto-chemistry, phyto-geography, cyto-genetics and physiology coupled with adequate computation.

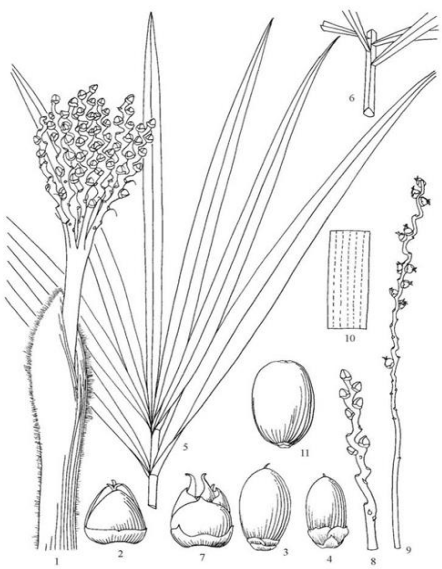
Herbarium: Plant collecting, Preservation and Documentation

- **To make a herbarium specimen, the plant is collected, and notes are made about it. The plant is then pressed until dry between blotters that absorb moisture and mounted onto a herbarium sheet with a suitable label, and stored in steel cabinet arranged into some system of classification.**
- **Herbarium techniques involve : (i) Collection, (ii) Drying, (iii) Poisoning, (iv) Mounting, (v) Stitching, (vi) Labelling, and (vii) Deposition.**
- **Flora = it is the documentation of plants occurring in a particular region.**
- The FLORA is the main Resources of Taxonomic Information
- **A HERBARIUM is a collection of dried plants systematically named and arranged for ready reference and study.**



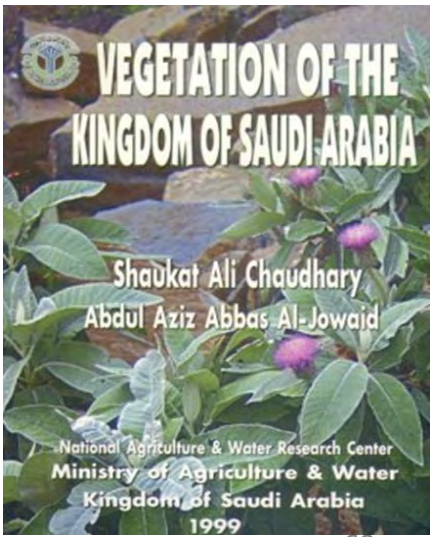






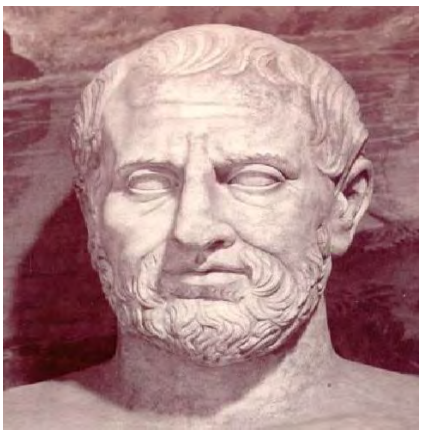
Phoenix dactylifera Linnaeus, Sp. Pl. 2: 1188. 1753.

Stems solitary or clustered and then with few shoots, to 30 m tall, to 50 cm in diam., rough with persistent, diamond-shaped leaf bases. Leaves 3-5 m; sheath and petiole to 1 m; rachis 1-2 m; acanthophylls many per side of rachis; pinnae to 200 per side of rachis, linear, irregularly arranged and spreading in different planes; middle pinnae to 40 × 2 cm. Male inflorescences erect, to 1 m, with many rachillae, these ca. 30 cm; female inflorescences erect, becoming pendulous, to 2 m, with to 150 rachillae, these to 40 cm. Fruits variable in shape, usually oblong, to 7 × 3 cm, brown or black; endosperm homogeneous.



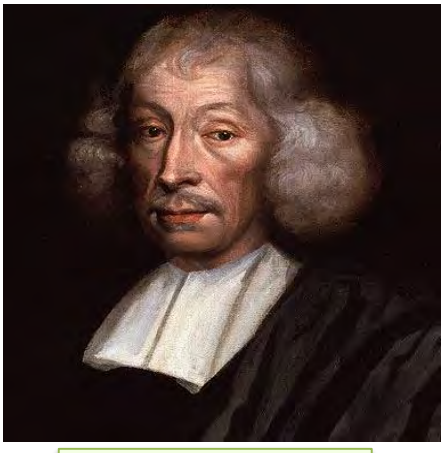
SYSTEM OF PLANT CLASSIFICATION

Preliterate Mankind / Folk taxonomies:



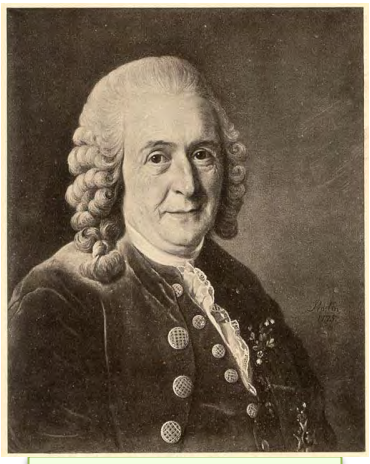
Theophrastus (372 BC to 287 BC):

Andrea Cesalpino (1519-1603)



John Ray (1627-1705)

J. P. de Tournefort (1656-1708)



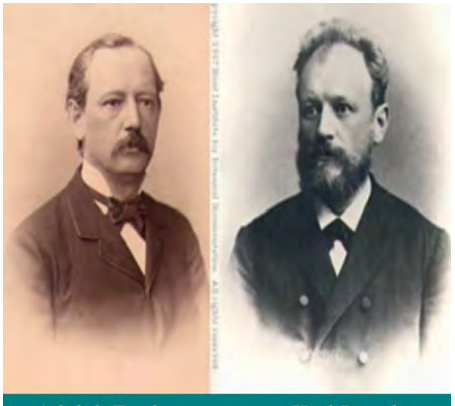
Carolus Linneaus (1753)

Michel Adanson (1727-1806)

Jean B.P. Lamarck (1744-1829)

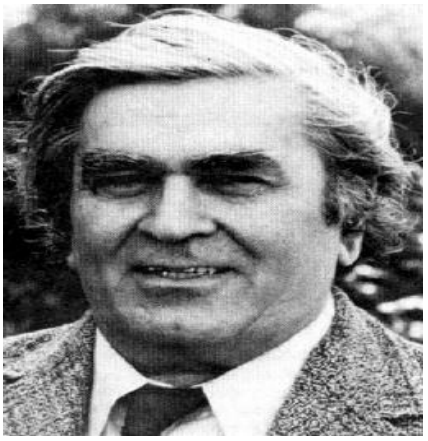
Antoine Laurent de Jussieu (1748-1836)

de Candolle (1778–1841)

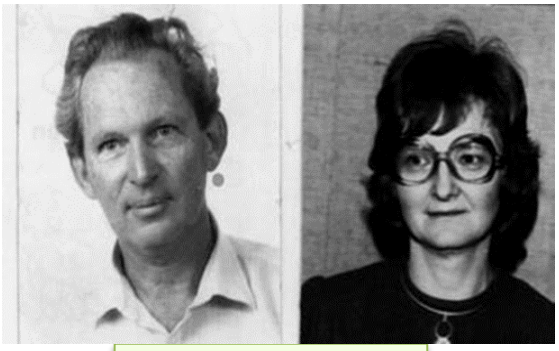


Charles E. Bessey (1845-1915)

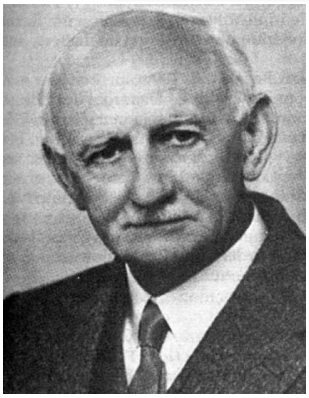
Auther Cronquist 1968



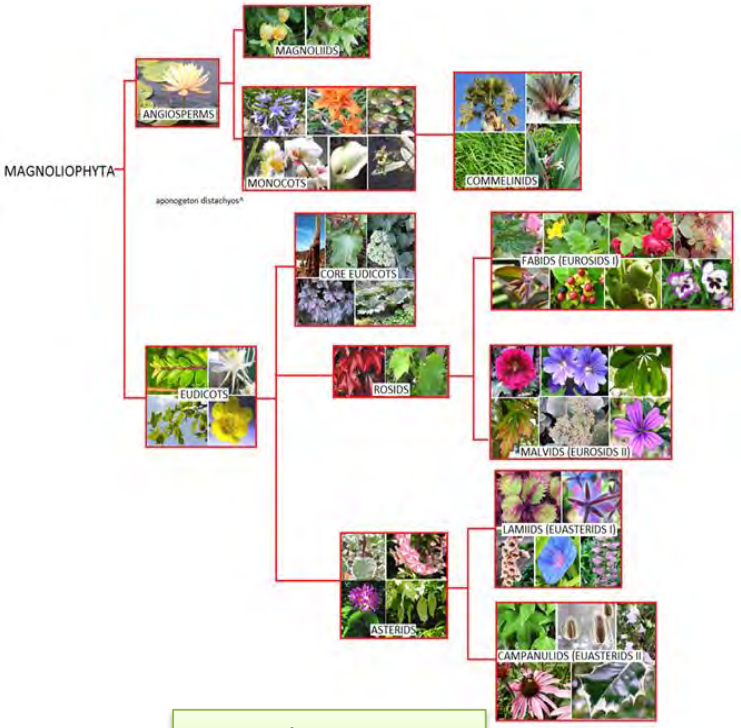
Armen Takhtajan 1969



Rolf Dahlgren (1932-87)



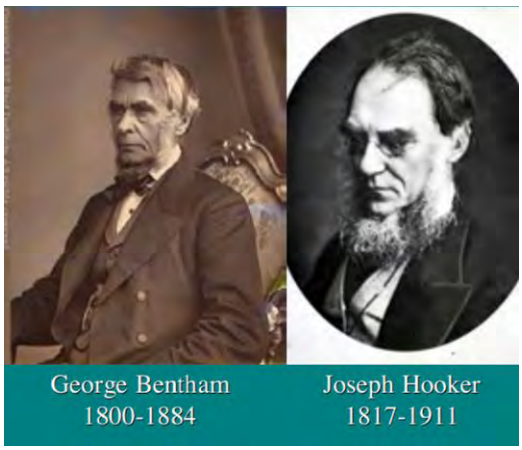
John Hutchinson (1884-1972)



APG Angiosperm Phylogeny Group (1998)

Bentham and Hooker System of Plant Classification

- Bentham and Hooker, two English botanists, represented the most well developed natural system of plant classification. The classification was published in a three-volume work Genera plantarum (1862-83).
- Hooker supervised the publication of Index Kewensis (2 volumes, 1893), listing the names of all known species and their synonyms.
- Many important herbaria of the world have specimens arranged according to Bentham and Hooker system of plant classification.



- Bentham and Hooker recognized three class:

Class Dicotyledones:

Subclass POLYPETALE with three series

Series 1. THALAMIFLORÆ, Series 2. DISCIFLORÆ, Series 3. CALYCIFLORÆ;

Subclass DICOTYLEDONES (GAMOPETALÆ) with three series that is Series 1. INFERÆ, Series 2. HETEROMERÆ, Series 3. BICARPELLATÆ, and

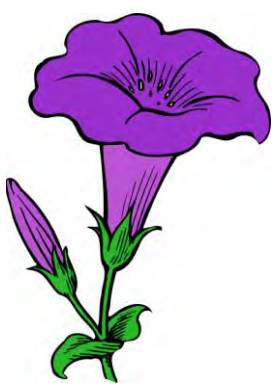
Subclass DICOTYLEDONES MONOCHLAMIDEÆ.

Class Gymnospermeæ (Gymnosperms are placed between Dicotyledons and Monocotyledons)

Class Monocotyledones

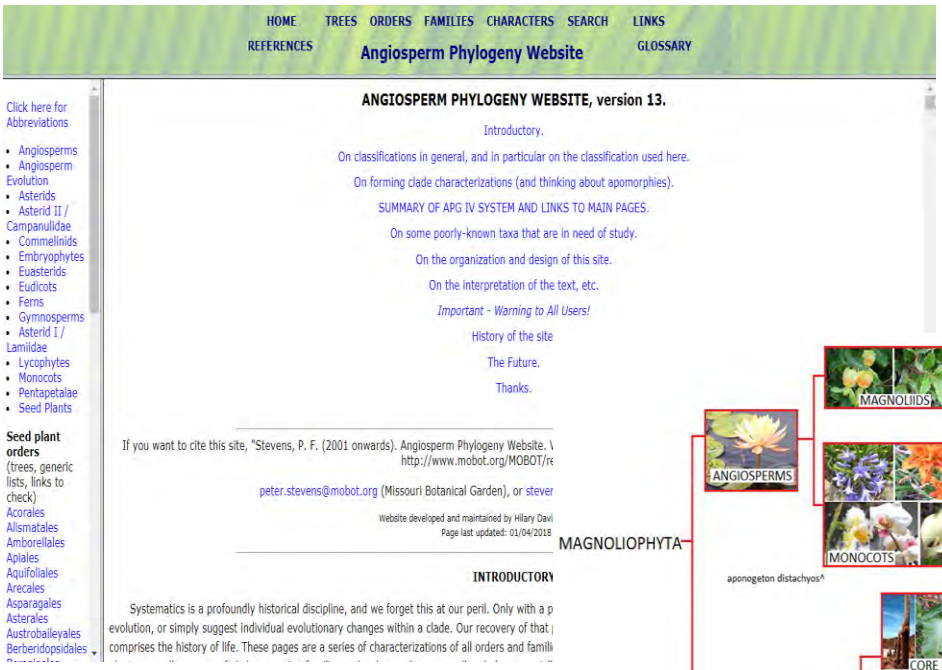

A Flower

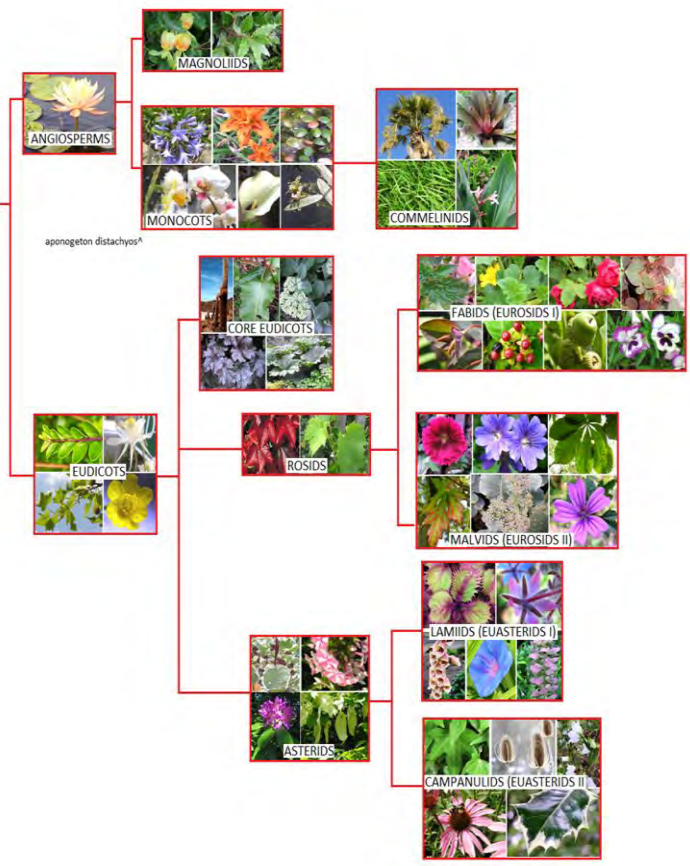

Corolla
(whorl of petals)



Angiosperm Phylogeny Group (APG)

- ❖ **The APG system of flowering plant classification is the modern, mostly molecular-based, system of plant taxonomy for flowering plants (angiosperms) being developed by the Angiosperm Phylogeny Group (APG).**
- ❖ **The APG was first published in 2008.**
- ❖ **Currently the APG IV system recognizes a total of 64 angiosperm orders and 416 families.**
- ❖ **The families in APG classification have been grouped into 40 putative monophyletic orders under a small number of informal monophyletic higher groups: monocots, commelinoids, eudicots, core eudicots, rosids, eurosids I, eurosids II, asterids, euasterids I and euasterids II**





SCIENTIFIC NOMENCLATURE / BOTANICAL NOMENCLATE :

Nomenclature deals with the application of a correct name to a plant or a taxonomic group.

- ❖ **We have millions of species distributed in different geographical regions of the world.**
- ❖ **The Scientific names (Botanical name and Zoological name)of the living organism (Plants and Animals) are necessary because the same common name is used for different plants / Animals in different areas of the world.**
- **Swedish Botanist Carolus Linnaeus introduced Binomial Nomenclature.**
- **The Binomial nomenclature uses two Latin words to indicate the genus and the species. The first word is the genus and the second word is the species. Example- the botanical name of Dates is *Phoenix dactylifera***

Species Concept

- o **Species is the basic unit of classification**
- o **Plants in the same species consistently produce plants of the same types**
- o **The name of the plants must should be written in italics. For example Phoenix dactylifera**

TAXONOMIC RANKS OF LAND PLANTS	ENDING	EXAMPLE TAXON
Kingdom	(various)	**Plantae**
Phylum [Division]	**-phyta**	**Magnoliophyta**
Subphylum [Subdivision]	-phytina	Magnoliophytina
Class	**-opsida**	**Asteropsida**
Subclass	-idae	Asteridae
Order	**-ales**	**Asterales**
Suborder	-ineae	Asterineae
Family	**-aceae**	**Asteraceae**
Subfamily	-oideae	Asteroideae
Tribe	-eae	Heliantheae
Subtribe	-inae	Helianthinae
Genus	(various)	***Helianthus***
Subgenus	(various)	*Helianthus*
Section	(various)	*Helianthus*
Series	(various)	*Helianthus*
Species [abbr. sp. (sing.), spp. (pl.)]	(various)	***Helianthus annuus***
Subspecies [abbr. subsp. or ssp. (sing.), subspp. or sspp. (pl.)]	(various)	*Helianthus annuus* ssp. *annuus*
Variety [abbr. var. (sing.), vars. (pl.)]	(various)	*Helianthus annuus* var. *annuus*
Form [abbr. f.]	(various)	*Helianthus annuus* f. *annuus*

❖ **Binomial Nomenclature and Carolus Linneaus System of Plant Classification**

❖ **Taxonomic Systems of Classification: Ideally our systems of classification should allow us to place similar species of plants together in the same category.**

❖ **There are two types of Classification Schemes:**

❑ **Artificial taxonomy was a system of grouping unrelated plant species by a common criteria (i.e. a flowers sexual organs)**

❑ **Natural classification reflects relationships among taxon**

➢ **Carolus Linneaus was a Swedish botanist.**

➢ **Carolus Linneaus traveled to Lapland (Blue Lake, CA) and collected large number of plants.**

➢ **Carolus Linneaus introduced Binomial Nomenclature.**

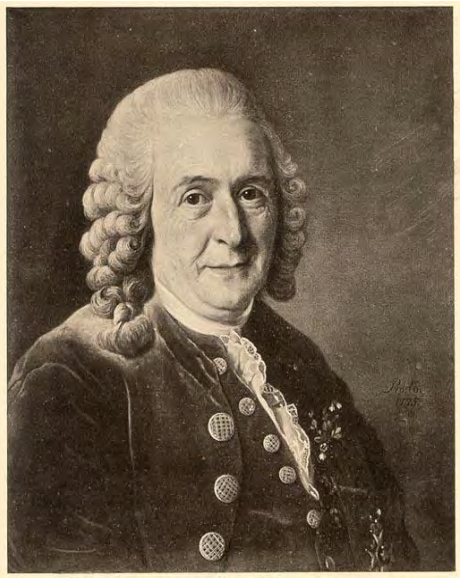
Binomial nomenclature = Uses two Latin words to indicate the genus and the species. The first word is the genus and the second word is the species. Example- the botanical name of dates is *Phoenix dactylifera*

➢ **Carolus Linneaus published the book 'Species Plantarum' in 1753.**

➢ **Carolus Linneaus classified the plants based on the plant's method of reproduction and structure of reproductive parts.**

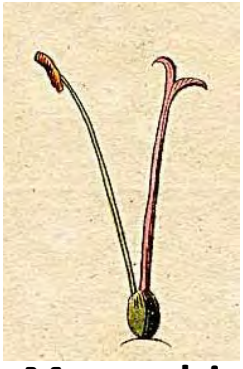
➢ **Produced his sexual system of classification (Artificial classification)**

➢ **Carolus Linneaus divided plants into 24 classes. The Classes in the Linneaus is based largely on the amount, union and length of stamens**



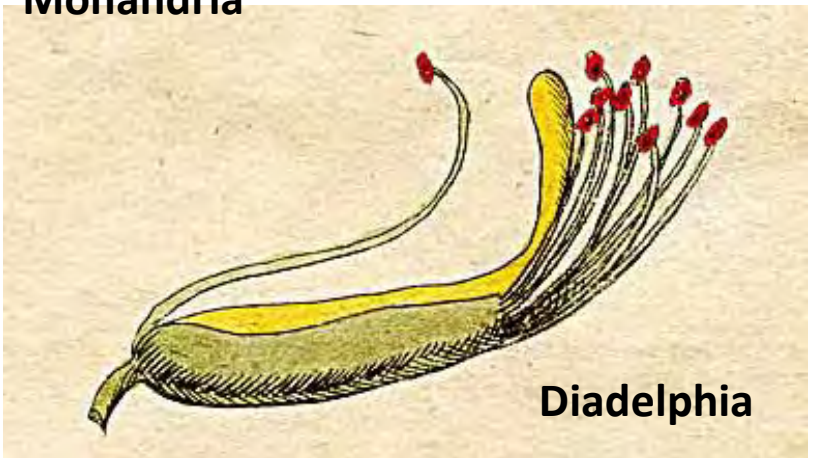
Classes

1. Monandria- stamen one
2. Diandria- stamens two
3. Triandria- stamens three
4. Tetrandria- stamens four
5. Pentandria- stamens five
6. Hexandria- stamens six
7. Heptandria- stamens seven
8. Octandria- stamens eight
9. Ennandria- stamens nine
10. Decandria- stamens ten
11. Dodecandria- stamens 11-19
12. Icosandria- stamens 20 or more, on the calyx
13. Polyandria- stamens 20 or more, on the receptacle
14. Didynamia- stamens didynamous; 2 short, 2 long
15. Tetradynamia- stamens tetradynamous; 4 long, 2 short
16. Monadelphia- stamens monadelphous; united in 1 group
17. Diadephia- stamens diadelphous; united in 2 groups
18. Polyadelphia- stamens polyadelphous; united in 3 or more groups
19. Syngenesia- stamens syngenesious; united by anthers only
20. Gynandria- stamens united with the gynoecium
21. Monoecia- plants monoecious
22. Dioecia- plants dioecious
23. Polygamia- plants polygamous
24. Cryptogamia- flowerless plants

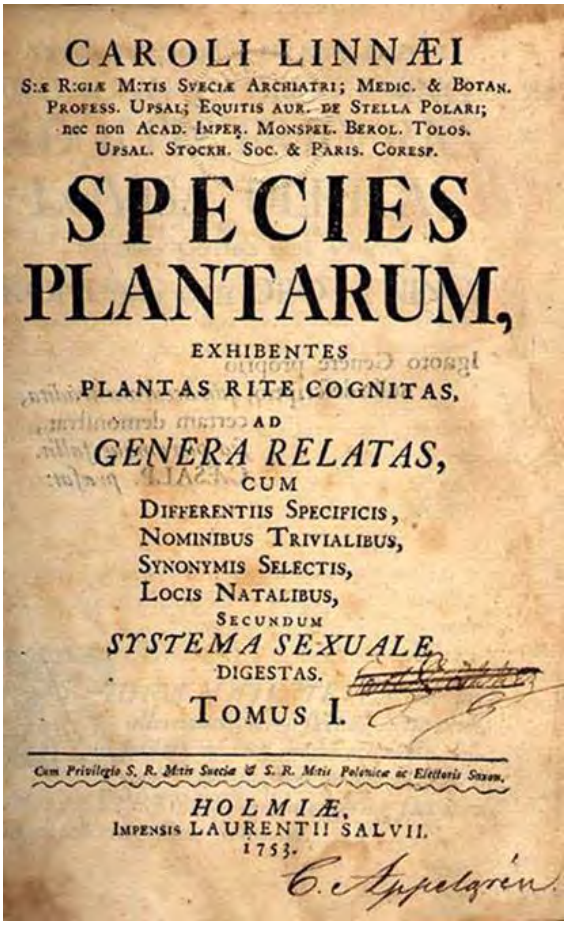


Monandria

Triandria



Diadelphia


CAROLI LINNÆI
S:æ R:giæ M:tis Sveciæ Archiatri; Medic. & Botan. Profess. Upsal; Equitis aur. de Stella Polari; nec non Acad. Imper. Monspel. Berol. Tolos. Upsal. Stockh. Soc. & Paris. Coresp.

SPECIES PLANTARUM,

EXHIBENTES
PLANTAS RITE COGNITAS,
AD
GENERA RELATAS,
CUM
DIFFERENTIIS SPECIFICIS,
NOMINIBUS TRIVIALIBUS,
SYNONYMIS SELECTIS,
LOCIS NATALIBUS,
SECUNDUM
SYSTEMA SEXUALE
DIGESTAS.

TOMUS I.

Cum Privilegio S. R. Mtis Suecia & S. R. Mtis Polonice ac Electoris Saxon.

HOLMIÆ,
IMPENSIS LAURENTII SALVII.
1753.

International Code of Nomenclature (ICN)

The current activity of botanical nomenclature is governed by the International Code of Nomenclature (ICN) published by the International Association of Plant Taxonomy (IAPT).

The Code is divided into 3 divisions:

I. Principles

II. Rules and recommendations

III. Provisions for the governance of the Code

Principles of ICBN

- **Botanical Nomenclature is independent of Zoological Nomenclature. The Code applies equally to the names of taxonomic groups treated as plants whether or not these groups were originally so treated.**
- **The application of names of taxonomic groups is determined by means of nomenclatural types / TYPIFICATION.**
- **Nomenclature of a taxonomic group is based upon Priority Of Publication.**
- **Each taxonomic group with a particular circumscription, position and rank can bear Only One Correct Name, the earliest that is in accordance with the rules.**
- **Scientific names of taxonomic groups are treated as LATIN, regardless of derivation.**
- **The rules of nomenclature are Retroactive, unless expressly limited.**

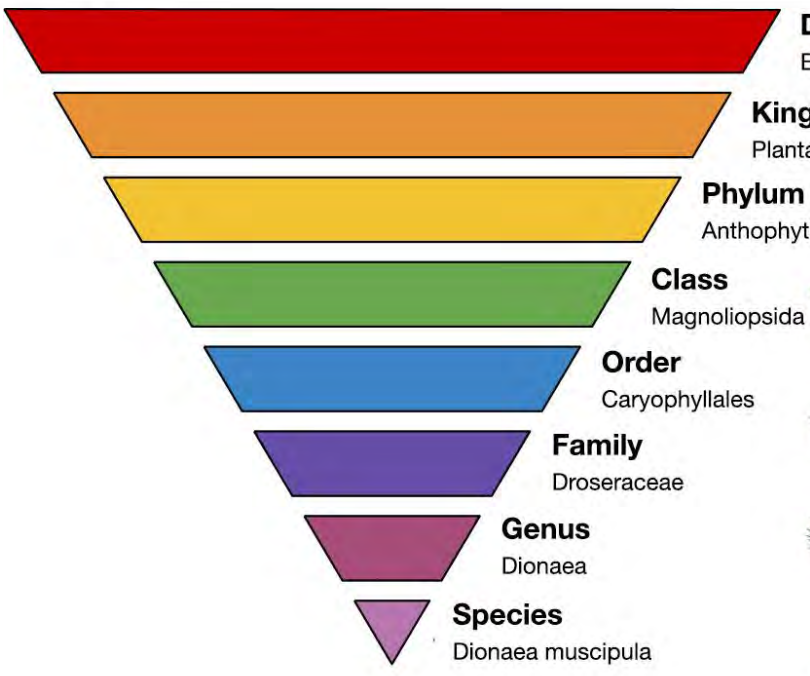
- **Generic Name:** The Generic name is usually a noun and singular, which is spelled or written with a capital letter.
- **Specific Epithet:** The specific epithet is often an adjective and it is written with a small initial letter.
- **In the hand written manner, both the generic names and specific epithet should be underlined, while if printed it should be in italics.**

Synonyms and related terminology

- **Synonyms:** A name rejected due to misuse or difference in taxonomic judgement.
- **Basionym:**
  - The basionym is the first name ever given to a taxon. Further studies and revisions may reject the basionym as the most correct one, but it still is useful as a nomenclatural reference for that species.
  - Also, according to the priority rules of the ICBN, after a taxonomic revision that results in a species being reclassified in another genus, the specific epithet must remain the same as the one in the Basionym.
  - A short example: Linnaeus classified the Tea Plant as *Thea sinensis*. Some decades later, Sweet noticed that the genus *Thea* was not really different from the genus *Camellia*, and renamed all the *Theas* as *Camellias*. *Thea sinensis* became *Camellia sinensis*, because he had to keep the specific epithet the same as the original name (Basionym) for that species, given by Linnaeus.
- **Homonym:** A case in which two or more identical names are baed on different type , of which only one can be a legitimate name , is called as homonym.
- **Tautonym:** A case in which name of genus and the name of the species is the same.

Names of Taxa

Rank	*Ending*	*Example*
Kingdom	-bionta	Chlorobionta
Division	-phyta	Magnoliophyta
	-mycota (Fungi)	Eumycota
Subdivision	-phytina	Pterophytina
	-mycotina (Fungi)	Eumycotina
Class	-opsida	Magnoliopsida
	-phyceae (Algae)	Chlorophyceae
	-mycetes (Fungi)	Basidiomycetes
Subclass	-opsidae	Pteropsidae
	-idae (Seed plants)	Rosidae
	-physidae (Algae)	Cyanophysidae
	-mycetidae (Fungi)	Basidiomycetidae
Order	-ales	Rosales
Suborder	-ineae	Rosineae
Family	-aceae	Rosaceae
Subfamily	-oideae	Rosoideae
Tribe	-eae	Roseae
Subtribe	-inae	Rosinae
Genus	*-us, -um, -is, -a, -on*	*Pyrus, Allium, Arabis, Rosa, Polypogon*
Subgenus		*Cuscuta* subgenus *Eucuscuta*
Section		*Scrophularia* section *Anastomosanthes*
Subsection		*Scrophularia* subsection *Vernales*
Series		*Scrophularia* series *Lateriflorae*
Species		*Rosa canina*
Subspecies		*Crepis sancta* subsp. *bifida*
Varietas		*Lantana camara* var. *varia*
Forma		*Tectona grandis* f. *punctata*



Typification: Type Specimen is the one representative of the taxon.

- **Holotype:** A specimen designated by the author in the original publication (nomenclatural type).
- **Isotype:** A duplicate specimen of the holotype collected at the same time and place (may be in other herbarium).
- **Lectotype:** A specimen chosen from the author's original material when no holotype has been designated.
- **Neotype:** A specimen selected when all original specimens have been destroyed

Author Citation, Effective Publication and Principle of Priority

Author Citation

- **For a name to be complete, it should be accompanied by the name of the author or authors who first published the name validly. The names of the authors are commonly abbreviated, Example L. for Carolus Linnaeus**
- ***Aizoon canariense* L.**
- ***Tribulus macropterus* var. *arabicus* (Hosni) Al-Hemaid & J. Thomas**

Basic structure of a taxonomic Research papers / Recent publication of a new species in taxonomic journal

Ann. Bot. Fennici 53: 37–39 ISSN 0003-3847 (print) ISSN 1797-2442 (online)
Helsinki 4 January 2016 © Finnish Zoological and Botanical Publishing Board 2016

Silene langshanensis (Caryophyllaceae), a new specie
from Inner Mongolia, China

Li-Qing Zhao[1,*], Zhi-Ming Xin[2] & Yi-Zhi Zhao[1]

[1] College of Life Science, Inner Mongolia University, Hohhot 010021, China (*corresp
author's e-mail: zhaotieniu@126.com)
[2] Experimental Center for Desert Forestry, Chinese Academy of Forestry, Dengkou,
Mongolia 015200, China

Received 22 Apr. 2015, final version received 9 Oct. 2015, accepted 9 Oct. 2015

Zhao L.Q., Xin Z.M. & Zhao Y.Z. 2016: *Silene langshanensis* (Caryophyllaceae), a new sp
from Inner Mongolia, China. — *Ann. Bot. Fennici* 53: 37–39.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin *sp. nova* (Caryophylla
is described and illustrated from Inner Mongolia, China. It appears to be most cl
related to *S. scabrifolia* of *Silene* sect. *Holopetalae*. *Silene langshanensis* can b
tinguished by the basally pubescent carpophore, petals with obtuse auricles, stem
leaves with dense, short hairs, and by the glabrous calyx.

In total, there are about 600 species of *Silene s. lato* (Caryophyllaceae) (Zhou *et al.* 2001). They are distributed mainly in the northern temperate regions, but occur also in Africa and South America (Zhou *et al.* 2001). Among these species, 110 are known from China, of which 67 are endemic. Twenty of the endemics (nine species of *Silene* s. *stricto*, nine of *Melandrium*, one of *Cucubalus* and one of *Lychnis*) are found in Inner Mongolia.

In September 2008 and later, in 2014, the authors Zhao and Xin collected specimens of *Silene* from Langshan in Bayannaoer (Inner Mongolia) from desert steppe communities on mountain slopes at 1150–1400 m a.s.l. After careful study, we concluded that the specimens represented an undescribed species of *Silene*.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin, *sp. nova* (Fig. 1)

HOLOTYPE: China. Inner Mongolia, Bayannaoer, Dengkou, Mt. Langshan, 40°43'58.4"N, 106°22'28.5"E, on stony mountain slopes, 1371 m a.s.l., 9 June 2014 *Li-Qing Zhi-Ming Xin, Shuai Qin & Long Chen N14.001* (HIM
PARATYPES (all HIMC[1]): Same location as holotype, 2014 *Li-Qing Zhao, Zhi-Ming Xin, Shuai Qin & Long N14.002, N14.003, N14.004*; Mt. Langshan, 40°39'2 106°23'14.1"E, on stony mountain slopes, alt. 1185 m 10 September 2008 *Li-Qing Zhao & Zhi-Ming Xin N08*

Herbs perennial, 20–50 cm tall. Roots n lignified. Plants cespitose with erect stems, branched or sparsely branched, pubescent, parts glabrescent. Basal leaves oblance 20–60 × 2–6 mm, both surfaces pubescent, gins ciliate, base attenuate into a long pe apex acute; cauline leaves 3–7 pairs, linear ceolate or linear-lanceolate, with short, ax sterile branches or sometimes elongated fl ing axillary branches. Flowers in a racemi like thyrse; cymes alternating (resulting suppression of opposite cyme at same no opposite, 1-flowered (rarely 2), peduncles nearly equal or shorter than pedicels. Pedicel 2–6 cm, glabrescent, bracts ovate-lanceolate, ciliate, base connate, apex acuminate. Calyx tubular, green, sometimes violet, narrowly campanulate, slightly inflated in fruit, 10–13 × 4–5 mm, glabrous; calyx teeth broadly triangular-ovate, apex obtuse or acute, margin membranous, ciliate. Carpophore 2–3 mm, basally pubescent. Petals ca. 15 mm, claws cuneate, glabrous, auricles obtuse; limbs yellowish green, narrowly obovate, margin entire or emarginate, coronal scales absent. Stamens and styles prominently exserted; filaments glabrous. Styles 3. Ovary ovate-elliptic, ca. 5 mm long. Capsule 6-valved. Flowering and fruiting in June–September.

Silene langshanensis is placed in the section *Holopetalae*, of which five species are now known from China (Zhou *et al.* 2001). It resembles *S. scabrifolia* (*see* Shishkin 1936: 677) but can be distinguished by several characters (Table 1). The following key should be useful when identifying species of sect. *Holopetalae* in China.

38 Zhao et al. • ANN. BOT. FENNICI Vol. 53

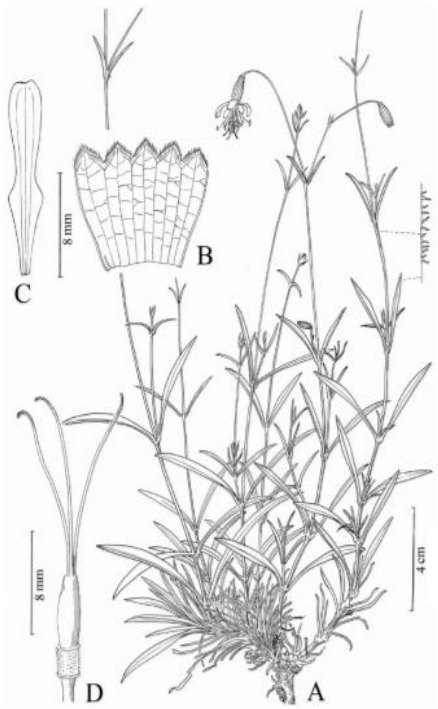


Fig. 1. *Silene langshanensis* (from the holotype, drawn by Ping Ma). — **A**: Habit. — **B**: Calyx. — **C**: Petal. — **D**: Pistil and carpophore.

Effective publication in the journal, available to Botanist

Date of valid publication (principles of priority): If the same species will be published by some one else after this date then the publication will be not valid. (/Principles of Priority).

Abstract / Summary / Synopsis.

Previously it was required to write in Latin.

Botanical name in Latin

Rank indicated

Type Specimen indicated

Specimens examined

Taxonomic Description

Ann. Bot. Fennici 53: 37–39
Helsinki 4 January 2016
ISSN 0003-3847 (print) ISSN 1797-2442 (online)

Silene langshanensis (Caryophyllaceae), a new species from Inner Mongolia, China

Li-Qing Zhao[1,*], Zhi-Ming Xin[2] & Yi-Zhi Zhao[1]

[1] *College of Life Science, Inner Mongolia University, Hohhot 010021, China (*corresponding author's e-mail: zhaotieniu@126.com)*
[2] *Experimental Center for Desert Forestry, Chinese Academy of Forestry, Dengkou, Inner Mongolia 015200, China*

Received 22 Apr. 2015, final version received 9 Oct. 2015, accepted 9 Oct. 2015

Zhao L.Q., Xin Z.M. & Zhao Y.Z. 2016: *Silene langshanensis* (Caryophyllaceae), a new species from Inner Mongolia, China. — *Ann. Bot. Fennici* 53: 37–39.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin *sp. nova* (Caryophyllaceae), is described and illustrated from Inner Mongolia, China. It appears to be most closely related to *S. scabrifolia* of *Silene* sect. *Holopetalae*. *Silene langshanensis* can be distinguished by the basally pubescent carpophore, petals with obtuse auricles, stems and leaves with dense, short hairs, and by the glabrous calyx.

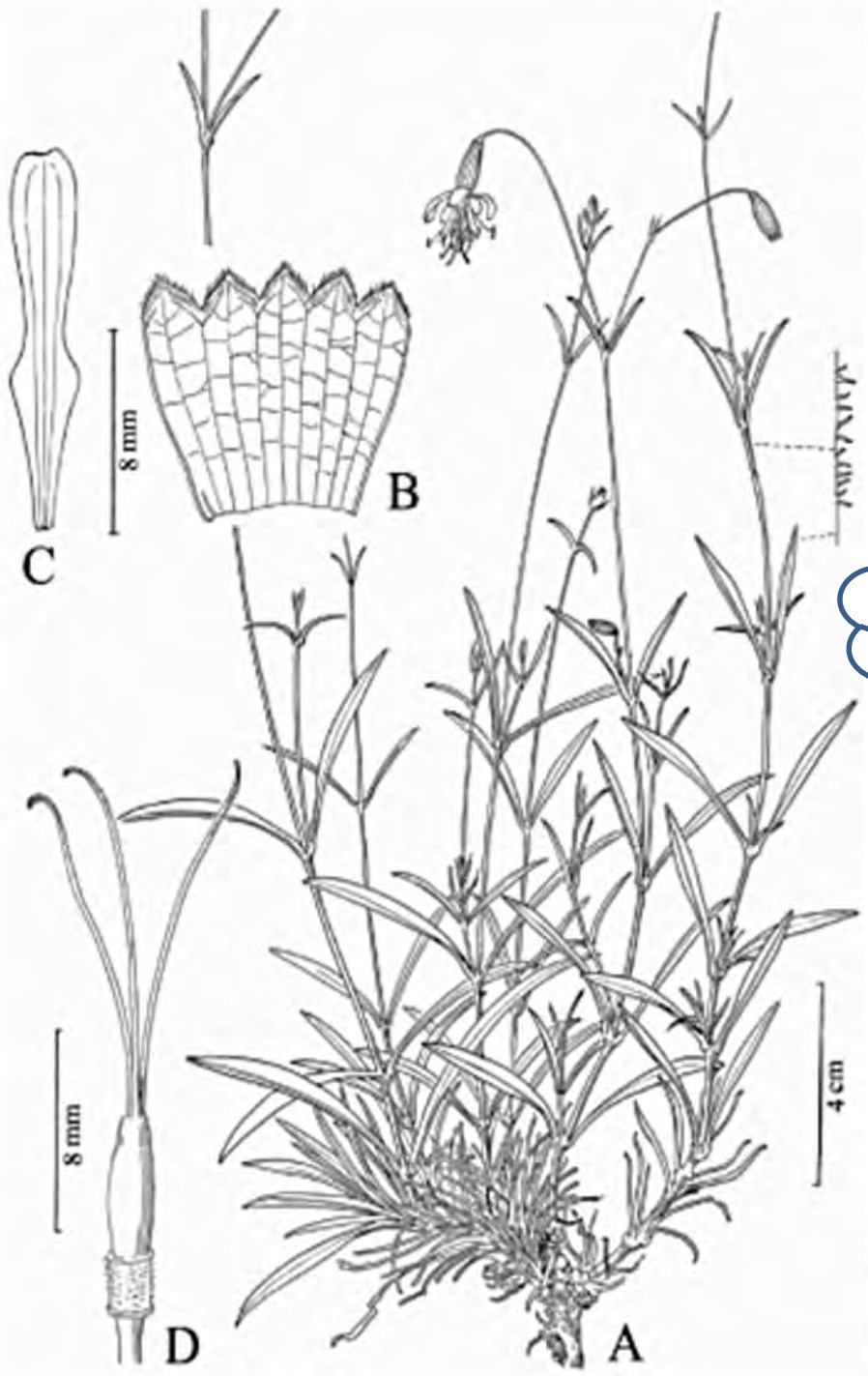
In total, there are about 600 species of *Silene s. lato* (Caryophyllaceae) (Zhou *et al.* 2001). They are distributed mainly in the northern temperate regions, but occur also in Africa and South America (Zhou *et al.* 2001). Among these species, 110 are known from China, of which 67 are endemic. Twenty of the endemics (nine species of *Silene s. stricto*, nine of *Melandrium*, one of *Cucubalus* and one of *Lychnis*) are found in Inner Mongolia.

In September 2008 and later, in 2014, the authors Zhao and Xin collected specimens of *Silene* from Langshan in Bayannaoer (Inner Mongolia) from desert steppe communities on mountain slopes at 1150–1400 m a.s.l. After careful study, we concluded that the specimens represented an undescribed species of *Silene*.

Silene langshanensis L.Q. Zhao, Y.Z. Zhao & Z.M. Xin, *sp. nova* (Fig. 1)

HOLOTYPE: China. Inner Mongolia, Bayannaoer, Dengkou, Mt. Langshan, 40°43′58.4″N, 106°22′28.5″E, on stony mountain slopes, 1371 m a.s.l., 9 June 2014 *Li-Qing Zhao, Zhi-Ming Xin, Shuai Qin & Long Chen N14-001* (HIMC). — PARATYPES (all HIMC!): Same location as holotype, 9 June 2014 *Li-Qing Zhao, Zhi-Ming Xin, Shuai Qin & Long Chen N14-002, N14-003, N14-004*; Mt. Langshan, 40°39′27.7″N, 106°23′14.1″E, on stony mountain slopes, alt. 1185 m a.s.l., 10 September 2008 *Li-Qing Zhao & Zhi-Ming Xin N08-001*.

Herbs perennial, 20–50 cm tall. Roots robust, lignified. Plants cespitose with erect stems, multi-branched or sparsely branched, pubescent, upper parts glabrescent. Basal leaves oblanceolate, 20–60 × 2–6 mm, both surfaces pubescent, margins ciliate, base attenuate into a long petiole, apex acute; cauline leaves 3–7 pairs, linear-oblanceolate or linear-lanceolate, with short, axillary sterile branches or sometimes elongated flowering axillary branches. Flowers in a racemiform-like thyrse; cymes alternating (resulting from suppression of opposite cyme at same node) or opposite, 1-flowered (rarely 2), peduncles nearly equal or shorter than pedicels. Pedicel 2–6 cm, glabrescent; bracts ovate-lanceolate, ciliate, base connate, apex acuminate. Calyx tubular, green,



Line drawing

Fig. 1. *Silene langshanensis* (from the holotype, drawn by Ping Ma). — **A**: Habit. — **B**: Calyx. — **C**: Petal. — **D**: Pistil and carpophore.

sometimes violet, narrowly campanulate, slightly inflated in fruit, 10–13 × 4–5 mm, glabrous; calyx teeth broadly triangular-ovate, apex obtuse or acute, margin membranous, ciliate. Carpophore 2–3 mm, basally pubescent. Petals ca. 15 mm, claws cuneate, glabrous, auricles obtuse; limbs yellowish green, narrowly obovate, margin entire or emarginate, coronal scales absent. Stamens and styles prominently exserted; filaments glabrous. Styles 3. Ovary ovate-elliptic, ca. 5 mm long. Capsule 6-valved. Flowering and fruiting in June–September.

Silene langshanensis is placed in the section *Holopetalae*, of which five species are now known from China (Zhou *et al.* 2001). It resembles *S. scabrifolia* (*see* Shishkin 1936: 677) but can be distinguished by several characters (Table 1). The following key should be useful when identifying species of sect. *Holopetalae* in China.

Table 1. Main morphological differences between *Silene langshanensis* and *S. scabrifolia*.

Character	*S. langshanensis*	*S. scabrifolia* (= *S. komarovii*)
Stem	densely pubescent, upper part glabrescent when flowering	pubescent in lower part, glabrous and viscid above
Basal leaves	oblanceolate, 20–60 × 2–6 mm	spatulate or lanceolate, 60–80 × 5–10 mm
Cyme	1-flowered (rarely 2)	multiflowered
Pedicel	20–60 mm long, glabrescent	5–10 mm long, sparsely pubescent
Calyx	narrowly campanulate, 10–13 × 4–5 mm, glabrous	tubular-clavate, 8–12 × 2–3 mm, glabrous or sparsely villous
Carpophore	shortly pubescent	glabrous
Petal	with obtuse auricles	without distinct auricles
Limbs	yellowish green	yellowish white

1. Leaves ovate-lanceolate, 15–30 mm wide ..... *S. kungessana*
1. Leaves lanceolate or linear, 1.5–10 mm wide 2
2. Leaves linear, 10–30 × 1.5–3 mm ..... *S. holopetala*
2. Leaves oblanceolate or lanceolate, 30–80 mm long, usually more than 4 mm wide 3
3. Stems usually not branched; calyx 6–9 mm; petals pinkish abaxially ..... *S. pseudotenuis*
3. Stems branched; calyx 8–13 mm; petals yellowish green or yellowish white 4
4. Stem pubescent in lower part, glabrous and viscid above; cymes multiflowered; petals yellowish white, without obvious auricles; carpophore glabrous ..... *S. scabrifolia*
4. Stem with dense short hairs, upper part glabrescent when flowering; cymes 1-flowered (rarely 2); petals yellowish green, with obtuse auricles; carpophore basally pubescent ..... *S. langshanensis*

Taxonomic Key for Identification

Acknowledgements

We are grateful to Ping Ma for the drawing. This study was financially supported by Natural Science Foundation of Inner Mongolia Autonomous Region (2014ZD02), National Key Basic Research Program of China (2014CB138802) and the Central Public-interest Scientific Institution Basal Research Fund (CAFYBB2014MA016).

References

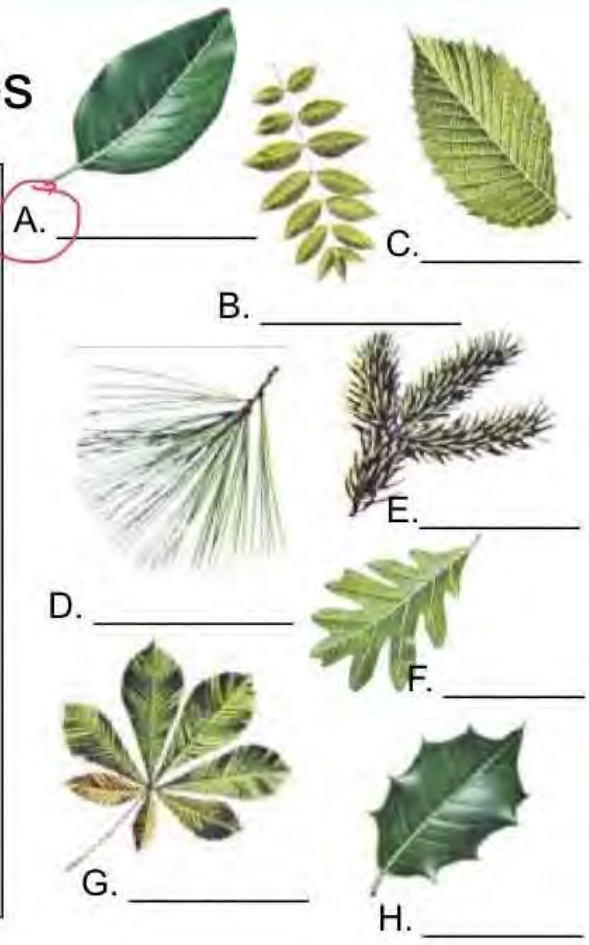
Shishkin B.K. [Шишкин Б.К.] (ed.) 1936: [*Flora USSR*], vol. 6. — Izdatel'stvo Akademii Nauk USSR, Leningrad. [In Russian].

Zhou L.H., Wu Z.Y., Lidén M. & Oxelman B. 2001: *Silene*. — In: Wu Z.Y. & Raven P.H. (eds.), *Flora of China*, vol. 6: 66–100. Science Press, Beijing & Missouri Botanical Garden Press, Saint Louis.

Taxonomic Key: **An identification device, consisting of contrasting statements used to narrow down the identity of a taxon**

Dichotomous Key For Leaves

1.	a. Needle leaves	go to 2
	b. Non-needle leaves	go to 3
2.	a. Needles are clustered	Pine
	b. Needles are in singlets	Spruce
3.	a. Simple leaves (single leaf)	go to 4
	b. Compound leaves (made of "leaflets")	go to 7
4.	a. Smooth edged	go to 5
	b. Jagged edge	go to 6
5.	a. Leaf edge is smooth	Magnolia
	b. Leaf edge is lobed	White Oak
6.	a. Leaf edge is small and tooth-like	Elm
	b. Leaf edge is large and thorny	Holly
7.	a. Leaflets attached at one single point	Chestnut
	b. Leaflets attached at multiple points	Walnut



Dichotomous Key

1a. It has fur.....**It belongs to the class Mammalia.**

1b. It does not have fur.....**Go to number 2.**

2a. It cannot change locations (move) on its own.....**Go to number 3.**

2b. It can change locations on its own.....**Go to number 4.**

3a. It has a flower.....**It belongs to the phylum Phanerogams (flowering plants).**

3b. It does not have a flower.....**It belongs to the phylum Cryptogams (non-flowering plants).**

4a. It has feathers and can fly.....**It belongs to the class Aves.**

4b. It cannot fly.....**Go to number 5.**

5a. It does not have a backbone.....**It belongs to the phylum Cnidaria.**

5b. It swims in water.....**It belongs to the phylum Chondrichthyes.**

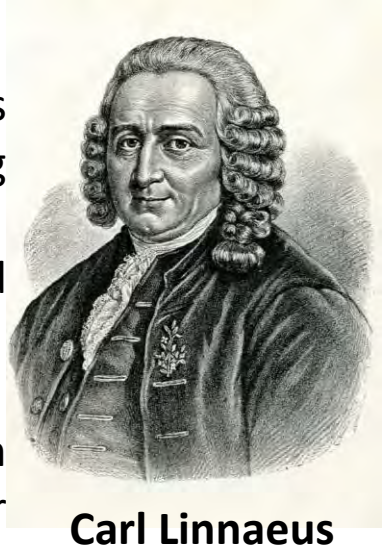
جامعة الملك سعود
King Saud University

Advance Experimental Taxonomy

Chapter 2- Taxonomic circumscription

Chapter 2- Taxonomic circumscription

Hierarchical Classification of Vascular Plants



Carl Linnaeus

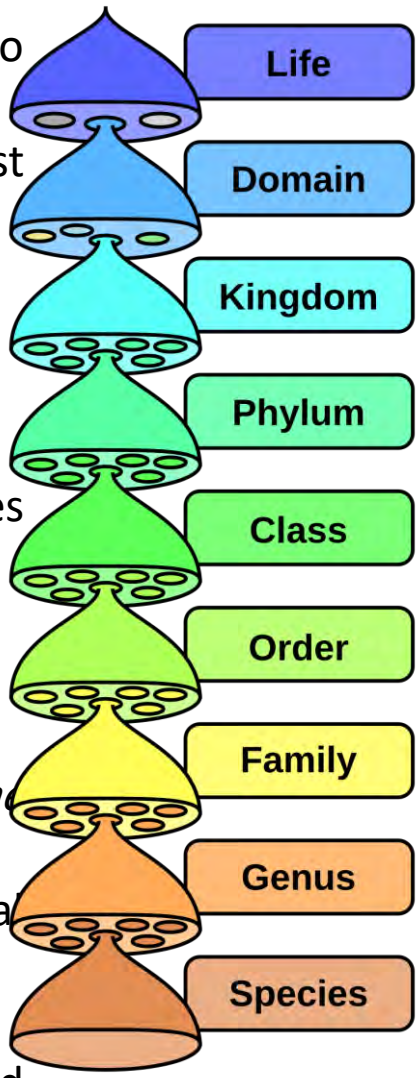
•**Need for Classification**:
- Over 250,000 species of vascular plants necessitate a systematic grouping mechanism.
- Enables structured study and documentation.

•**Classification Process**:
- Organisms are grouped based on resemblance (morphology, phylogeny, or phenetic relationships).
- Groups are progressively assembled into larger, inclusive groups.
- Final arrangement creates a single most inclusive group.

•**Taxonomic Hierarchy**:
- Groups (Taxa) arranged by inclusiveness:
  - **Bottom**: Least inclusive (e.g., species).
  - **Top**: Most inclusive (e.g., division).
- Fixed sequence of categories ensures consistency.

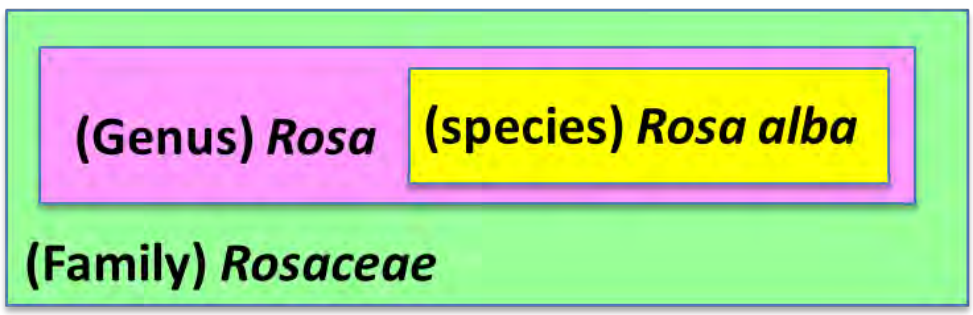
•**Naming Conventions**:
- Taxonomic groups named by category:
  - Example:
    - *Rosales* (Order), *Rosaceae* (Family).
- Names provide insight into hierarchical placement.

•**Outcome**:
- A complete taxonomic structure achieved through hierarchical classification.

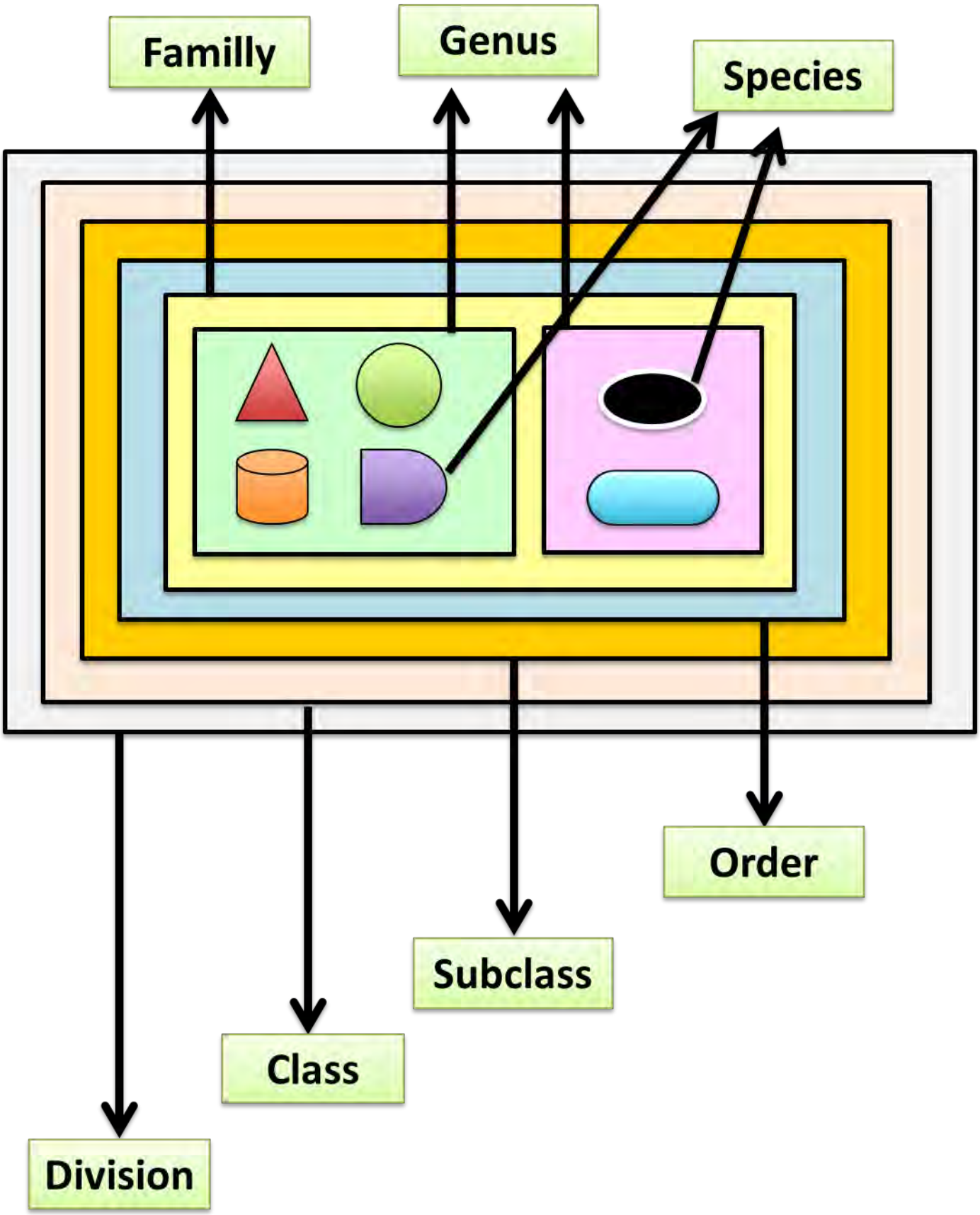


Taxonomic Groups, Categories, and Ranks

- **Key Concepts**:
  - **Taxonomic Categories**:
    - Abstract and artificial, like empty shelves in an almirah.
    - Gain meaning only when groups (taxa) are assigned to them.
    - Fixed positions in the hierarchy but do not directly represent nature.
  - **Taxonomic Ranks**:
    - Grammatical sense of a category's placement (e.g., genus → generic rank).
    - Act as partitions separating one category from another.
  - **Taxonomic Groups (Taxa)**:
    - Represent real, discrete biological entities in nature.
    - Objective and non-arbitrary.
    - Example: *Rosa alba* (species), *Rosa* (genus), *Rosaceae* (family).
- **Significance**:
  - Assigning groups to categories integrates natural entities into the hierarchical system.
  - Naming conventions (e.g., "-aceae" for families) establish clear hierarchical positions.



Concept of categories, groups and taxonomic structure illustrated in the form of a box-in-box figure



Dendrogram method for depicting the hierarchical system based on same hypothetical example

A dendrogram is a tree-like diagram used to illustrate the relationships between various taxonomic groups in a hierarchical classification system.

•**Key Points**:

1.Visual Representation:

1. Shows branching patterns to depict the hierarchy of groups and their relationships. Each branch point (node) represents a common ancestor or shared characteristics.

2.Levels of Hierarchy:

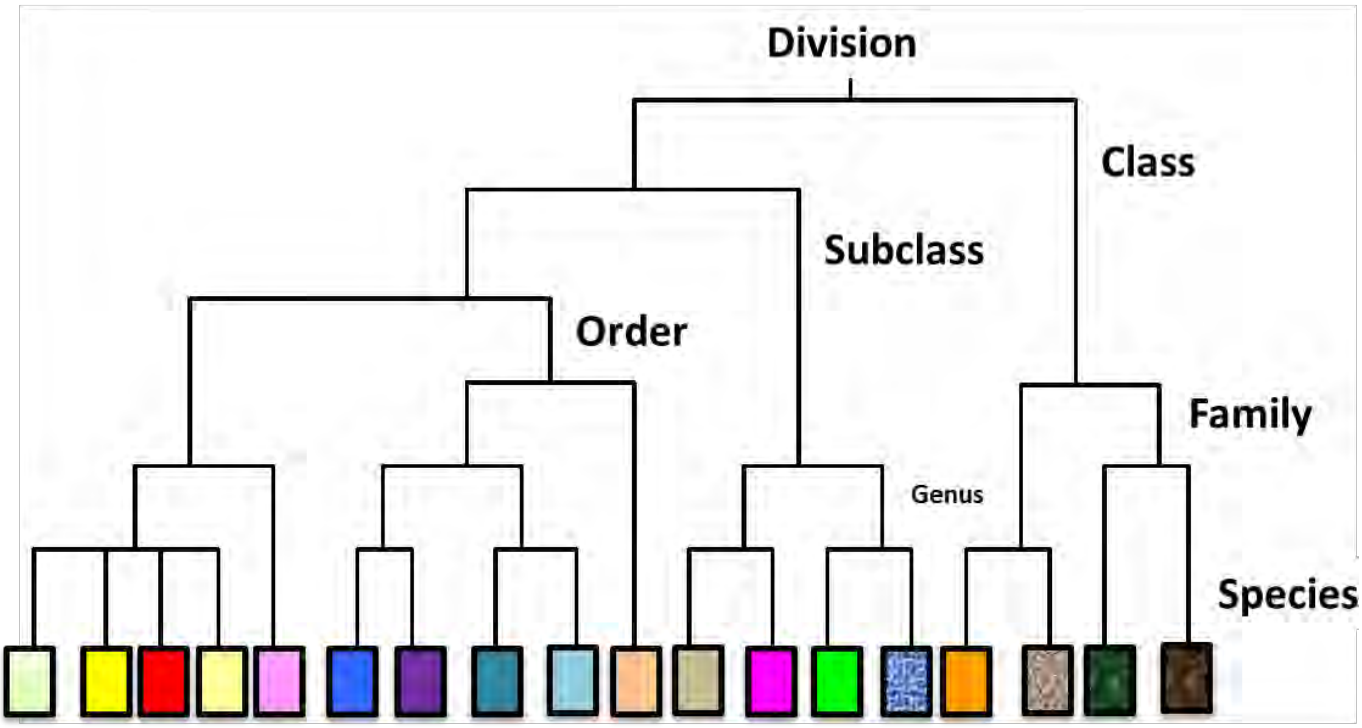
1. Branches split into progressively smaller groups, moving from the root (most inclusive category) to the tips (least inclusive groups like species).

3.Inclusive Groups:

1. Higher-level groups (e.g., families, orders) encompass multiple lower-level taxa (e.g., genera, species).

4.Shared Characteristics:

1. Proximity on the dendrogram indicates the degree of similarity or shared traits among taxa. Closely related taxa are connected by shorter branch lengths.



Key Characteristics of the Hierarchical Classification System

1.Inclusiveness of Categories:

1. Higher categories include more groups; lower categories are less inclusive.

2.Classification of Plants:

1. Plants are classified into groups, not categories.
2. Example: *Poa annua* → Member of *Poa* (genus), *Poaceae* (family), etc., but not a direct member of any category.

3.Taxa Relationships:

1. A taxon can belong to multiple taxa (e.g., *Urtica dioica* → member of *Urtica*, *Urticaceae*, etc.).
2. A taxon belongs to only one category (e.g., species).

4.Non-Hierarchical Composition of Categories:

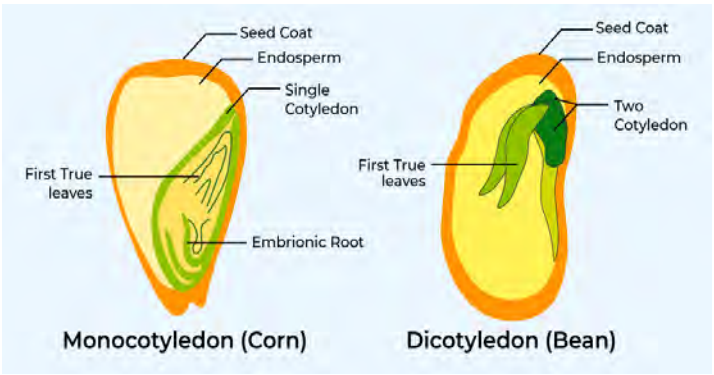
1. Categories do not comprise lower categories (e.g., family ≠ sum of genera).

5.Character Sharing in Hierarchy:

1. Lower taxon characters define higher taxa.
2. Example: Shared species traits define a genus; shared genera traits define a family.

6.Character Variation in Higher Taxa:

1. Higher taxa share fewer common characters.
2. Example: **Dicots**: Two cotyledons, pentamerous flowers, reticulate venation. **Monocots**: One cotyledon, trimerous flowers, parallel venation.
3. Exceptions exist: *Smilax* (monocot with reticulate venation) and *Plantago* (dicot with parallel venation).



trimerous

pentamerous

Utilization of Categories and Species Concept

Utilization of Categories:

- Categories have relative value and no meaning when empty.
- Assigning taxa to appropriate categories is key for meaningful classification.
- Proper utilization enhances the hierarchical system's accuracy.

Species Concept:

Darwin's Observation:

Species are widely understood but remain controversial to define.

Diversity of definitions stems from species being a human-constructed concept.

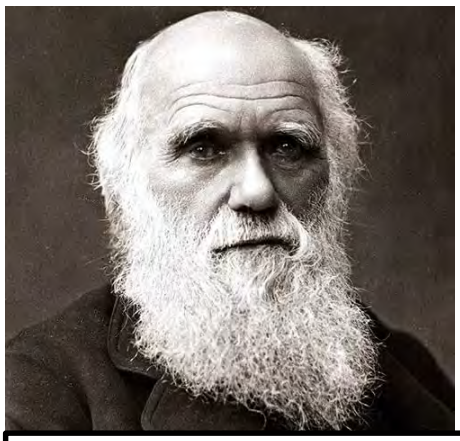
ICN Definition:

Species are classificatory units defined by trained biologists using available data.

Dual Meaning:

Naturally occurring groups of organisms (basic unit of evolution).

A taxonomic category governed by nomenclature rules.



Charles Darwin
12 February 1809- 19 April 1882

Species as the Basic Unit of Taxonomy

1.Fundamental Unit:

1. Species is the foundational level of classification, often lacking infraspecific names (e.g., families like Apiaceae and Liliaceae).

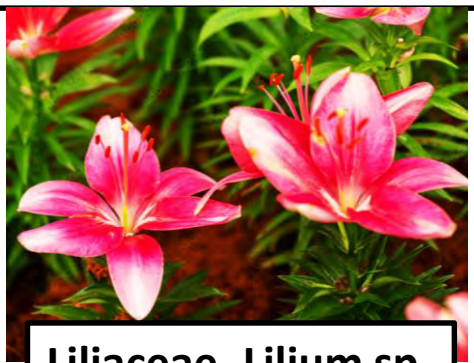
2.Directly Identifiable:

1. Species can be described and recognized independently of other taxonomic ranks.
2. Example: Herbarium sheets can be sorted into species without knowing the genera.

3.Non-Arbitrary Nature:

1. **Inclusion**: Members are continuous by appropriate criteria (e.g., morphology or gene flow).
2. **Exclusion**: Species are distinct from other groups by clear discontinuity.
3. Higher taxa are **non-arbitrary to exclusion** but **arbitrary to inclusion**, as they may include internally discontinuous species.

Apiaceae- Apium graveolens,



Liliaceae- Lilium sp.

Morphological variation in Ipomoea Species

Ideal Species: Characteristics and Reality

Definition:

A species with clear identity, no taxonomic issues, and distinct characteristics.

- **Expected Characteristics**:
  - Morphologically distinct and easily recognizable.
  - No internal variation (no subspecies, varieties, or formas).
  - Genetically isolated from other species.
  - Reproduces sexually.
  - At least partially outbreeding.
- **Reality**:
  - Ideal species are rare in the plant kingdom.
  - Most species deviate from one or more of these criteria.
- **Examples**:
  - Common ideal species: Families like Apiaceae, Asteraceae, and genera *Allium* and *Sedum*.

***Allium* and *Sedum*.**

Transmutation: An Ancient Greek Idea in Botany

- **Ancient Greek Concept**: The belief that plants like wheat, barley, and others could change into different species under certain conditions.
- **Prominent Botanists**: Botanists like Bobart supported the idea, claiming that plants like Crocus and Gladiolus could transform over time.
- **Robert Sharrock's Investigation (1660)**: He refuted the idea after investigation, explaining it as a result of unintentional mixing of seeds or propagules.
- **Modern Understanding**: Transmutation is now rejected as a fallacy, explained by accidental cross-pollination or mixing of seeds, not a biological transformation.
- **Case Study – Kashmir**: The author observed a mix-up between saffron and Iris reticulata plants, illustrating the misidentification and unintentional introduction of non-native species.

Saffron

Crocus and Gladiolus

Iris reticulata

Nominalistic Species Concept

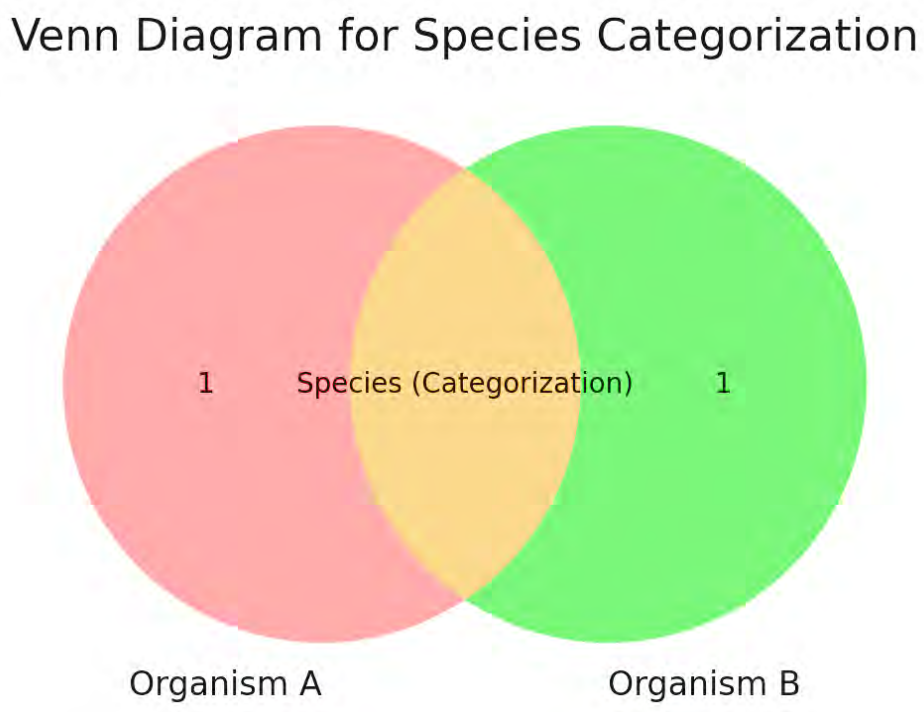
•**Overview**: Focuses on categorizing organisms within the framework of formal relations, not on their intrinsic properties.

•**Taxonomic Hierarchy**: Species are seen as categories within the broader classification system.

•**Binomial Nomenclature**: Species are assigned specific names in the binomial system, often used for formal identification.

•**Logically Sound, Scientifically Irrelevant**: While it makes sense logically, it does not serve much scientific purpose because the goal is merely to place individuals into species categories, regardless of their actual biological relationships.

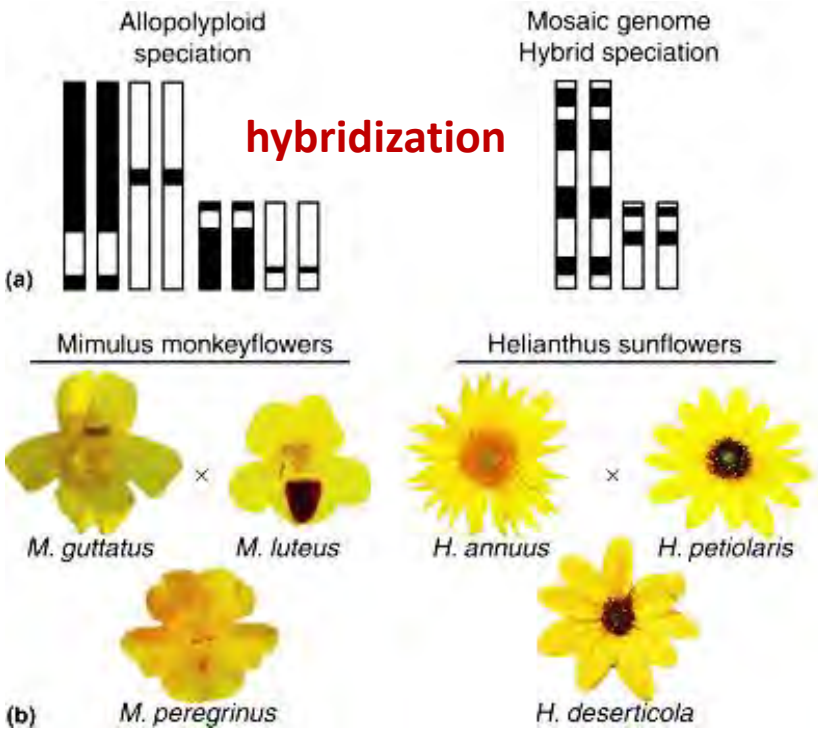
•**Academic Interest**: This concept is largely of academic significance today and no longer serves as a primary method for understanding species.



The overlapping region shows how two organisms can be grouped into the same species category under the Nominalistic Species Concept, even though their individual characteristics might not be the defining factor.

Typological Species Concept

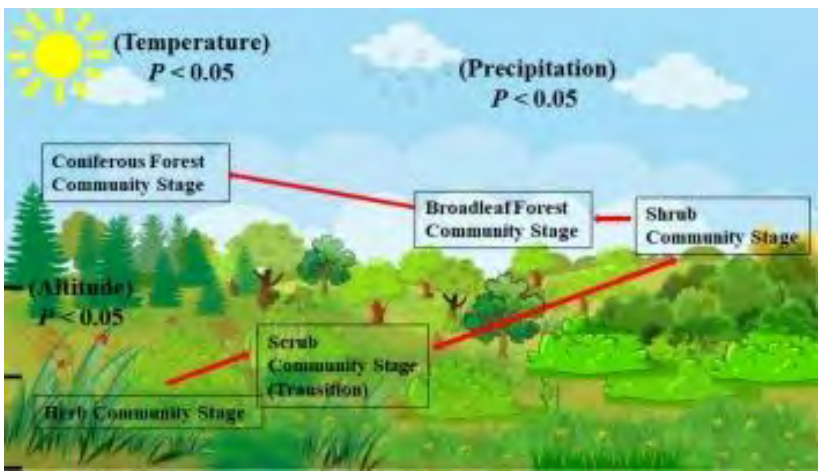
- **Origin**:
  - Proposed by John Ray (1686) and elaborated by Carl Linnaeus (*Critica Botanica*, 1737).
- **Key Ideas**:
  - **Fixity of Species**:
    - Species are fixed and created by the Almighty Creator.
    - Variations exist within species, but species themselves are immutable.
  - **True Breeding**:
    - Species breed true within their natural variation.
- **Linnaeus' Shift in Belief**:
  - Later recognized that species can arise through hybridization (*Fundamenta Fructificationis*, 1762).
  - Suggested original genera diversified over time into species and varieties.
- **Clarification**:
  - Typological concept differs from typification, a nomenclatural methodology for naming taxonomic groups.



Taxonomic Species Concept

- **Introduction**:
  - Developed from Lamarck's (1809) and Darwin's (1859) ideas.
  - Also known as the **Morphological Species Concept**.
- **Key Principles**:
  - Species are groups of individuals sharing common morphological features.
  - Distinct discontinuity in features separates one species from another.
  - Recognizes **continuous and discontinuous variation** within and between species.
- **Modification by Du Rietz (1930)**:
  - Introduced the **Morpho-Geographical Species Concept**:
    - Species are the smallest populations permanently separated by distinct discontinuities in biotypes.
    - Geographic distribution of populations is emphasized.

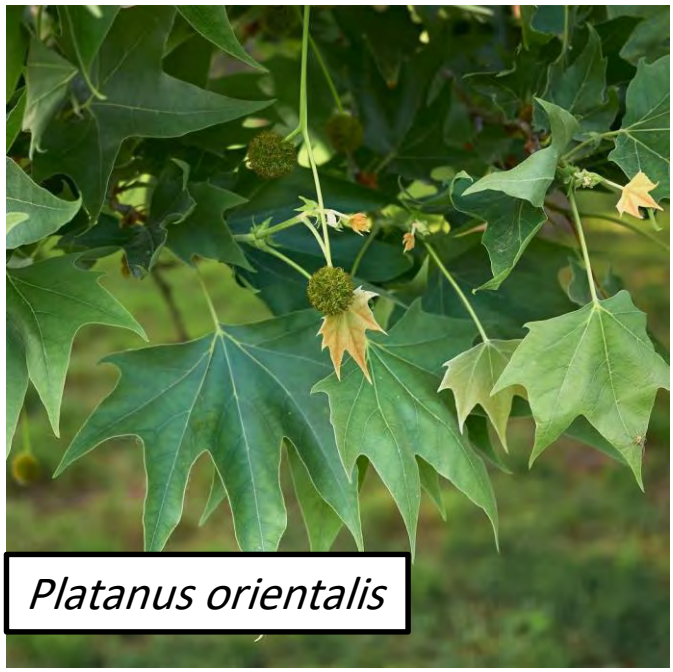
Asteraceae



Geographic distribution of species

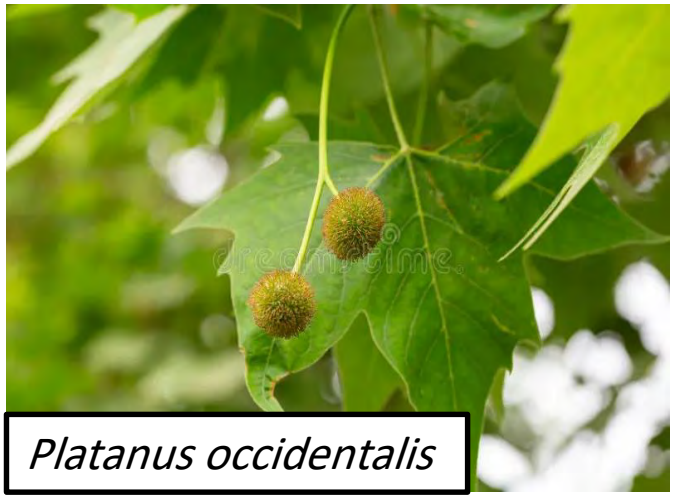
•**Vicariance in Species**:

- Species pairs morphologically distinct and adapted to specific climates, but capable of interbreeding in shared environments.
- Examples:
  - *Platanus orientalis* (Mediterranean) and *P. occidentalis* (Eastern USA).
  - *Catalpa ovata* (Japan/China) and *C. bignonioides* (America).
- Known as **vicarious species** or **vicariants**; the phenomenon is **vicariance**.



Platanus orientalis

Catalpa ovata



Platanus occidentalis

Catalpa bignonioides

Advantages and Drawbacks of the Taxonomic Species Concept

Advantages:

1.Effective for **field and herbarium identification** of plants.

2.Widely applied in taxonomy, aiding in species recognition.

3.Utilizes **easily observable morphological and geographical features**.

4.Even experimental taxonomists often **apply it indirectly**.

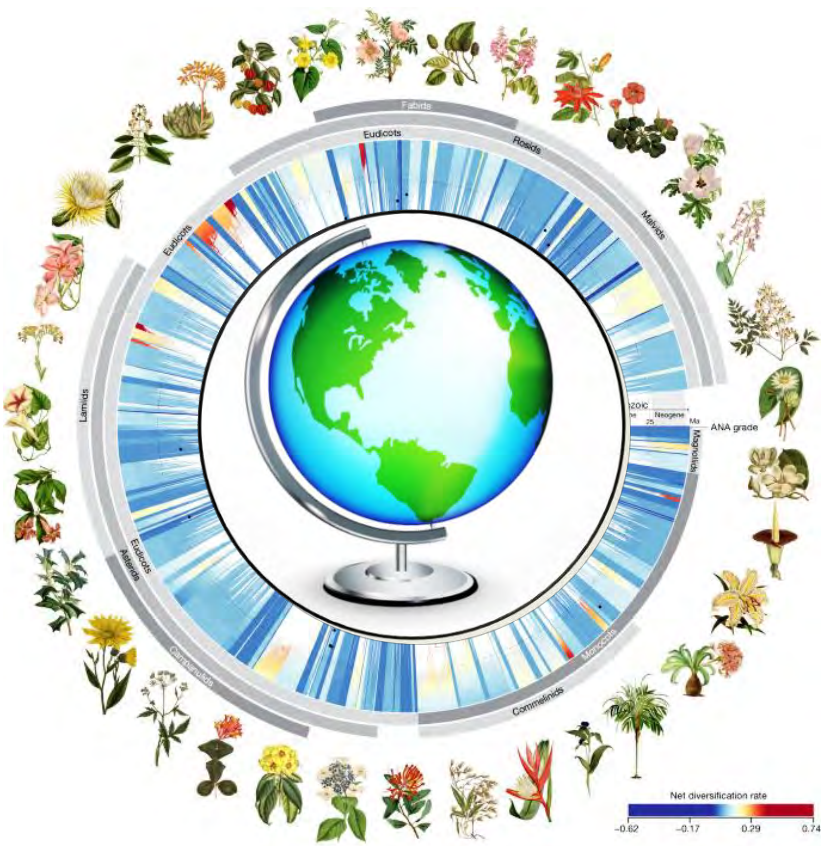
5.Many species recognized morphologically are later **experimentally confirmed**.

Drawbacks:

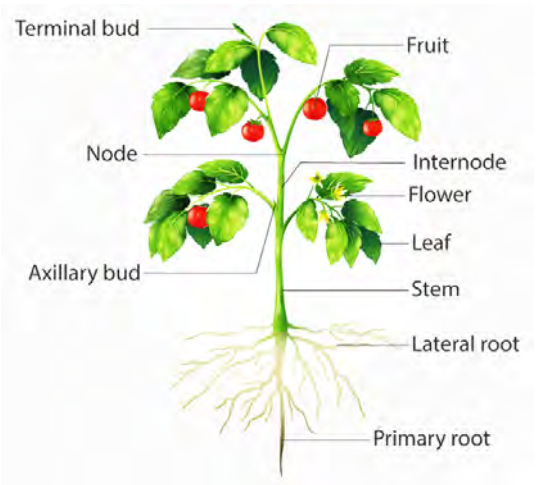
1.Subjective: Different taxonomists use varied sets of characters.

2.Requires expertise: Reliable characters are determined only with experience.

3.Neglects genetic relationships between plants.



Phylogenomics and the rise of the angiosperms



Biological Species Concept

Definition:

•Proposed by **Ernst Mayr (1942)**:

- **Species**: Groups of natural populations that are reproductively isolated from other such groups.
- Refined by Grant (1957): Species as a **community of cross-fertilizing individuals** linked by mating bonds and isolated by barriers to reproduction.
- Valentine & Love (1958): Defined species in terms of **gene exchange**.

Key Principles:

1.Populations within the same species interbreed.

2.Reproductive isolation distinguishes species.

3.Conspecific populations: Freely exchange genes.

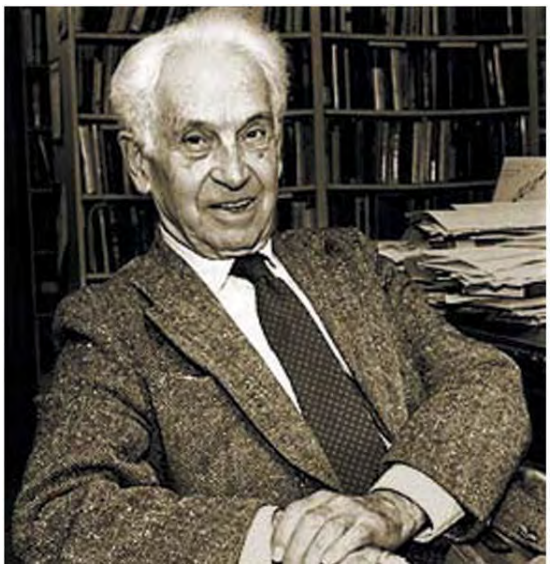
4.Distinct species: Reproductively isolated.

Advantages:

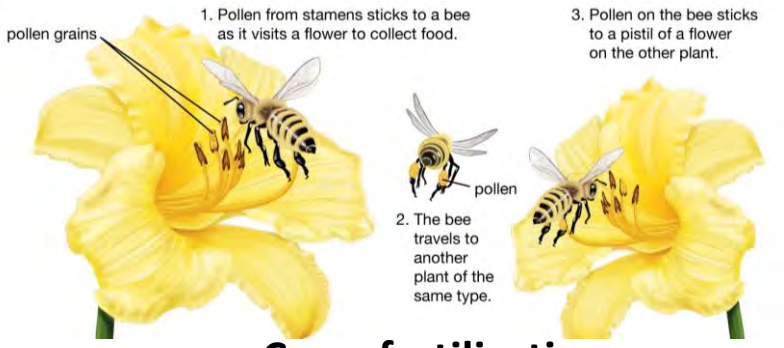
1.Objective: Applies the same criteria across plant groups.

2.Scientific basis: Maintains morphological differences due to reproductive isolation.

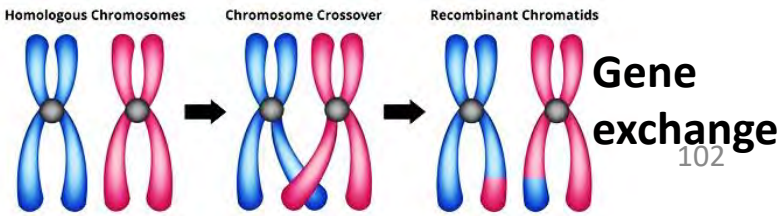
3.Practical: Based on feature analysis, not reliant on taxonomic experience.



Ernst Mayr: 1904 – 2005



Cross fertilization



Gene exchange

Challenges in Applying Biological Species Concept to Plants

Limitations:

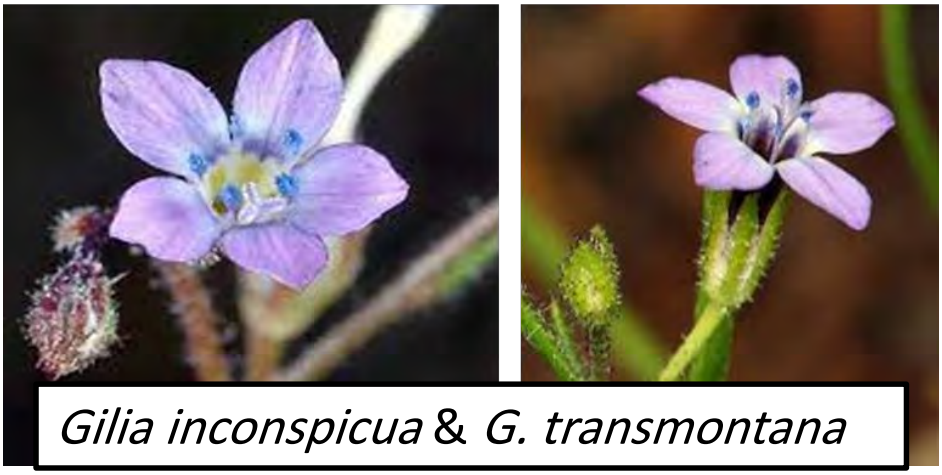
1.Vegetative Reproduction Dominates: Most plants reproduce asexually, making reproductive isolation inapplicable.

2.Experimental Relevance: Isolation tests under cultivation may not reflect natural populations.

3.Morphology vs. Reproductive Barriers:

1. **Compilospecies**: Morphologically distinct but not reproductively isolated (e.g., *Salvia mellifera* & *S. apiana*).
2. **Sibling Species**: Reproductively isolated but morphologically similar (e.g., *Gilia inconspicua* & *G. transmontana*).

Salvia mellifera & *S. apiana*.



Gilia inconspicua & *G. transmontana*

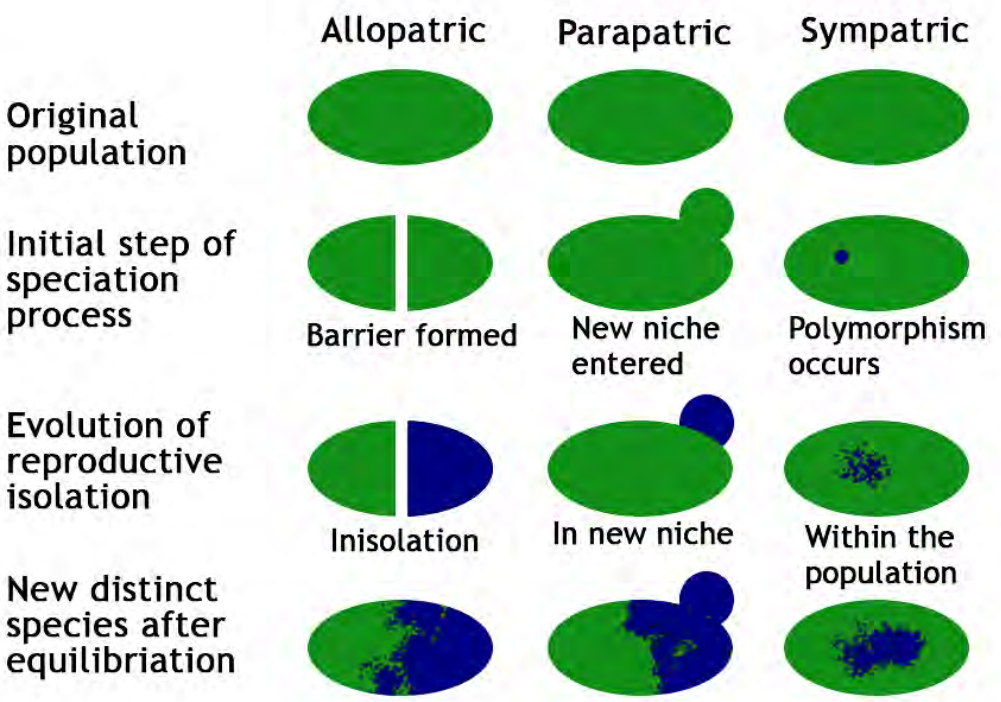
4. **Allopatric Populations**: Fertility-sterility tests lack significance in geographically separated populations.
5. **Practical Difficulties**: Conducting tests is time-consuming and data is limited for most species.
6. **Apomicts**: Reproductive barriers irrelevant in plants reproducing via apomixis.

Stebbins' Contribution (1950):

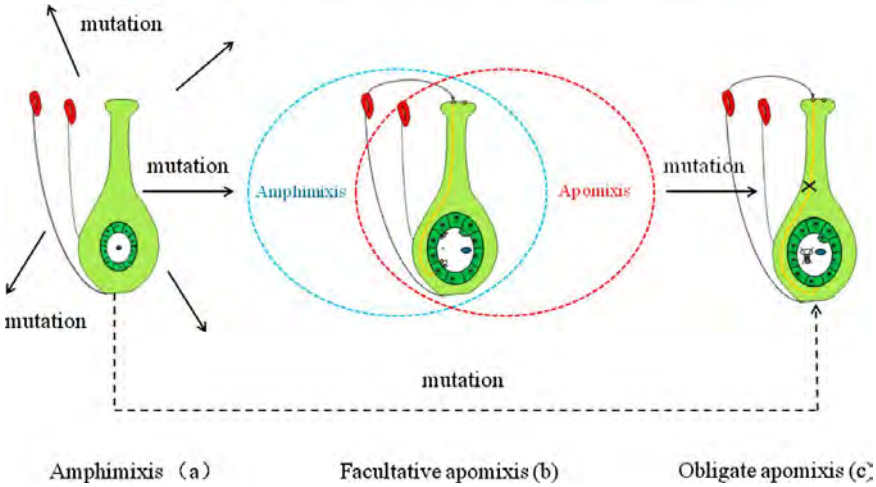
Species: Systems of populations with **genetically-based discontinuities** in variation patterns.

Isolation Mechanisms: Key for distinguishing species:

- **Sympatric species**: Occur in the same region.
- **Allopatric species**: Occur in different regions.



Apomicts



Evolutionary Species Concept

A **single lineage** of ancestor-descendant populations maintaining **identity** and possessing unique evolutionary tendencies and historical fate.

Key Features:

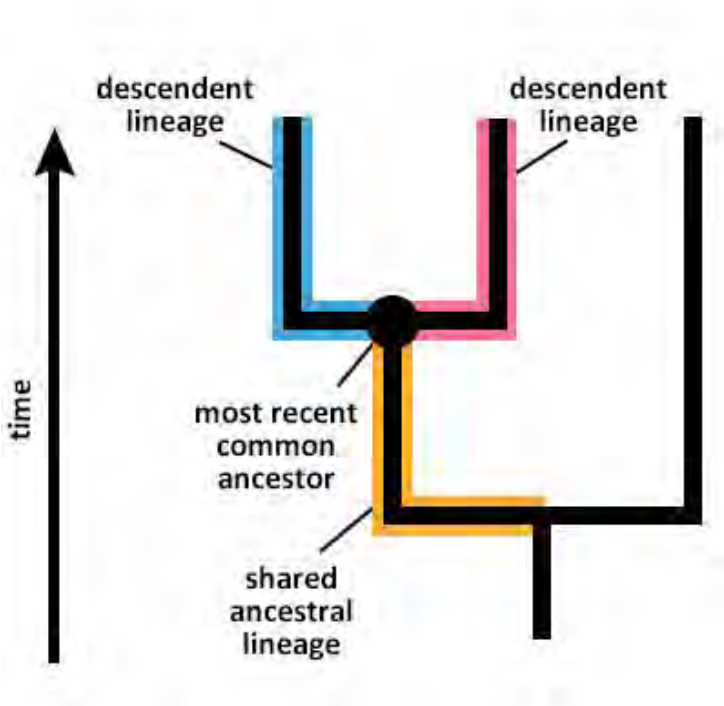
1.Broad Reproductive Scope: Compatible with both sexual and asexual species.

2.Species Identity:

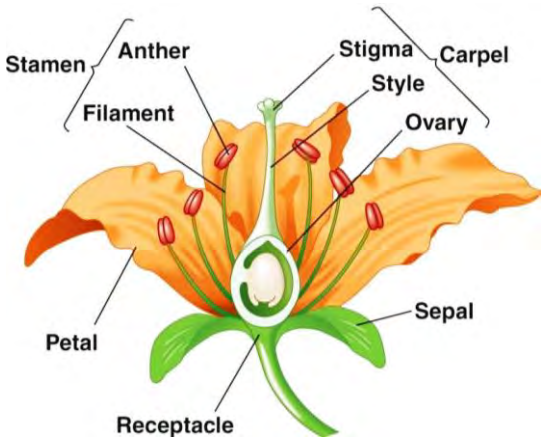
1. **Sexual Species**: Recognized through phenotypic, behavioral, and biochemical differences.
2. **Asexual Species**: Maintained through phenotypic, genotypic differences, or ecological roles.

3.Hybridization Consideration:

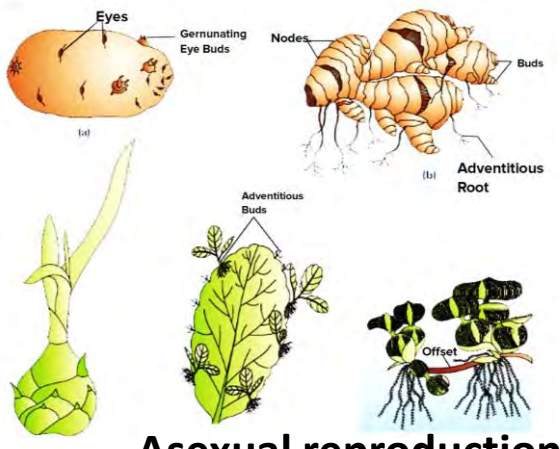
1. Hybridization does not negate species distinction if ecological and evolutionary roles remain separate.



Lineages are sequences of biological entities connected by ancestry-descent relationships



Sexual reproduction



Asexual reproduction

Microspecies:

•Populations with slight morphological differences, often restricted geographically.

•Types: **Clonal**: *Phragmites*

- **Agamospermous**: *Rubus*
- **Heterogamic**: *Oenothera biennis*
- **Autogamous**: *Erophila*

Cryptic Species: Morphologically similar but cytologically or physiologically distinct.

Advantages:

•Accounts for **lineage-based identity** and evolutionary trajectories.

•Applicable across diverse reproductive modes.

Microspecies vs. Cryptic Species:

•**Microspecies**: Differentiated by minor traits, often unstable over time.

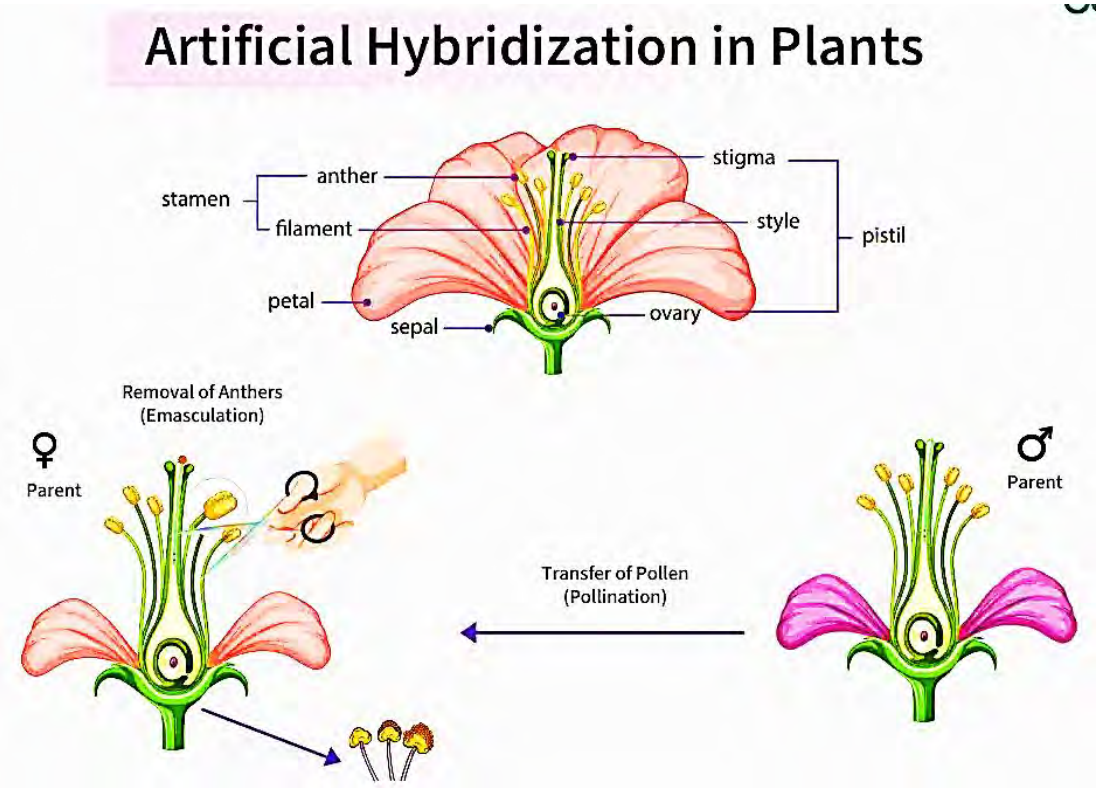
•**Cryptic Species**: Similar morphology, distinct cytology/physiology.

Biosystematic Species Concept

The concept of **biosystematic species** was introduced by Grant (1981), referring to species categories based on fertility relationships that are determined through artificial hybridization experiments. According to this concept, species can be classified into different categories based on their ability to interbreed and produce fertile offspring.

Related Terms:

- **Ecotype:** A group of organisms within a species, adapted to specific habitats. Can interbreed freely without loss of fertility or vigour.
- **Ecospecies:** Ecotypes that interbreed freely form an **ecospecies** (taxonomically equivalent to a species).
- **Coenospecies:** Related ecospecies with limited genetic exchange. Equivalent to a **subgenus**.
- **Comparium:** Related coenospecies capable of hybridization, directly or via intermediates. Equivalent to a **genus**.
- **Genera:** Groups with complete sterility barriers between species.



Infraspecific ranks

Basic Unit of Classification

•**Species**: Regarded as the basic unit of classification.

•Some works (e.g., *Flora of USSR*) do **not recognize infraspecific taxa**, while others (European, American, Asian Floras) do.

Infraspecific Ranks (Recognized by the International Code of Botanical Nomenclature)

1.Subspecies

1. Defined by Du Rietz (1930) as:
    1. A population of multiple biotypes.
    2. Forms distinct **regional facies** of a species due to geographical isolation.
2. Morphologically distinct but **interfertile populations**.

2.Variety (Latin: *Varietas*)

1. Defined by Du Rietz (1930) as:
    1. A population of multiple biotypes forming a **local facies**.
    2. Restricted to **smaller geographical areas**.
2. Often used for:
    1. Variations with unclear nature in early taxonomy.
    2. Variations within a **subspecies**.

3.Forma (Latin: *Forma*)

1. Defined as:
    1. **Sporadic variants** with minor taxonomic significance.
    2. Distinguished by one or a few linked characters.

Concept of Genus in Taxonomy

Definition of Genus

•A group of closely-related species.

•Represents a **phylogenetic relationship** (Rollins, 1953).

•Mayr (1957):

- A genus includes one species or a **monophyletic group**.
- Separated by a **decided discontinuity gap**.

Criteria for Recognizing a Genus

1.Natural Grouping

1. Should be monophyletic.
2. Consider cytogenetic, geographic, and morphological information.

2.Multiple Character Consideration

1. Avoid relying on a **single character**.
2. Adaptive traits (e.g., *Ranunculus* → *Batrachium* based on aquatic adaptations).

3.Size Flexibility

1. **Monotypic Genus**: Single species (e.g., *Leitneria*, *Ginkgo*).
2. **Polytypic Genus**: Many species (e.g., *Euphorbia* - 2100 species, *Astragalus* - 2000 species).

4.Clear Gap Between Genera

1. If unclear, merge as **subgenera or sections**.
2. Consider family context, genus size, and traditional usage.

5.Range-Wide Study

1. Examine species across their full distribution range.
2. Stability in one region may not apply elsewhere.

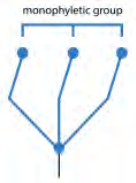
Adaptive traits (e.g., ***Ranunculus*** → ***Batrachium*** based on aquatic adaptations).

Irish Wildflowers - Water-crowfoots, Ranunculus subgenus Batrachium

Monotypic Genus: Single species (e.g., Leitneria, Ginkgo).
Polytypic Genus: Many species (e.g., Euphorbia - 2100 species, Astragalus - 2000 species).

Concept of Family in Taxonomy

A group of **closely-related genera**. Ancient concept recognized by both laymen and taxonomists.



Key Characteristics

1.Monophyletic Nature

1. Families ideally represent **monophyletic groups** (common evolutionary origin).

Umbels

2.Size Variability

1. Single-genus families:
    1. *Podophyllaceae*, *Hypecoaceae*.
2. Multi-genus families:
    1. *Asteraceae* (~1100 genera).

3.Stability in Classification

1. Broadly-conceived family concepts promote stability.
2. Examples:
    1. *Lamiaceae* and *Verbenaceae* remain distinct despite minimal discontinuity.
    2. *Rosaceae* is not split despite significant internal differences.

Notable Examples

•**Legumes** (*Fabaceae*), **Crucifers** (*Brassicaceae*), **Umbels** (*Apiaceae*), **Grasses** (*Poaceae*): Widely recognized for centuries.

Podophyllaceae

Legumes

Legumes

Grasses

Advance Experimental Taxonomy

Chapter 3- Tools and techniques for sequencing

The Cell, Nucleus, and DNA

The Cell: Basic unit of life, containing organelles that perform essential functions.

The Nucleus: Enclosed in a double membrane, contains **DNA** as **chromatin/chromosomes**.

Chromosomes & Genes:

Chromosomes: Thread-like structures carrying genetic information.

Genes: Units of heredity made of **DNA**, located on chromosomes.

DNA Structure:

Double-helix model by **Watson & Crick (1962)**.

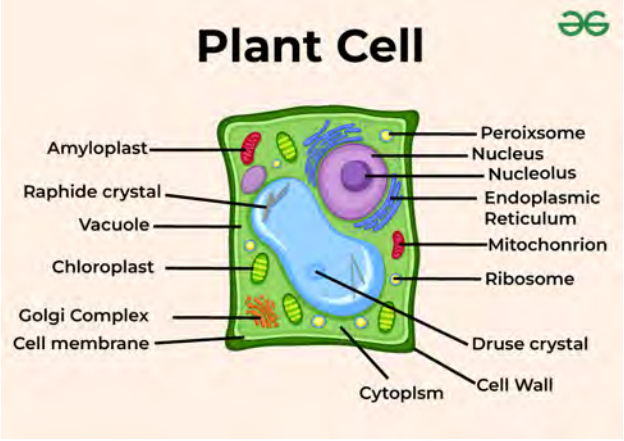
Composed of **four bases** (A, T, C, G) paired as **A-T & C-G**.

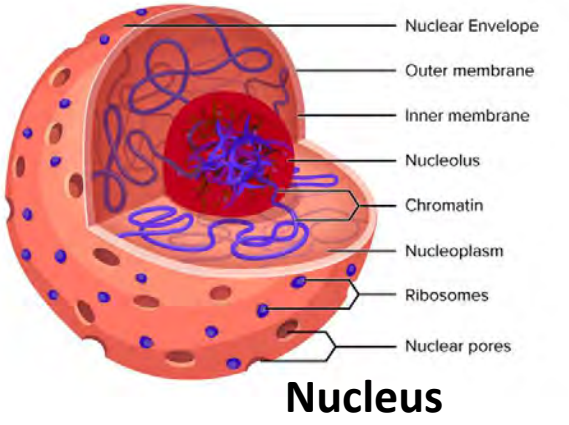
Made of **nucleotides** (Base + Sugar + Phosphate).

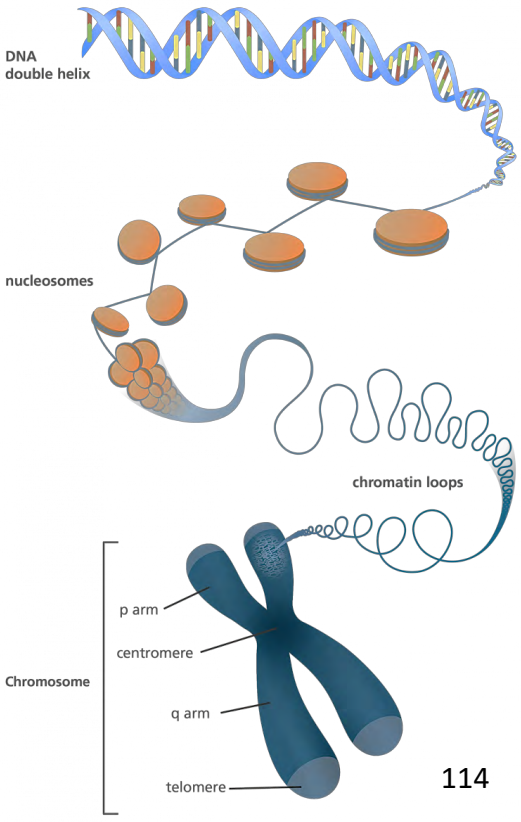
Types of DNA:

Nuclear DNA (in nucleus).

Mitochondrial DNA (mtDNA) & **Chloroplast DNA (cpDNA)** (in organelles).







DNA Replication & Gene Expression

DNA Replication:

Each strand acts as a template to form new DNA copies.

Gene Expression & Protein Synthesis:

Transcription: DNA → RNA.

Translation: RNA → Amino Acids → Proteins.

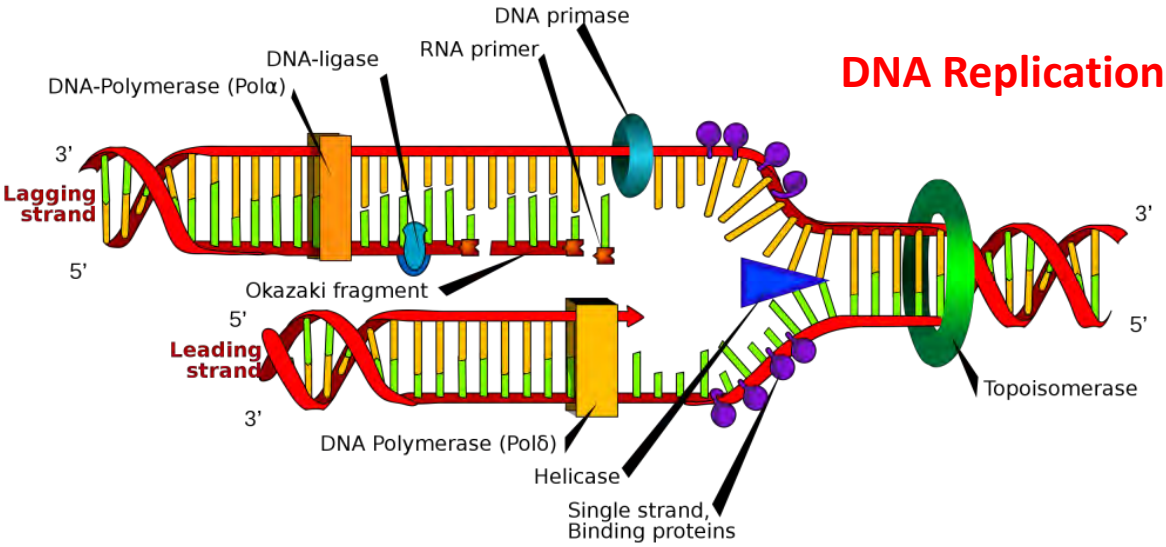
Proteins & Enzymes:

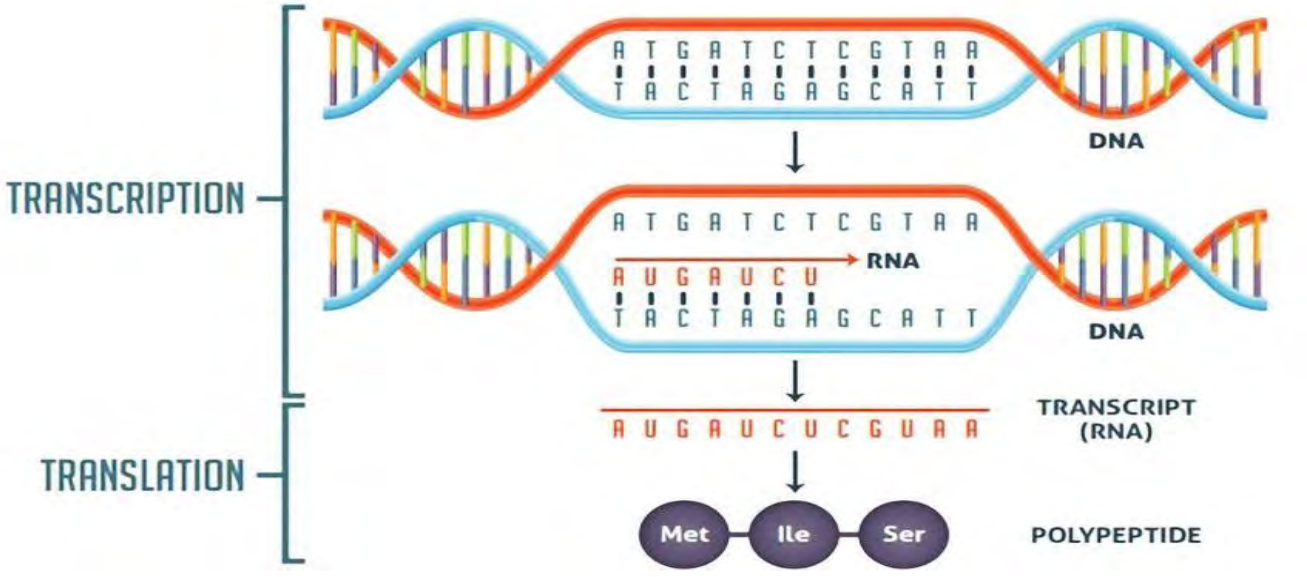
20 amino acids form proteins in a specific sequence.

Proteins form **enzymes** that regulate **biochemical reactions**.

Importance:

Proper biochemical reactions are essential for **cell function and life**.





Rosalind Franklin and DNA

Who Was Rosalind Franklin?

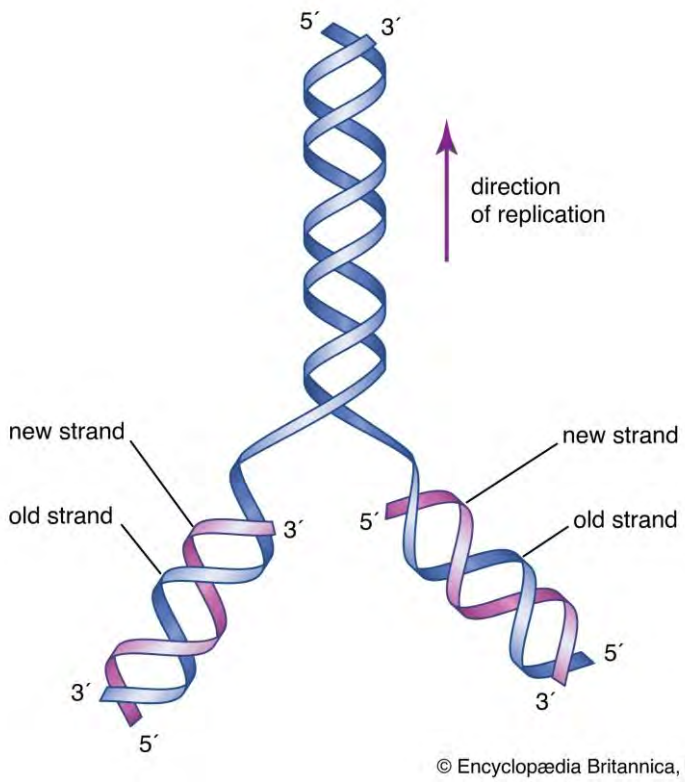
British **X-ray crystallographer** (1920–1958).
Best known for her work on **DNA structure**.
Also researched **RNA, viruses, coal, and graphite**.

Her Contribution to DNA Discovery

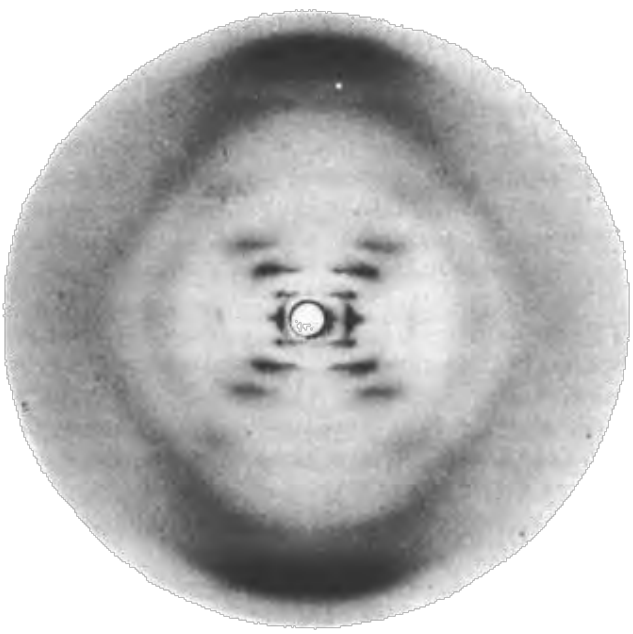
Used **X-ray diffraction** to capture images of DNA fibers.
Produced the famous **"Photo 51"**, which showed DNA's **double-helix structure**.
Her research was used by **Watson & Crick** to develop the DNA model, but she was not credited at the time.

Recognition & Legacy

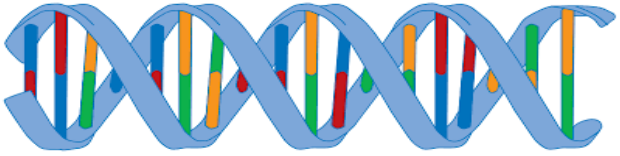
Did not receive the Nobel Prize, which was awarded in 1962 to Watson, Crick, and Wilkins.
Today, she is recognized as a key figure in **molecular biology**.
Her work paved the way for **genetic research, medicine, and biotechnology**.



© Encyclopædia Britannica, Inc.



Watson & Crick and DNA



Who Were Watson & Crick?

James Watson (American biologist) and **Francis Crick** (British physicist).

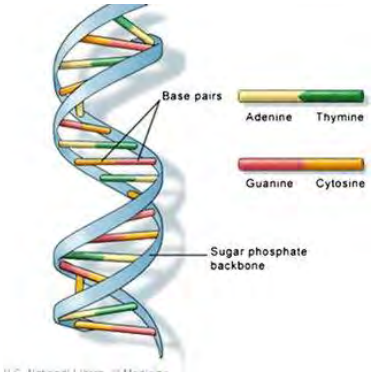
Worked together at the **University of Cambridge**.

Used **X-ray diffraction data**, including Rosalind Franklin's **Photo 51**, to model DNA's structure.

Their Discovery (1953)

Proposed the **double-helix model** of DNA.

Showed that DNA has two strands twisted like a **spiral staircase**.



Identified **base pairing rules**:

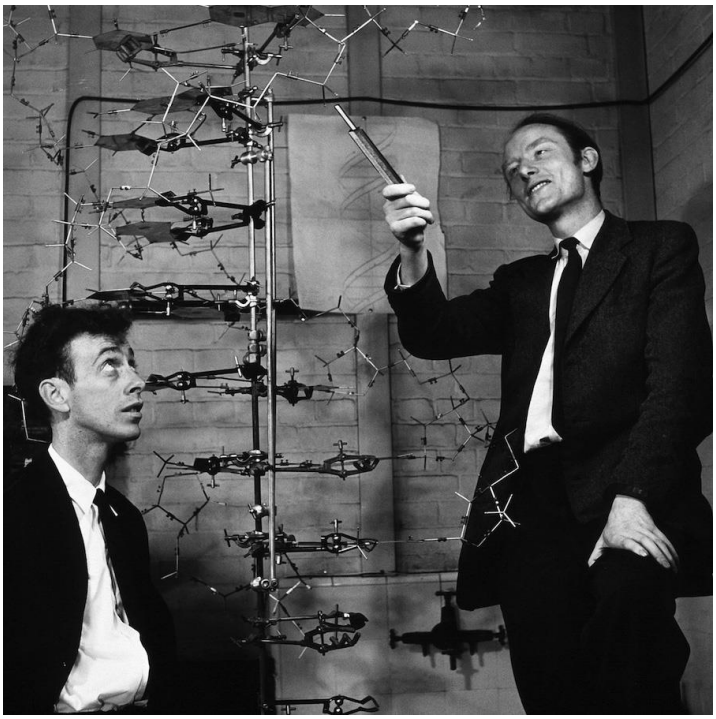
- **Adenine (A) pairs with Thymine (T)**
- **Cytosine (C) pairs with Guanine (G)**

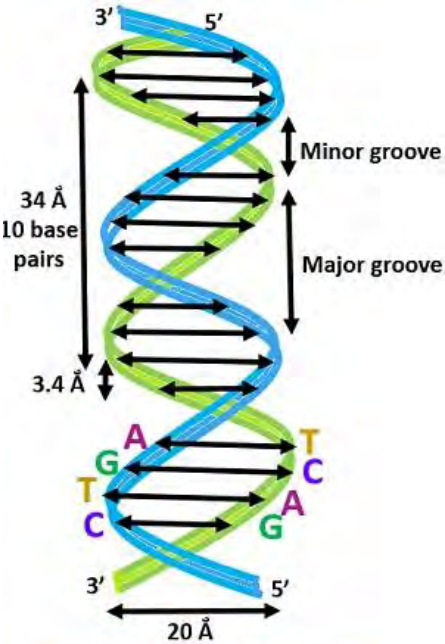
Published their findings in *Nature* (1953).

Recognition & Nobel Prize (1962)

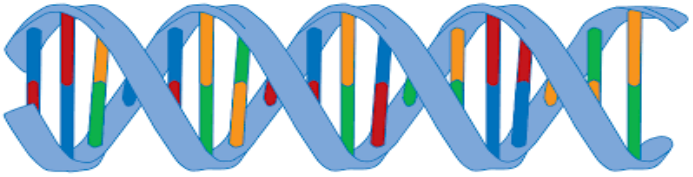
Awarded the **Nobel Prize in Physiology or Medicine** (with Maurice Wilkins).

Their discovery became the foundation of **modern genetics and biotechnology**.





STRUCTURE OF DNA BY WATSON AND CRICK



Maurice Wilkins and DNA

Who Was Maurice Wilkins?

British biophysicist (1916–2004).

Worked at **King's College London** on DNA structure.

Collaborated with **Rosalind Franklin** on X-ray diffraction studies.

His Contribution to DNA Discovery

Used **X-ray crystallography** to study DNA fibers.

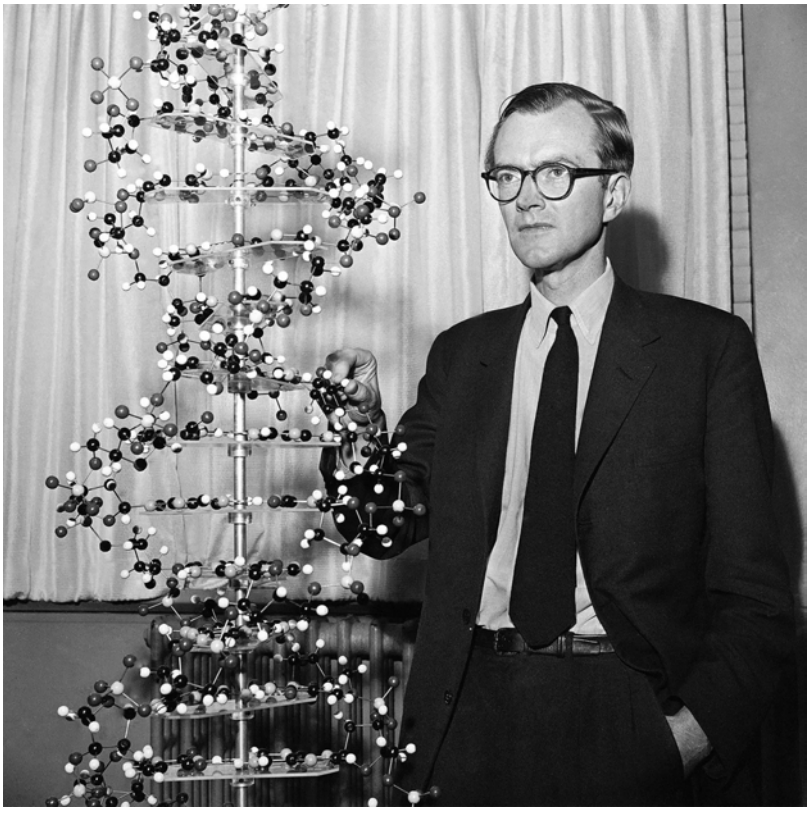
His work, along with **Franklin's Photo 51**, provided key evidence for the **double-helix structure**.

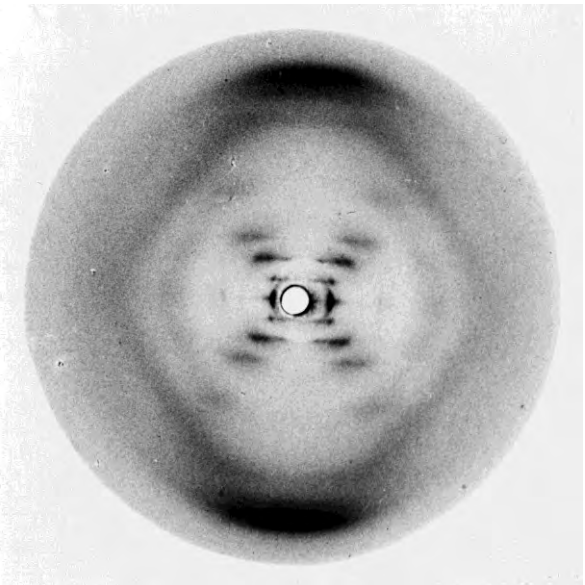
Shared Franklin's data (without her direct permission) with Watson & Crick.

Nobel Prize & Recognition

Awarded the **Nobel Prize in Physiology or Medicine (1962)** along with **Watson and Crick**.

His contributions helped confirm the **structural model of DNA**, which revolutionized **genetics and molecular biology**.





Tools Used in Molecular Taxonomy Sequencing

A. DNA Extraction & Preparation Tools

1. Qiagen DNeasy Kits

Silica-based column purification for fast and efficient DNA extraction.

Provides **high-purity DNA** suitable for PCR, sequencing, and other applications.

2. Promega Wizard Genomic DNA Purification Kits

Uses **resin-based purification** for genomic DNA isolation.

Suitable for a variety of sources, including **blood, tissue, and cells**.

3. Phenol-Chloroform Extraction

Traditional **organic extraction method** for DNA purification.

Uses **phenol, chloroform, and isoamyl alcohol** to separate DNA from proteins and lipids.

Produces **high-quality, high-yield DNA** but requires careful handling of toxic chemicals.

4. Nanodrop Spectrophotometer *(For DNA Quantification)*

Measures **DNA concentration and purity** using UV absorption (260/280 nm ratio).

Requires only **1-2 µL** of sample.

Provides a **quick and easy** assessment of DNA quality.

5. Qubit Fluorometer *(For DNA Quantification)*

Uses **fluorescent dyes** for **highly specific DNA quantification**.

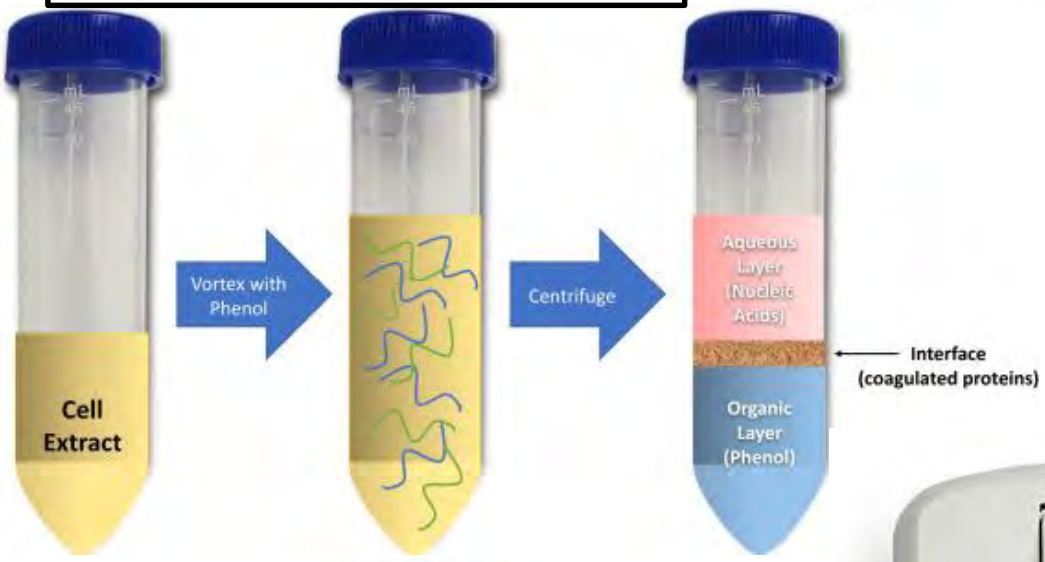
More accurate than Nanodrop for **low-concentration DNA samples**.

Minimizes interference from RNA and contaminants.

Qiagen DNeasy Kits

Promega Wizard Genomic DNA Purification Kits

Phenol-Chloroform Extraction



Nanodrop Spectrophotometer

Qubit Fluorometer

B. PCR Amplification & Primer Design

1. Thermo Fisher DreamTaq Polymerase

High-performance Taq polymerase for **PCR amplification**.

Provides **high yield and specificity** with an optimized buffer system.

Suitable for routine **genotyping, cloning, and sequencing** applications.

2. TaqMan Polymerase

Hot-start polymerase designed for **real-time quantitative PCR (qPCR)**.

Uses **fluorescent probes** to enable precise **quantification of DNA/RNA**.

Highly specific, reducing **non-specific amplification**.

3. Primer3 (Primer Design Software)

Open-source software for **designing primers** for PCR and sequencing.

Allows users to customize **melting temperature (Tm), GC content, and product size**.

Used in **gene expression, SNP genotyping, and mutagenesis studies**.

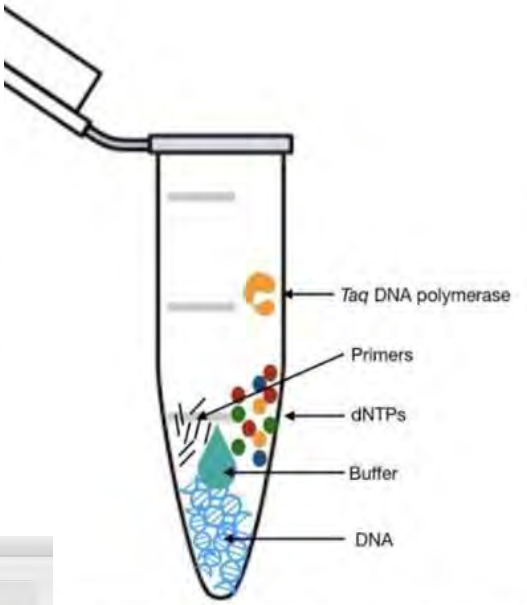
4. OligoAnalyzer (IDT)

Web-based tool by **Integrated DNA Technologies (IDT)** for **primer analysis**.

Checks **GC content, Tm, self-dimerization, and hairpin formation**.

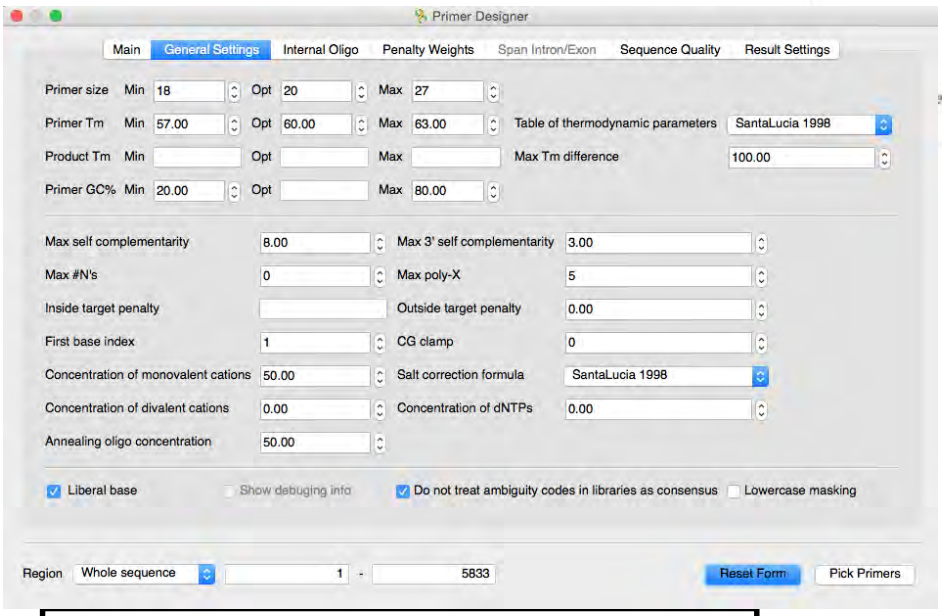
Helps optimize primers for **PCR, qPCR, and sequencing applications**.

Thermo Fisher DreamTaq Polymerase

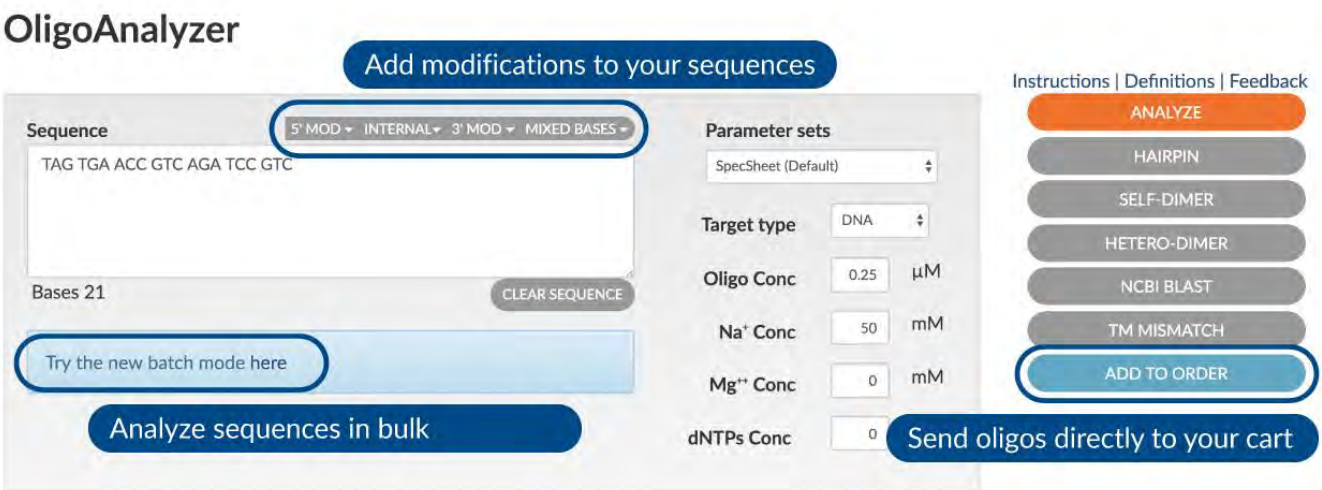


erent component while performing PCR reaction. Image copyright to ©Genetic Education Inc.

TaqMan Polymerase



Primer3 (Primer Design Software)



OligoAnalyzer (IDT)

C. Sequencing Platforms

1. Sanger Sequencing

Gold standard for DNA sequencing, used for **small-scale projects**.

Uses **chain termination method** (dideoxy sequencing).

Applied Biosystems (ABI) 3730xl DNA Analyzer:

High-throughput capillary electrophoresis system.

Commonly used for **gene sequencing, mutation detection, and DNA barcoding**.

2. Next-Generation Sequencing (NGS) Platforms *(High-Throughput & Scalable)*

Allows **massive parallel sequencing** for **whole-genome, exome, and transcriptome studies**.

Illumina Sequencing (Most widely used):

MiSeq – Small-scale targeted sequencing.

HiSeq – High-throughput sequencing for larger projects.

NovaSeq – Ultra-high throughput for large genome studies.

Ion Torrent:

Uses **semiconductor sequencing technology**.

Faster turnaround but less widely used than Illumina.

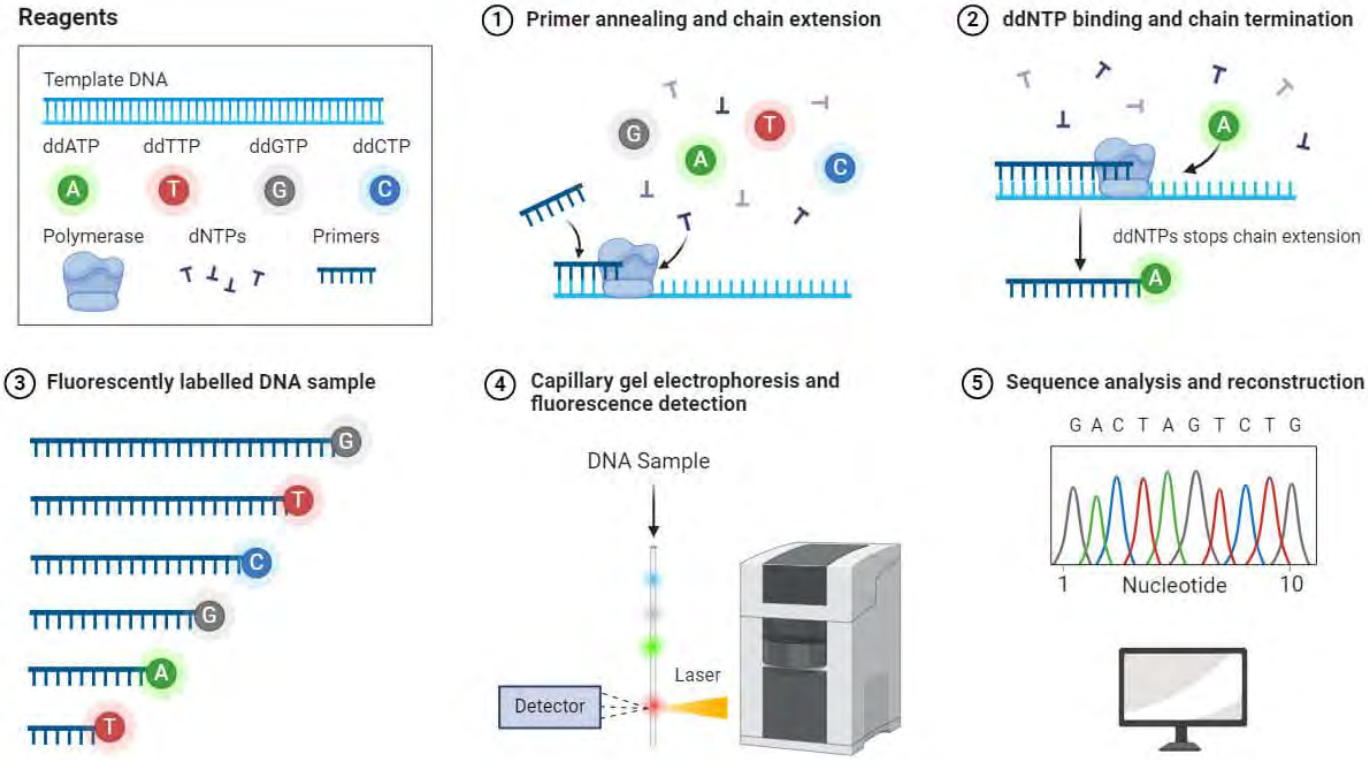
Roche 454 (Discontinued):

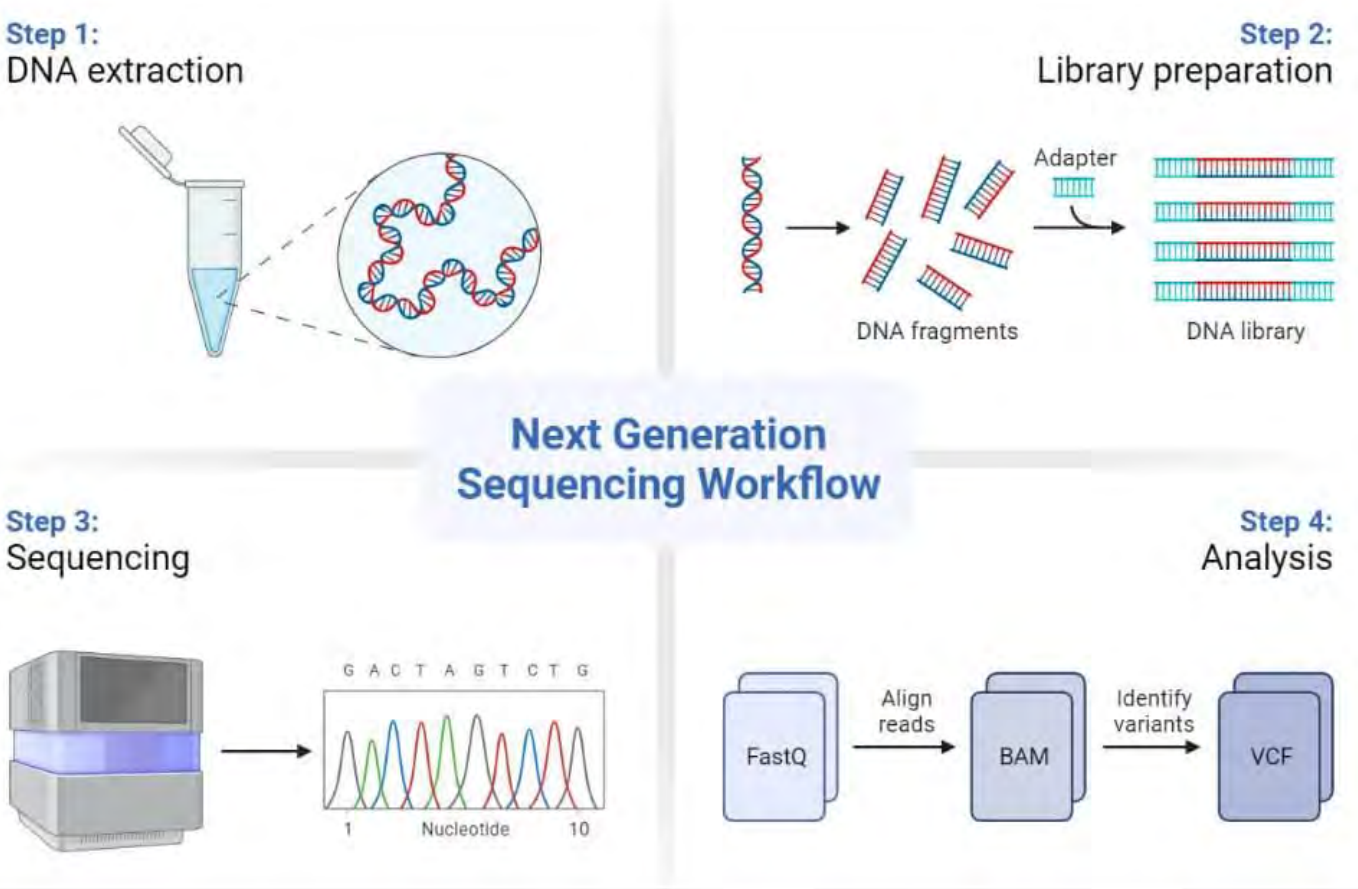
Early pyrosequencing platform, replaced by newer technologies.

SOLiD Sequencer:

Uses **two-base encoding** for high accuracy but lower adoption.

Sanger Sequencing





3. Third-Generation Sequencing Platforms *(Long-Read Sequencing & Real-Time Data)*

PacBio Sequel & Sequel II (SMRT Sequencing):

Single-molecule real-time (SMRT) sequencing allows **long reads**.

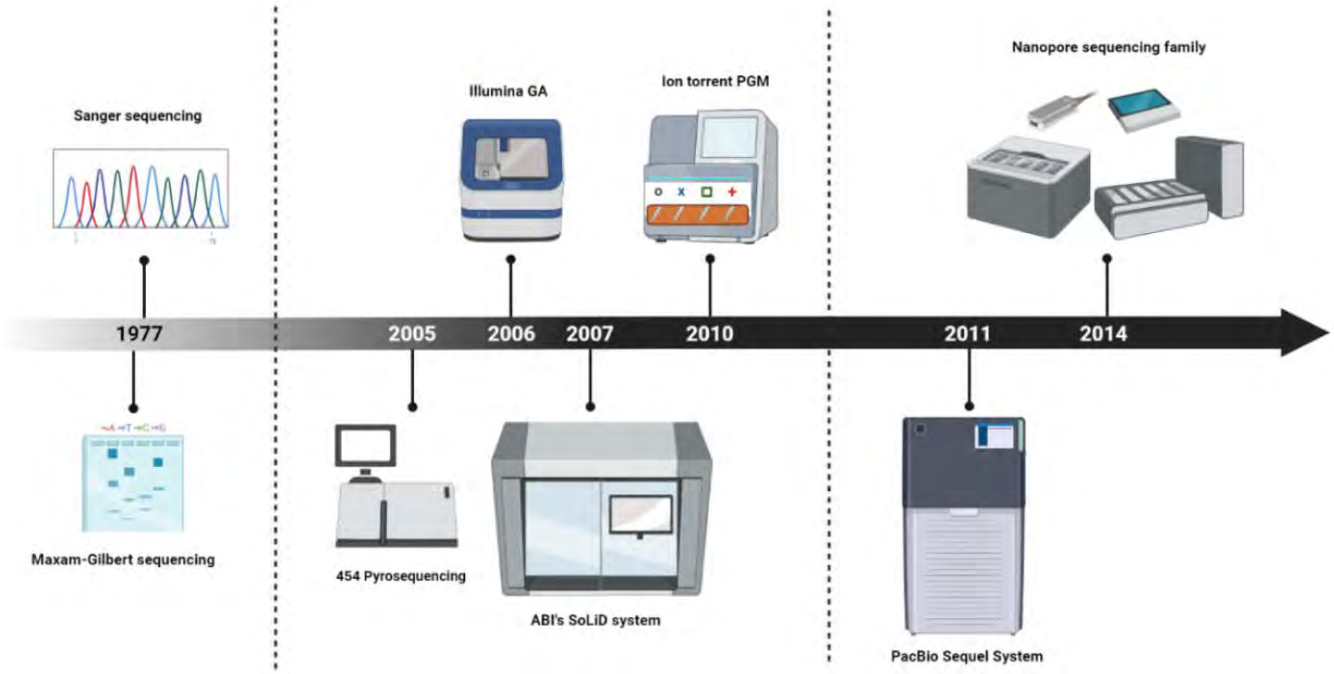
Used for **complex genome sequencing, structural variants, and full-length transcripts**.

Oxford Nanopore Technologies:

MinION – Portable, real-time sequencing device.

PromethION – High-throughput system for **large-scale sequencing projects**.

Can sequence **ultra-long DNA fragments**, useful for **structural variants and**



D. Bioinformatics & Data Analysis Tools

1. Sequence Assembly & Alignment *(For comparing and analyzing DNA sequences)*

MEGA (Molecular Evolutionary Genetics Analysis) – Used for **sequence alignment, phylogenetics, and evolutionary analysis**.

Clustal Omega – Fast and accurate **multiple sequence alignment tool**.

MUSCLE (Multiple Sequence Alignment) – High-accuracy tool for **aligning large sequence datasets**.

BLAST (Basic Local Alignment Search Tool) – Identifies **similar sequences** in large databases, essential for **genome annotation**.

2. Phylogenetic Analysis *(For constructing evolutionary relationships between species)*

BEAST (Bayesian Evolutionary Analysis) – Bayesian statistical approach for **evolutionary modeling and molecular dating**.

RAxML (Randomized Axelerated Maximum Likelihood) – Fast and efficient **phylogenetic tree construction** using **maximum likelihood**.

PhyML – Another maximum likelihood tool for **analyzing large datasets**.

MrBayes – Uses **Bayesian inference** for phylogenetic tree estimation.

Sequence Assembly & Alignment

3. Genomic & Metagenomic Analysis *(For analyzing microbial and complex DNA samples)*

QIIME – Used for **microbial community sequencing** and **16S rRNA analysis**.

Kraken – Assigns taxonomic labels to sequences for **rapid classification**.

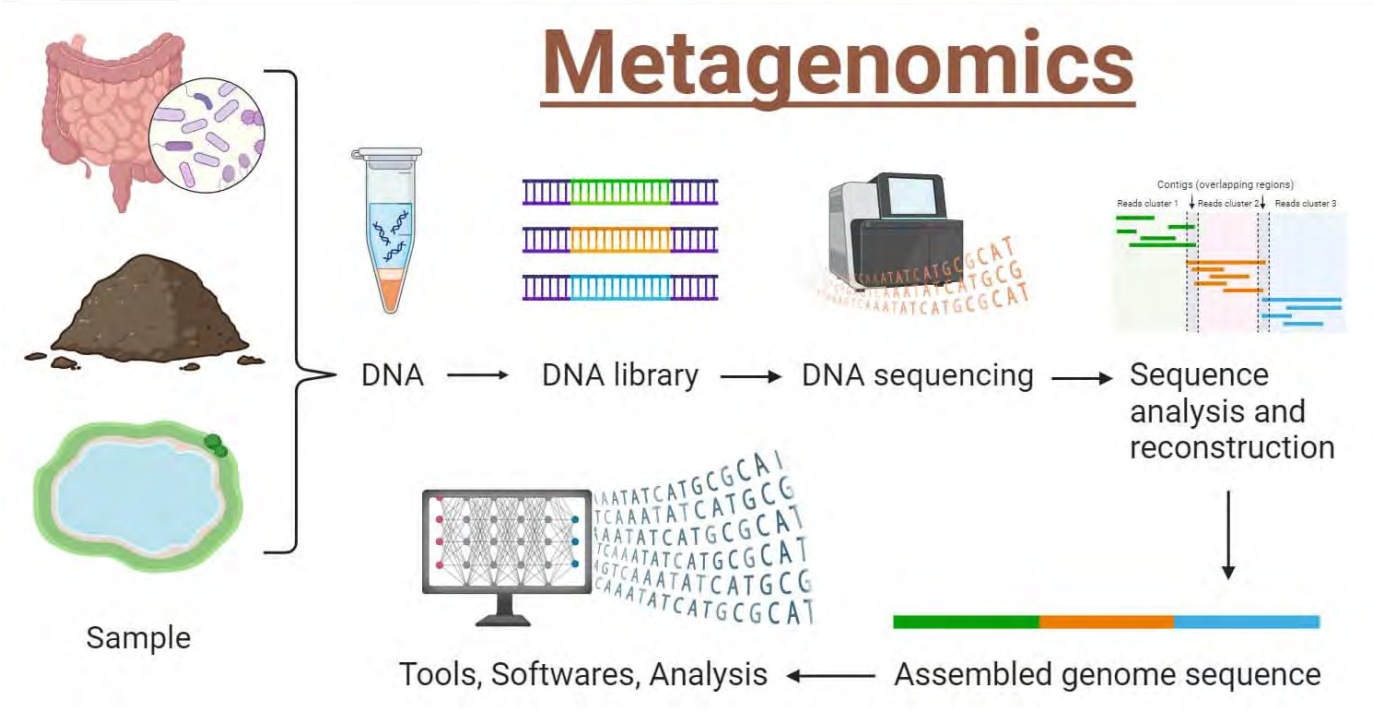
MetaPhlAn – Performs **metagenomic profiling** to determine microbial composition.

4. Variant Calling & Annotation *(For identifying genetic variations and mutations)*

GATK (Genome Analysis Toolkit) – Industry-standard tool for **detecting genetic variants** (SNPs, Indels).

SAMtools – Processes and manipulates **aligned sequence data (BAM/CRAM files)**.

BCFtools – Works with **variant calling and filtering** in sequencing data.



The most commonly used sequencing techniques depend on the application, but in general, the most widely used ones are:

1. Next-Generation Sequencing (NGS) (Most Popular Overall)

Illumina Sequencing (Dominates the market due to its high accuracy, scalability, and cost-effectiveness)

16S rRNA Sequencing (Widely used for microbial taxonomy and metagenomics)

RNA Sequencing (RNA-Seq) (Essential for transcriptomics and gene expression studies)

Whole Genome Sequencing (WGS) (Increasingly popular for comprehensive genetic studies)

2. Sanger Sequencing (Still Used for Small-Scale Applications)

Used for DNA Barcoding and small gene sequencing projects

Common in clinical and research labs for single-gene analysis

3. Third-Generation Sequencing (For Long-Read Applications)

PacBio SMRT Sequencing (Used for complex genomes, structural variants, and full-length transcripts)

Oxford Nanopore Sequencing (Gaining popularity for real-time sequencing and portable applications)

4. Metagenomic Sequencing (Increasing Use in Microbial Studies)

Used for studying complex microbial communities (e.g., gut microbiome, environmental microbiology)

5. DNA Barcoding (Widely Used in Species Identification and Taxonomy)

Common for biodiversity studies and species identification using standardized genetic markers

1. Introduction to Next-Generation Sequencing (NGS)

What is NGS?

High-throughput sequencing technology that allows **massive parallel sequencing**.

Provides **faster, cheaper, and more accurate** DNA & RNA sequencing.

Used in **genomics, transcriptomics, and metagenomics**.

Key Advantages:

✔ High scalability (from small genes to whole genomes).

✔ Cost-effective compared to traditional methods.

✔ High accuracy with deep sequencing coverage.

Major NGS Platforms

1. Illumina Sequencing (Most Widely Used)

Platforms: **MiSeq, HiSeq, NovaSeq**.

Uses **SBS (Sequencing by Synthesis)** for high accuracy.

Applications: Whole-genome sequencing, RNA-Seq, cancer genomics.

2. Ion Torrent Sequencing

Uses **semiconductor technology** (pH detection).

Faster sequencing but less accurate than Illumina.

Used for targeted sequencing & microbial genomics.

3. PacBio (Single-Molecule Real-Time - SMRT)

Provides **long-read sequencing** (great for structural variants & epigenetics).

Used in **complex genome studies**.

4. Oxford Nanopore Sequencing

Portable devices like **MinION, PromethION** for **real-time, ultra-long read sequencing**.

Useful for **real-time fieldwork & rapid pathogen detection**.

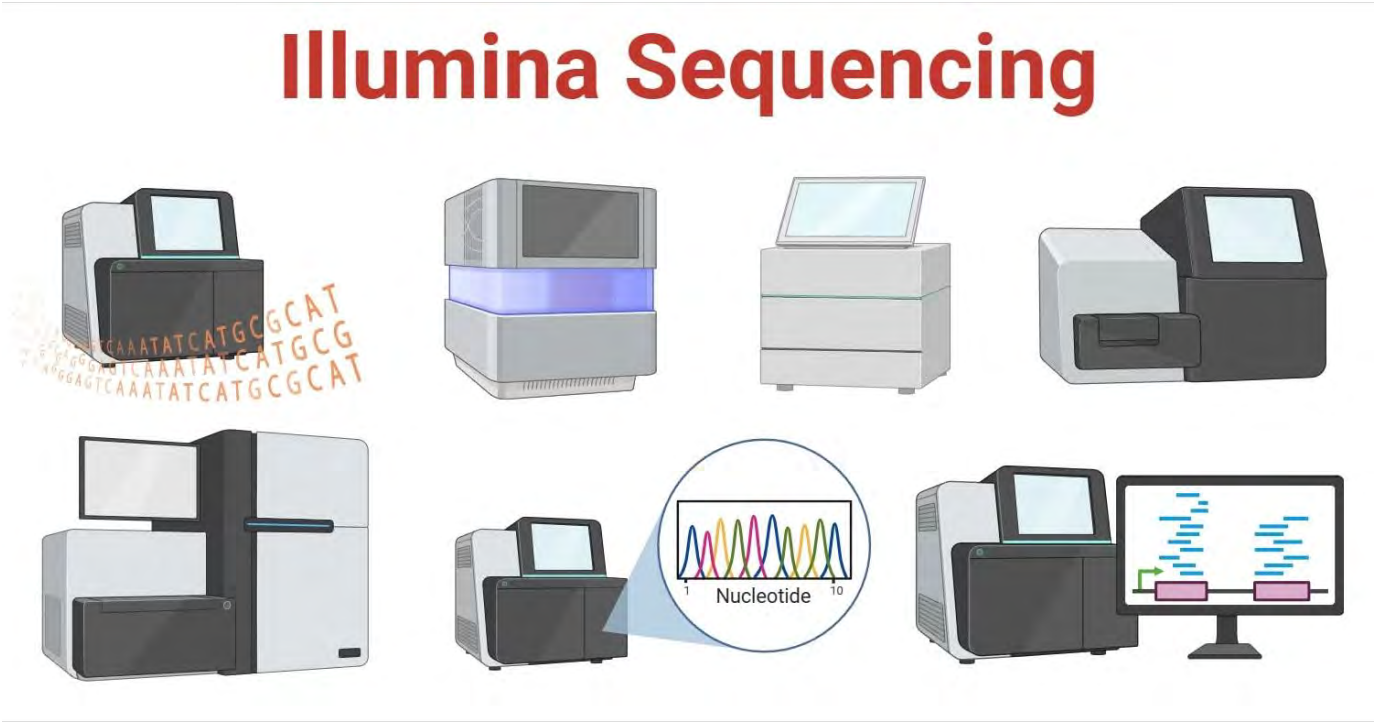
Applications of NGS

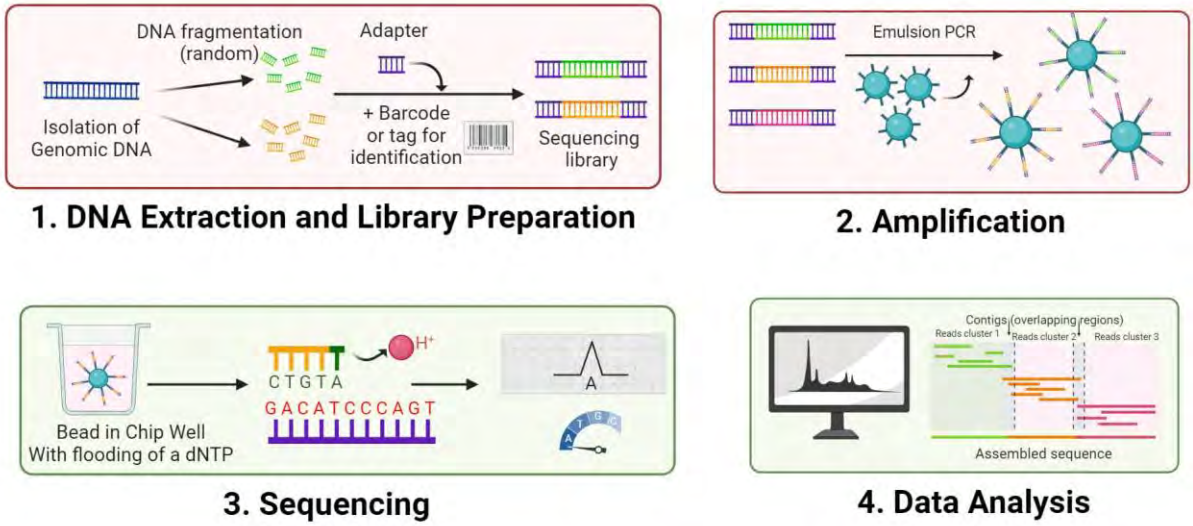
Clinical Genomics: Cancer research, genetic disorders, personalized medicine.

Microbial & Metagenomics: Identifying bacteria, viruses, and microbial communities.

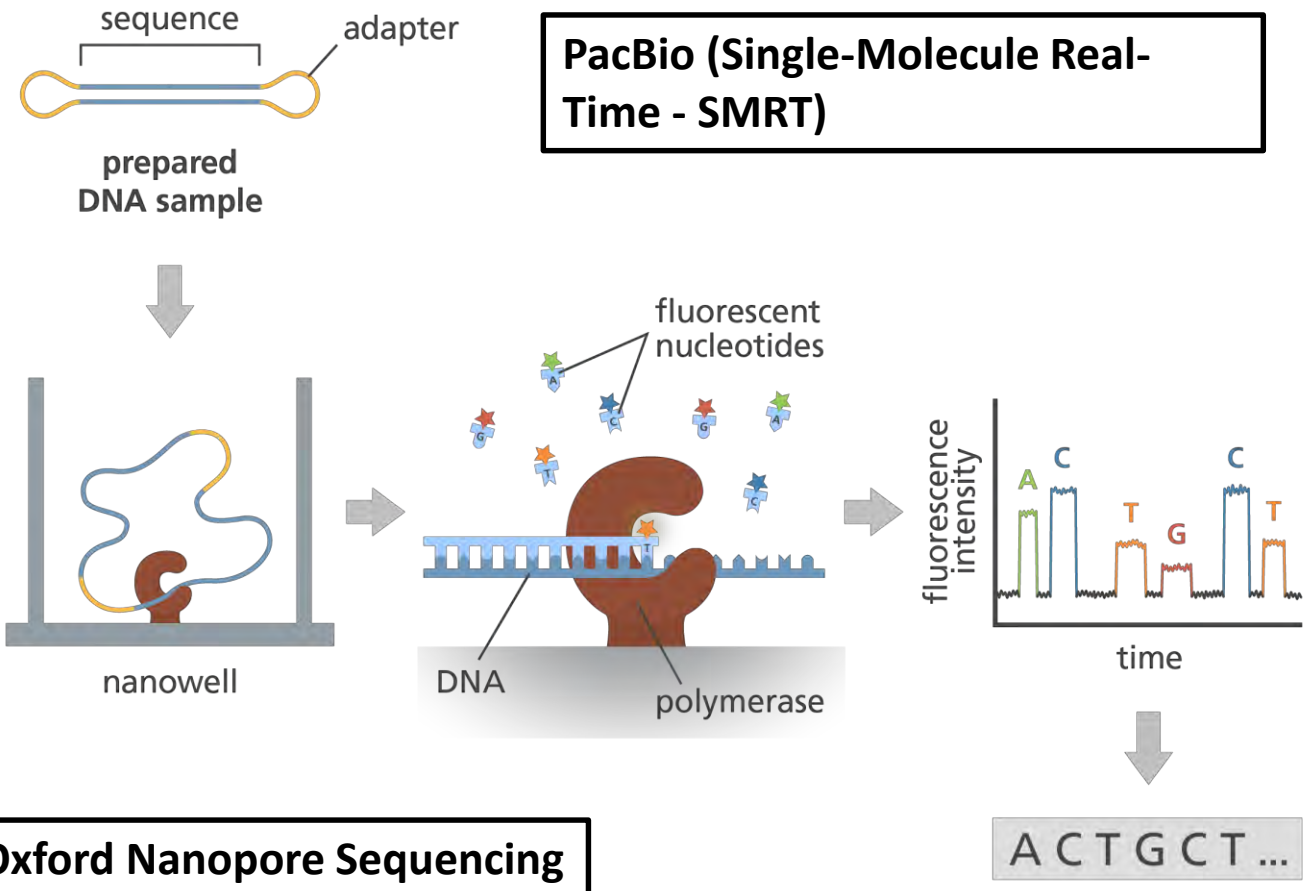
Evolutionary & Population Genetics: Studying genetic variations and ancestry.

Agrigenomics: Improving crop and livestock genetics.

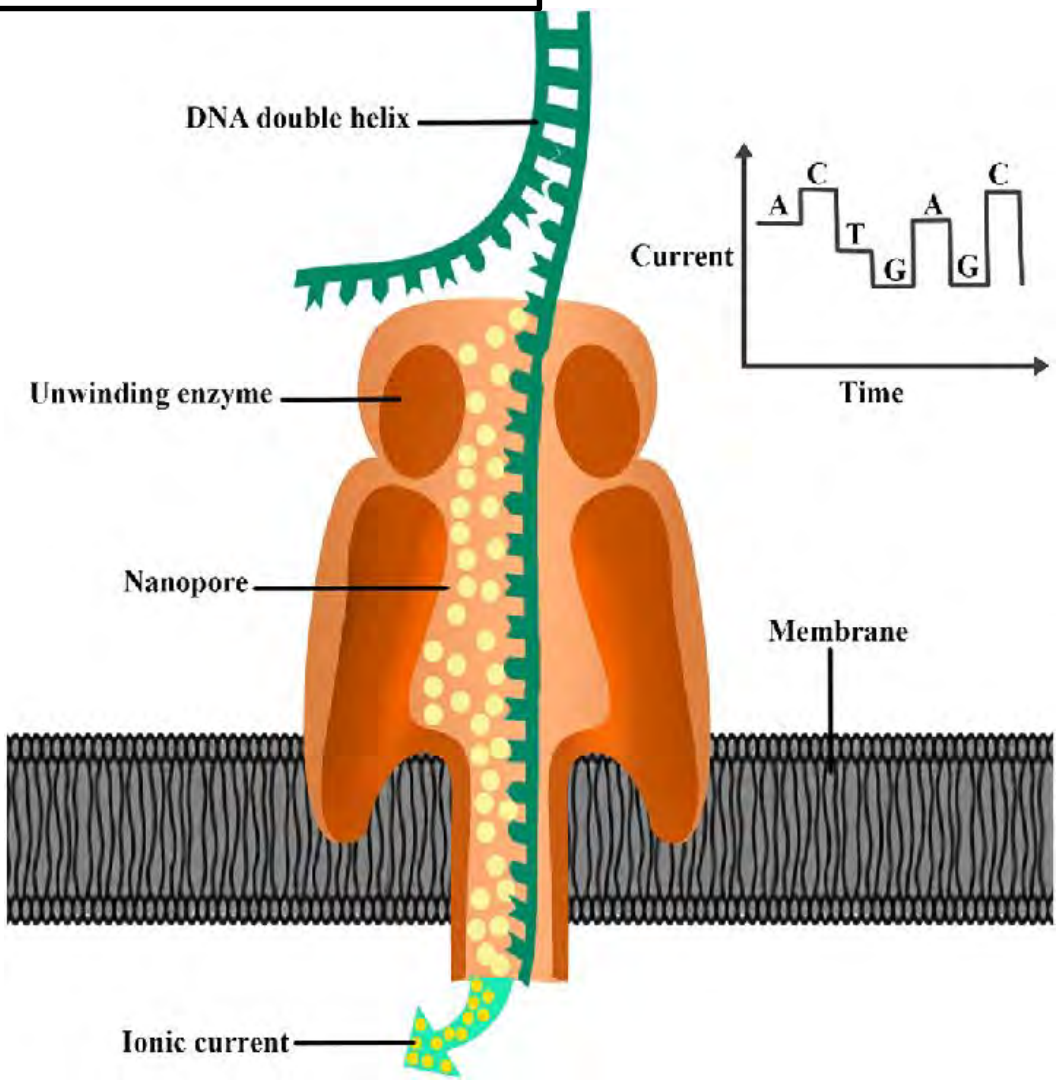




PacBio (Single-Molecule Real-Time - SMRT)



Oxford Nanopore Sequencing



2. Introduction to Sanger Sequencing

What is Sanger Sequencing?

Also known as **Dideoxy Chain Termination Sequencing**.
Developed by **Frederick Sanger in 1977**.
Uses **DNA polymerase, fluorescently labeled dideoxynucleotides (ddNTPs), and capillary electrophoresis** to determine DNA sequences.

Key Features:

Highly accurate for sequencing small DNA fragments.
Gold standard for mutation validation and clinical diagnostics.
Cost-effective for **low-throughput sequencing needs**.

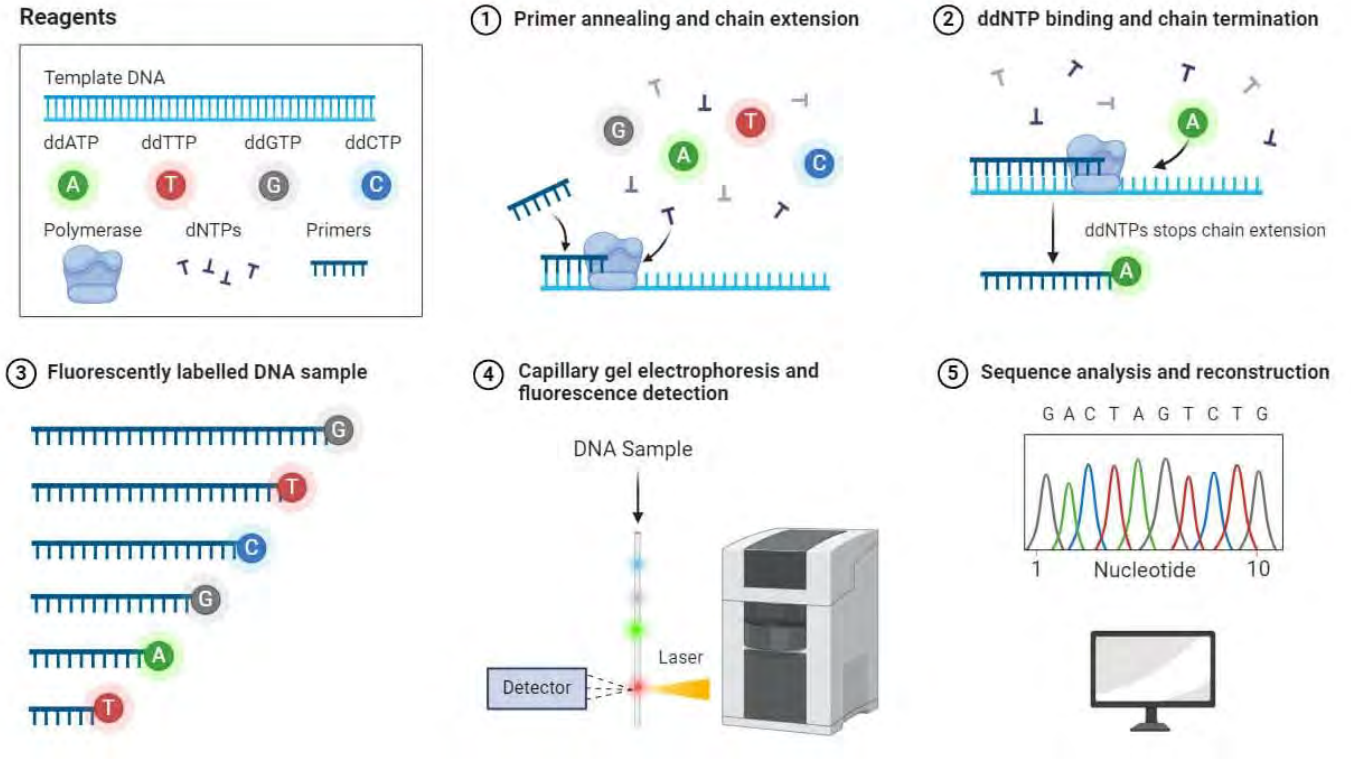
Sanger Sequencing Workflow

DNA Template Preparation – Extracted and purified DNA.
PCR Amplification – Specific region of DNA is amplified.
Dideoxy Sequencing Reaction – ddNTPs terminate chain elongation at specific bases.
Capillary Electrophoresis – DNA fragments are separated based on size.
Data Analysis – Fluorescent signals are read to determine the DNA sequence.

Advantages of Sanger Sequencing:

High accuracy (99.99%) – Ideal for mutation confirmation.
Long read lengths (up to 1,000 bp) – Good for small genes and single-gene sequencing.
Reliable for clinical applications (e.g., hereditary disease testing).

Sanger Sequencing



Limitations:

Low throughput – Not suitable for whole-genome sequencing.
Expensive for large-scale sequencing projects.
Slower compared to Next-Generation Sequencing (NGS).

Applications of Sanger Sequencing

- **Clinical Diagnostics** – Mutation detection in genetic disorders.
- **DNA Barcoding** – Species identification in biodiversity research.
- **Microbial Identification** – 16S rRNA sequencing for bacteria classification.
- **Validation of NGS Data** – Used to confirm mutations found in high-throughput sequencing studies.

3. Introduction to Third-Generation Sequencing

What is Third-Generation Sequencing?

Advanced sequencing technology that provides **long-read capabilities**.
Unlike **Next-Generation Sequencing (NGS)**, it sequences DNA **in real-time** without the need for amplification.
Used for **complex genome structures, structural variants, and epigenetic studies**.

Key Features:

Long-read sequencing (10,000 – 100,000+ bp per read).
Real-time sequencing for faster results.
Can detect epigenetic modifications (e.g., DNA methylation).

Major Third-Generation Sequencing Platforms

1. PacBio (Single-Molecule Real-Time - SMRT) Sequencing Technology:

Sequencing Technology: Uses **circular consensus sequencing (CCS)** with fluorescent-labeled nucleotides.

Advantages:
Long reads (10,000–100,000 bp).
High accuracy with **HiFi reads**.
Excellent for **structural variant detection & epigenetics**.

Applications: Whole-genome sequencing, transcriptomics, complex genome analysis.

2. Oxford Nanopore Sequencing Technology:

DNA passes through **protein nanopores**, and electrical signals are used to read sequences.

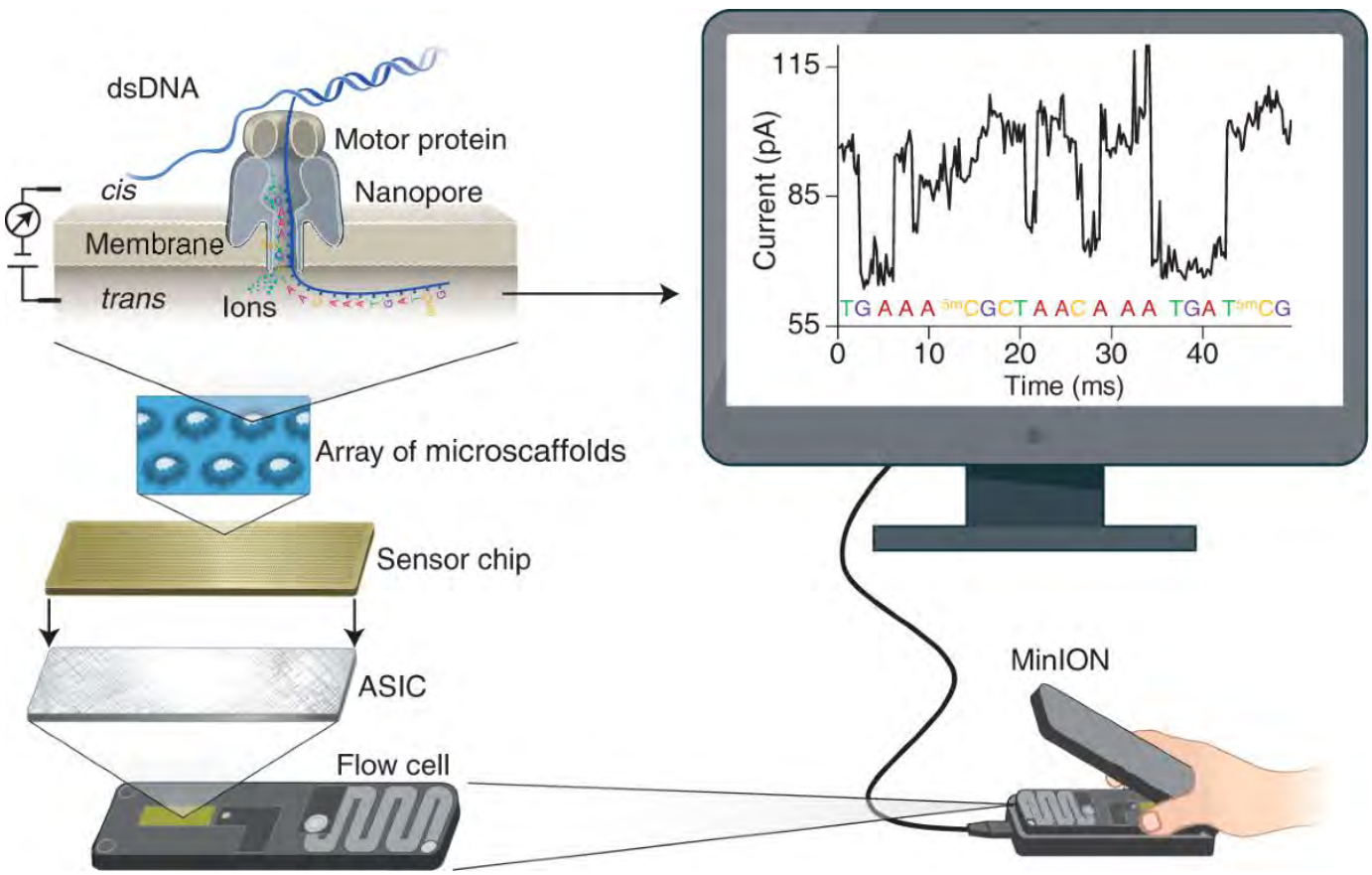
Advantages:

Ultra-long reads (up to 2 million bp).

Portable devices (MinION, PromethION) for field applications.

Can sequence **RNA directly** without conversion to cDNA.

Applications: Pathogen detection, metagenomics, real-time field sequencing.



Advantages & Limitations of Third-Generation Sequencing

Advantages:

Long-read sequencing enables detection of **large structural variations**.

No amplification bias – direct sequencing of native DNA.

Real-time sequencing for quick results.

Portable options (Nanopore MinION) for on-site analysis.

Limitations:

Higher error rates than NGS (but improving with HiFi sequencing).

Expensive compared to traditional short-read sequencing.

Requires **high-quality sample preparation** for best results.

Applications of Third-Generation Sequencing

Structural Variant Detection – Identifies complex genetic rearrangements.

Epigenetics – Direct detection of DNA methylation & modifications.

Microbial & Metagenomics – Studies microbial communities & pathogen identification.

Plant & Animal Genomics – Deciphers complex genomes (e.g., wheat, endangered species).

Clinical & Medical Research – Cancer genomics, rare disease research, transcriptomics.

4. Introduction to Metagenomic Sequencing

What is Metagenomic Sequencing?

A **high-throughput sequencing approach** that analyzes genetic material from **environmental samples** (e.g., soil, water, gut microbiome).

Identifies and characterizes **unculturable microorganisms**, providing insights into microbial diversity and function.

Uses **Next-Generation Sequencing (NGS) and Third-Generation Sequencing (TGS)** for analysis.

Key Features

Culture-independent – Studies entire microbial communities without isolation.

High-resolution taxonomic and functional profiling.

Useful for human health, environmental science, and biotechnology.

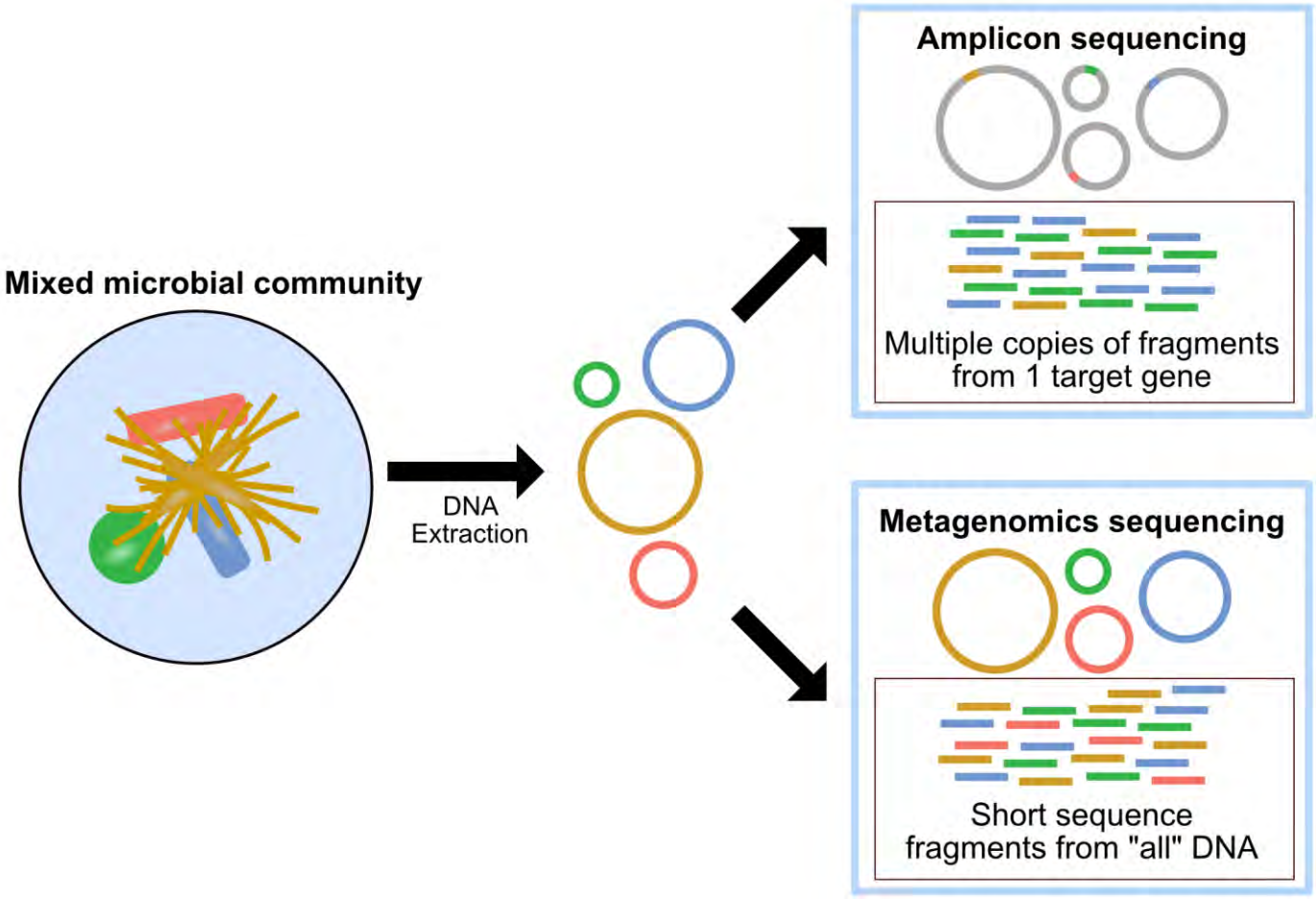
Types of Metagenomic Sequencing Approaches

1. Amplicon Sequencing (Marker-Gene Sequencing)

Focuses on **specific gene markers** (e.g., 16S rRNA for bacteria, ITS for fungi).

Identifies microbial **taxonomy & diversity** in a given sample.

Common tools: **QIIME, Mothur, SILVA database**.



2. Shotgun Metagenomic Sequencing

Sequences **entire microbial genomes** present in a sample.

Provides insights into **microbial functions, resistance genes, and metabolic pathways**.

Requires **high computational power** for analysis.

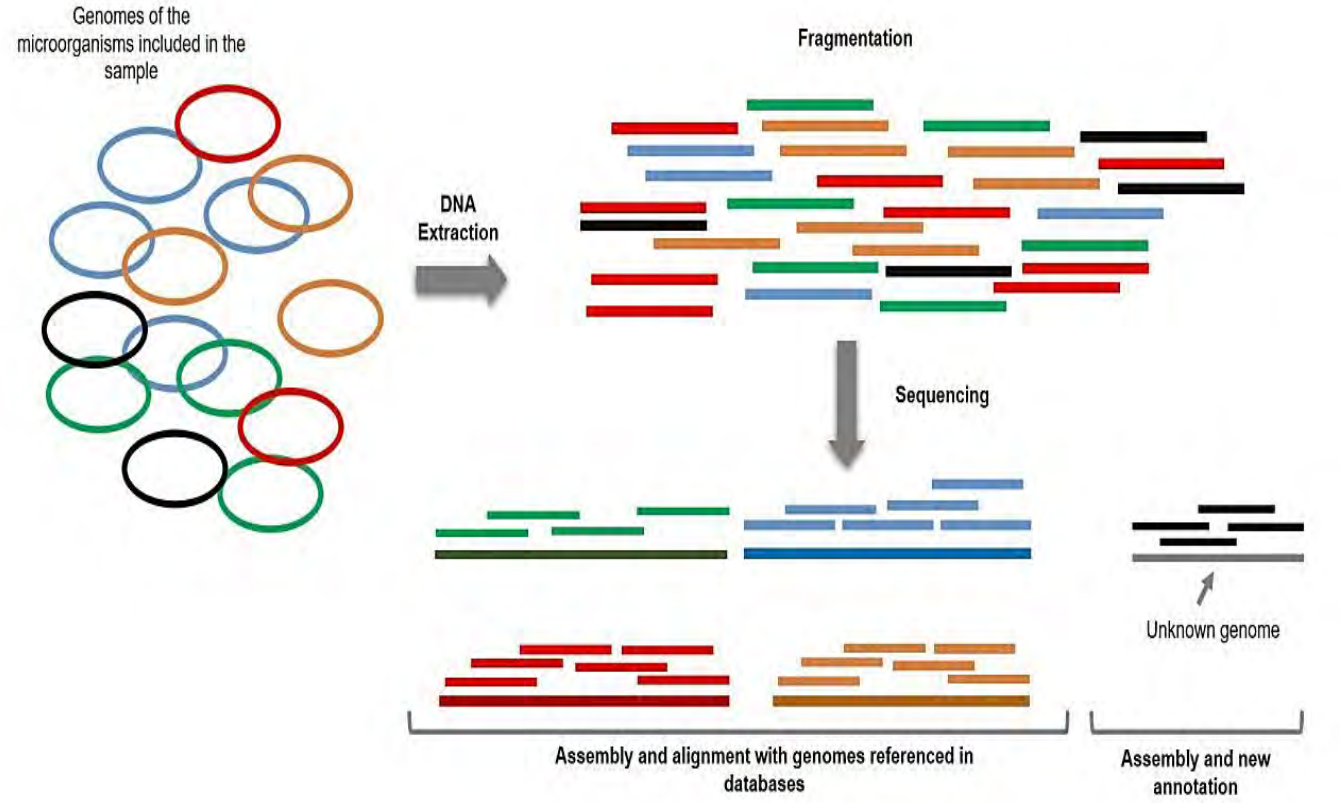
Common tools: **Kraken, MetaPhlAn, HUMAnN**.

3. Metatranscriptomics (RNA-Sequencing of Microbial Communities)

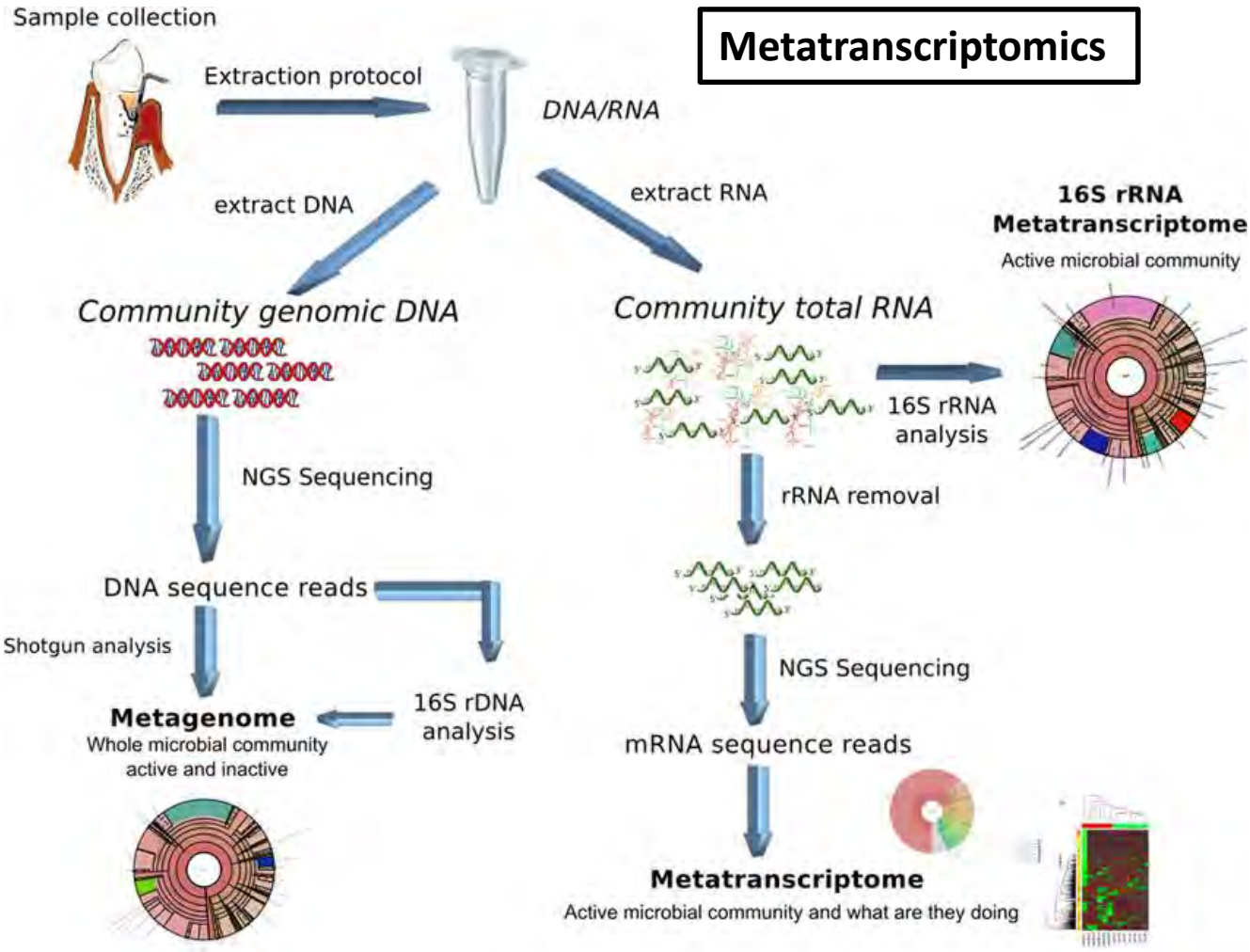
Studies **active microbial genes** (expression profiling).

Helps understand microbial **functions in response to environmental changes**.

Shotgun Metagenomic Sequencing



Metatranscriptomics



Advantages & Limitations of Metagenomic Sequencing

Advantages:

Unbiased microbial identification (detects both known & unknown species).

Provides functional insights into microbial metabolism & antibiotic resistance.

Reveals host-microbe interactions in human health and disease.

Applicable to diverse environments (soil, oceans, human gut, etc.).

Limitations:

High sequencing cost for shotgun metagenomics.

Computationally intensive (requires bioinformatics pipelines).

Potential contamination issues affecting results.

Applications of Metagenomic Sequencing

Human Microbiome Studies – Gut, oral, skin microbiome & disease correlations.

Environmental Microbiology – Identifies microbes in soil, oceans, extreme environments.

Antimicrobial Resistance (AMR) Research – Detects antibiotic resistance genes.

Biotechnology & Industrial Applications – Enzyme discovery, biofuel production.

Astrobiology & Space Research – Identifying microbial life in extreme conditions.

5. Introduction to DNA Barcoding

What is DNA Barcoding?

A molecular technique for **species identification** using **short, standardized DNA regions**.
Compares unknown DNA sequences to a **reference database** for species classification.
Based on **highly conserved yet variable genes**:

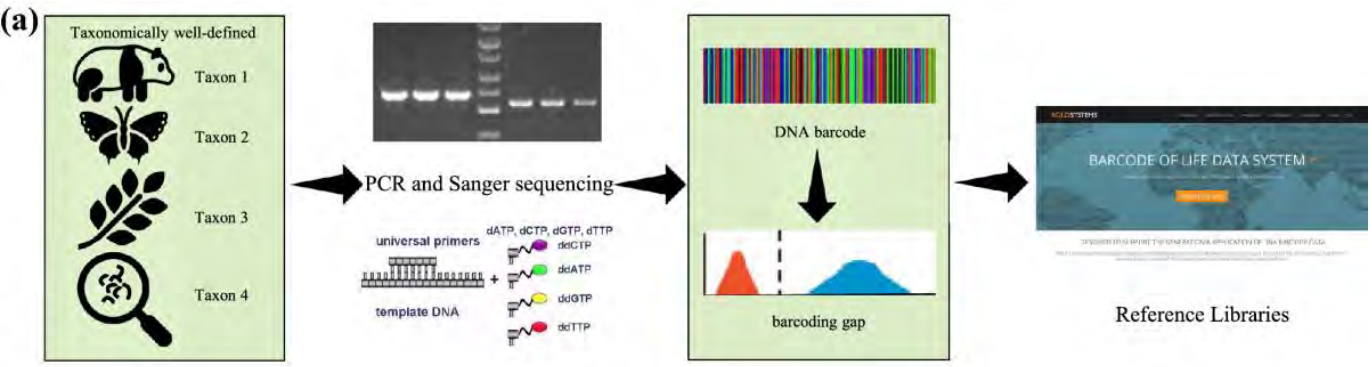
- **Animals: Cytochrome c oxidase subunit I (COI) gene** (mtDNA).
- **Plants: rbcL & matK genes** (chloroplast DNA).
- **Fungi: Internal Transcribed Spacer (ITS) region**.

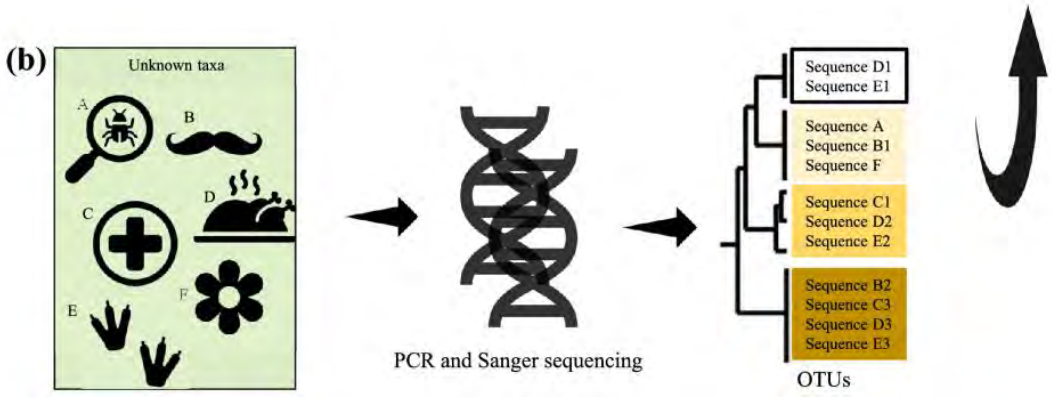
Key Features:

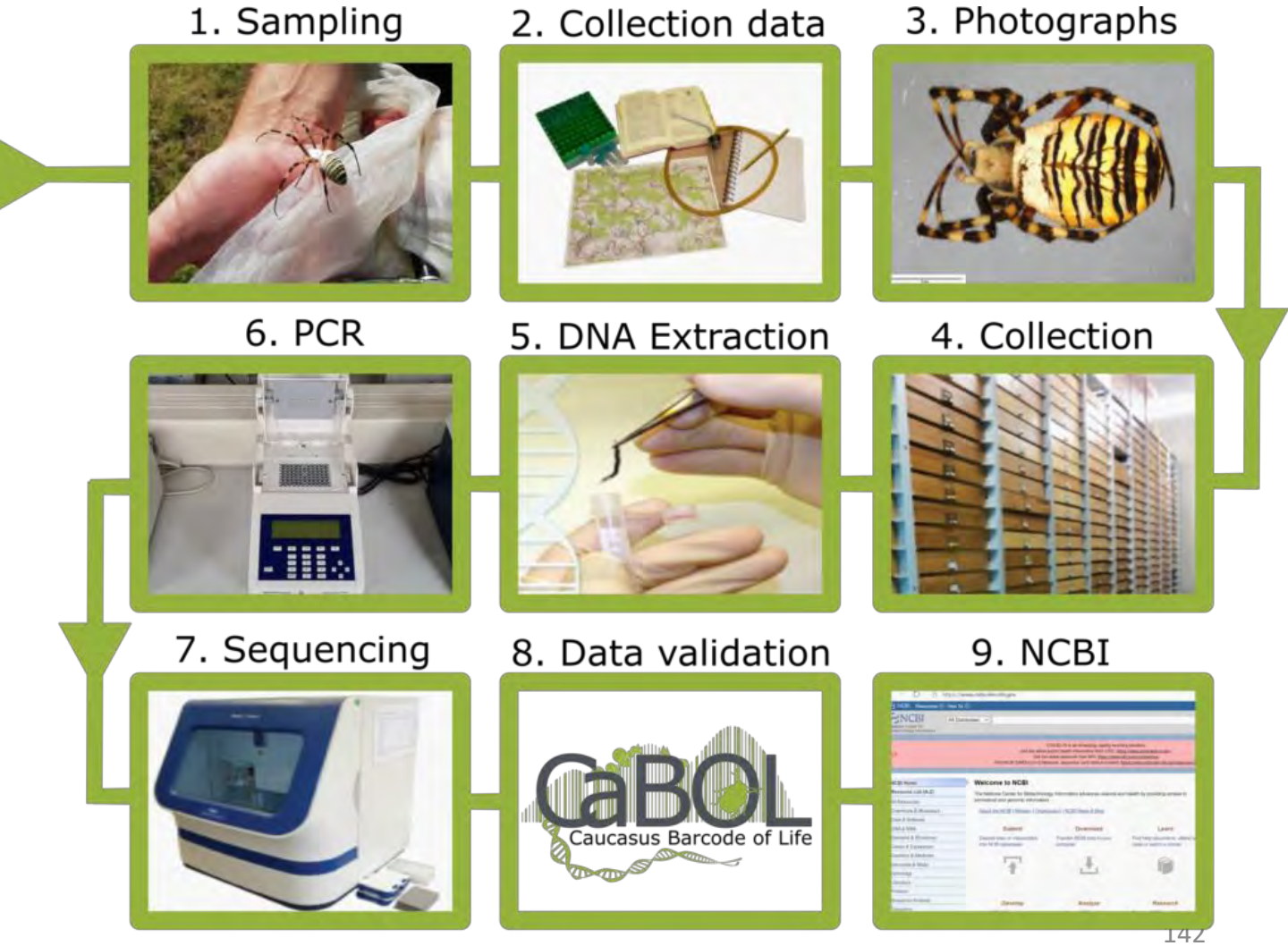
Fast and accurate species identification.
Works on degraded DNA (useful for forensic and conservation studies).
Standardized across global databases (BOLD, GenBank).

DNA Barcoding Workflow

Sample Collection – Tissue, leaf, soil, water, or preserved specimen.
DNA Extraction – Isolating genetic material from the sample.
PCR Amplification – Specific barcode gene is amplified using primers.
Sequencing – **Sanger Sequencing** or **Next-Generation Sequencing (NGS)**.
Data Analysis – Sequence comparison using **BOLD (Barcode of Life Database), GenBank (NCBI)**.
Species Identification – Matches unknown DNA with known reference sequences.


(a)
Taxonomically well-defined
Taxon 1
Taxon 2
Taxon 3
Taxon 4
PCR and Sanger sequencing
universal primers
dATP, dCTP, dGTP, dTTP
ddCTP
ddATP
ddGTP
ddTTP
template DNA
DNA barcode
barcoding gap
BARCODE OF LIFE DATA SYSTEM
Reference Libraries


(b)
Unknown taxa
A
B
C
D
E
F
PCR and Sanger sequencing
Sequence D1
Sequence E1
Sequence A
Sequence B1
Sequence F
Sequence C1
Sequence D2
Sequence E2
Sequence B2
Sequence C3
Sequence D3
Sequence E3
OTUs


1. Sampling
2. Collection data
3. Photographs
6. PCR
5. DNA Extraction
4. Collection
7. Sequencing
8. Data validation
9. NCBI
CaBOL
Caucasus Barcode of Life

Advantages & Limitations of DNA Barcoding

Advantages:

Universal & standardized method for species identification.

Works with minimal or degraded DNA samples.

Enables discovery of cryptic species (visually similar but genetically distinct).

Supports biodiversity conservation efforts.

Limitations:

Limited to known reference sequences – Cannot identify truly novel species.

Hybrid species & horizontal gene transfer can cause misidentification.

Plants require multiple barcodes (rbcL & matK), making analysis complex.

Applications of DNA Barcoding

Biodiversity & Conservation – Identifying endangered species, monitoring ecosystems.

Wildlife & Trade Regulation – Detecting illegal wildlife trafficking & food fraud.

Microbial Ecology – Identifying microbial species in complex environments.

Food Authentication – Verifying seafood & herbal medicine authenticity.

Forensic Science – Species identification from remains (e.g., bones, hair).

Advance Experimental Taxonomy

Chapter 4- Choosing molecular marker for phylogenetic analyses

Introduction to Molecular Systematics

What is Molecular Systematics?

A branch of systematics that uses molecular data (DNA, RNA, proteins) to study evolutionary relationships.

Helps classify organisms based on genetic similarities and differences.

Resolves evolutionary relationships at all taxonomic levels.

Importance:

More accurate than morphological classification.

Useful in phylogenetics, species identification, and conservation genetics.

Helps understand evolutionary history and biodiversity.

Common Molecular Markers:

Marker Type	Example	Used For
Mitochondrial DNA	COI, Cyt b	Species identification, animal phylogenetics
Nuclear DNA	ITS, 18S, 28S	Deep evolutionary relationships
Chloroplast DNA	rbcL, matK	Plant phylogenetics
Whole Genome	SNPs, RADseq	High-resolution phylogenetics

Techniques, Applications, and Challenges

Techniques in Molecular Systematics:

DNA Extraction & PCR Amplification – Isolating and amplifying genetic material.
DNA Sequencing – Sanger sequencing, Next-Generation Sequencing (NGS).
Sequence Alignment – BLAST, Clustal Omega, MUSCLE.
Phylogenetic Tree Construction – Maximum Likelihood, Bayesian Inference.

Applications:

Evolutionary Biology: Understanding species divergence.
Taxonomy: Refining classifications.
Conservation Genetics: Identifying endangered species.
Forensics: DNA-based species identification.

Challenges:

Contaminated or incomplete data.
Horizontal gene transfer (HGT) in microbes.
Computational complexity in large datasets.

Introduction to Plant Genome

What is a Plant Genome?

The complete set of genetic material (DNA) present in a plant cell.

Contains nuclear, mitochondrial, and chloroplast DNA.

Encodes genes responsible for plant growth, development, and adaptation.

Components of a Plant Genome:

Nuclear Genome:

Largest part of the genome, organized into chromosomes.

Contains genes regulating metabolism, growth, and reproduction.

Chloroplast Genome (cpDNA):

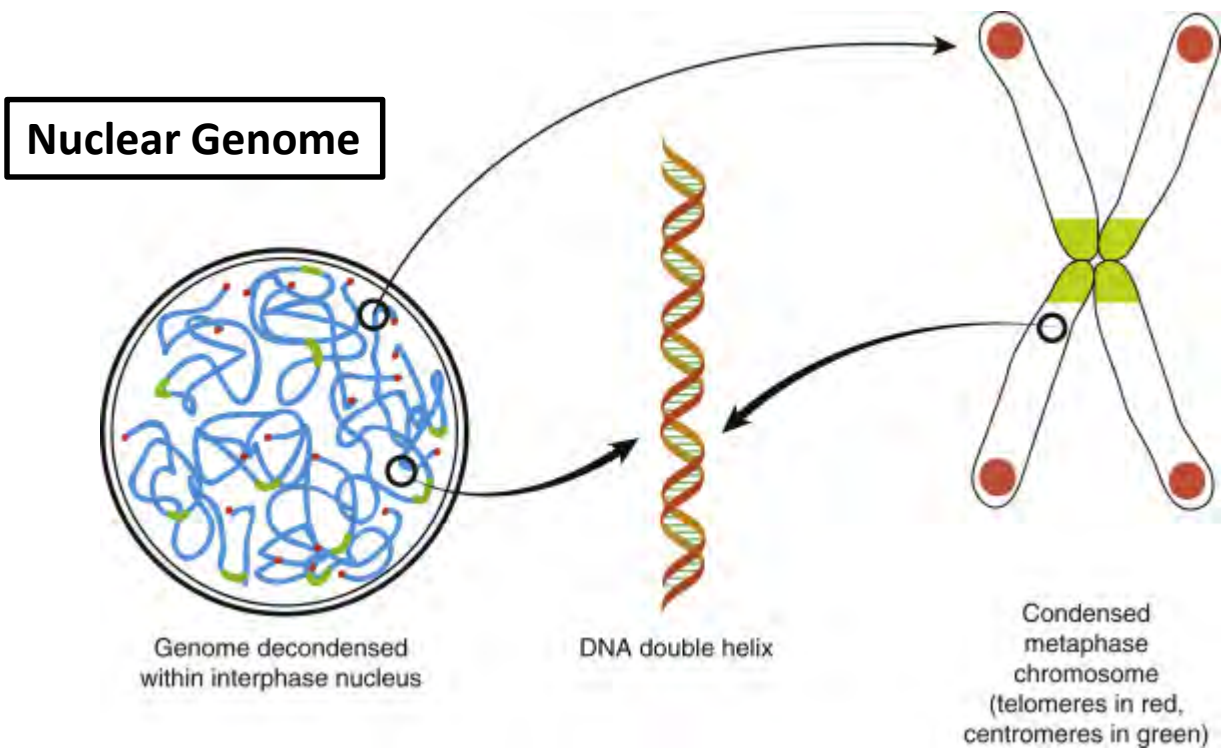
Circular DNA present in chloroplasts.

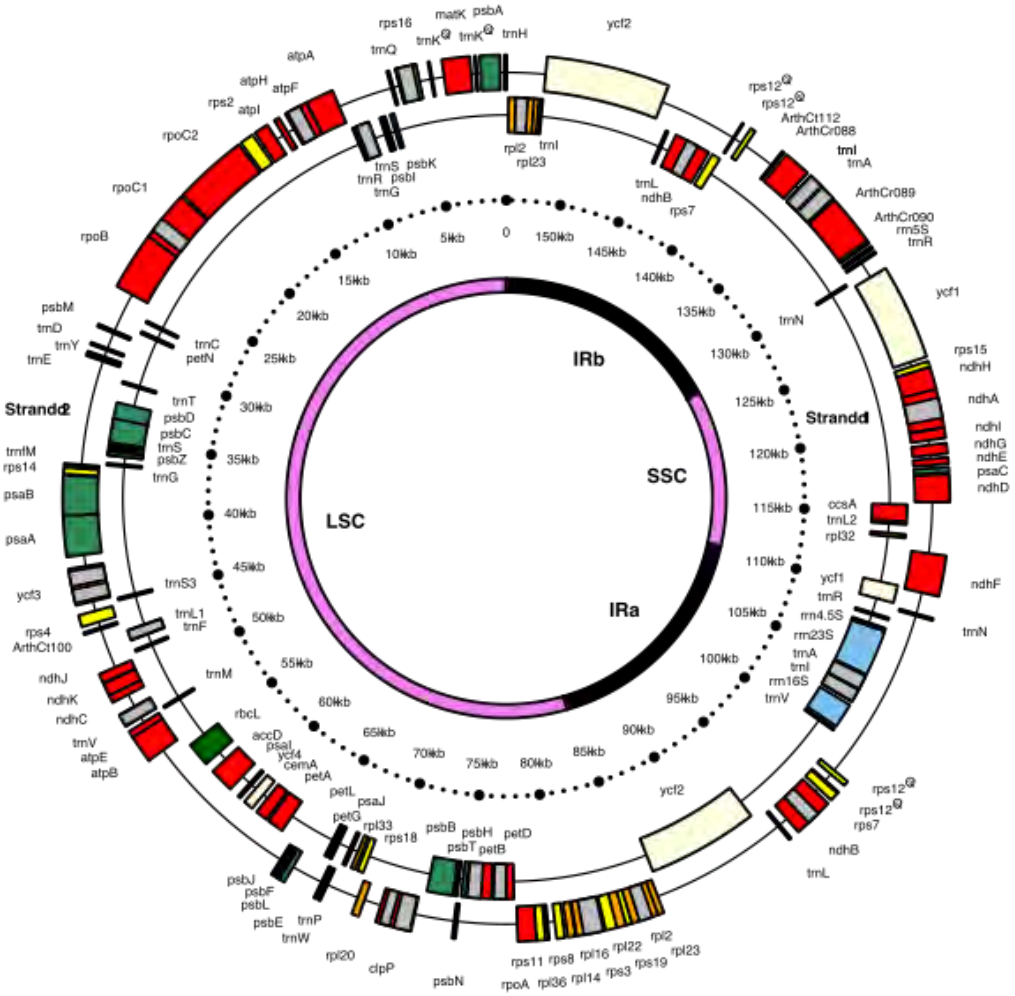
Essential for photosynthesis and energy production.

Mitochondrial Genome (mtDNA):

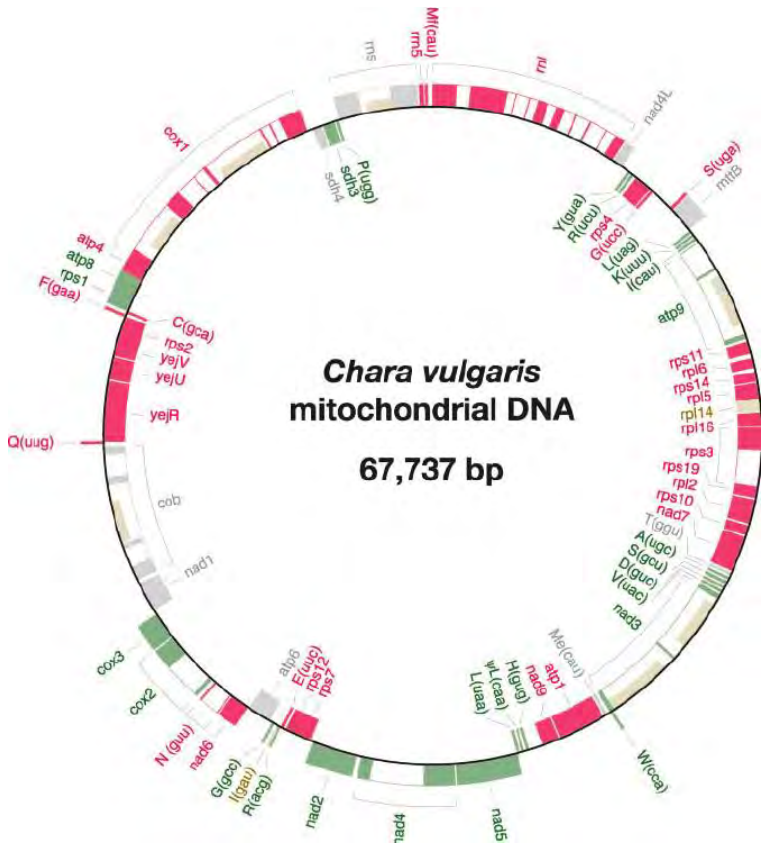
Circular DNA found in mitochondria.

Involved in respiration and energy conversion.





The 154 kb chloroplast DNA map of a model flowering plant (Arabidopsis thaliana: Brassicaceae) showing genes and inverted repeats.



Plant Genome Structure, Techniques & Applications

Genome Size & Complexity:

Varies widely among plant species (e.g., Arabidopsis thaliana: ~135 Mb, Wheat: ~16 Gb).

Contains coding (genes) and non-coding (regulatory) regions.

Techniques for Plant Genome Study:

DNA Sequencing:

Sanger sequencing (small-scale)

Next-Generation Sequencing (NGS) for whole-genome sequencing

Genomic Mapping & Assembly:

Helps identify genes and regulatory elements.

CRISPR & Genetic Engineering:

Gene editing for crop improvement.

Applications of Plant Genomics:

Crop Improvement: Enhancing yield, disease resistance, and stress tolerance.

Biodiversity Conservation: Understanding genetic diversity and evolution.

Medicinal Plants: Identifying genes responsible for bioactive compounds.

Climate Adaptation: Developing drought- and heat-resistant crops.

Molecular Markers in Plant Genomics

What are Molecular Markers?

- DNA sequences used to identify genetic variation among individuals.
- Do not get affected by environmental factors, making them reliable for genetic studies.

Types of Molecular Markers:

RFLP (Restriction Fragment Length Polymorphism)

Uses restriction enzymes to cut DNA at specific sequences.
Labor-intensive but provides high accuracy.

SSR (Simple Sequence Repeats / Microsatellites)

Short, repeating DNA sequences used for high-resolution genetic mapping.
Commonly used in plant breeding and population genetics.

AFLP (Amplified Fragment Length Polymorphism)

Uses PCR and restriction enzymes to detect DNA polymorphisms.
Useful for studying genetic diversity.

SNP (Single Nucleotide Polymorphism)

Single base-pair variations in DNA sequences.
Highly abundant and used in genome-wide association studies (GWAS).

ISSR (Inter-Simple Sequence Repeats)

Uses microsatellite regions to amplify DNA fragments.
Cost-effective and used for genetic fingerprinting.

Applications of Molecular Markers:

Crop improvement and plant breeding.
Phylogenetic and evolutionary studies.
DNA fingerprinting for species identification.
Studying genetic diversity and conservation biology.

Chloroplast Genome in Plants

What is the Chloroplast Genome?

The genetic material found in chloroplasts, essential for photosynthesis.

Circular, double-stranded DNA (cpDNA) with 120-160 kb size.

Encodes genes for photosynthesis, gene expression, and some metabolic functions.

Key Features of the Chloroplast Genome:

Inherited maternally in most plants (paternally in some species).

Highly conserved with low mutation rates compared to nuclear DNA.

Contains genes for ribosomal RNA (rRNA), transfer RNA (tRNA), and proteins involved in photosynthesis.

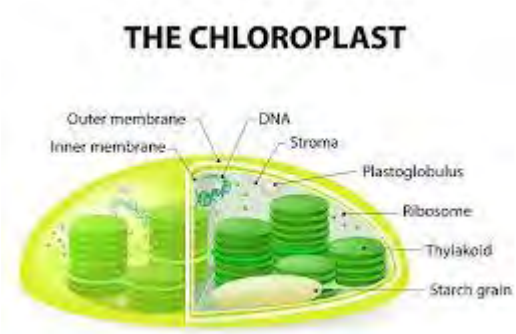
Applications of Chloroplast Genome Studies:

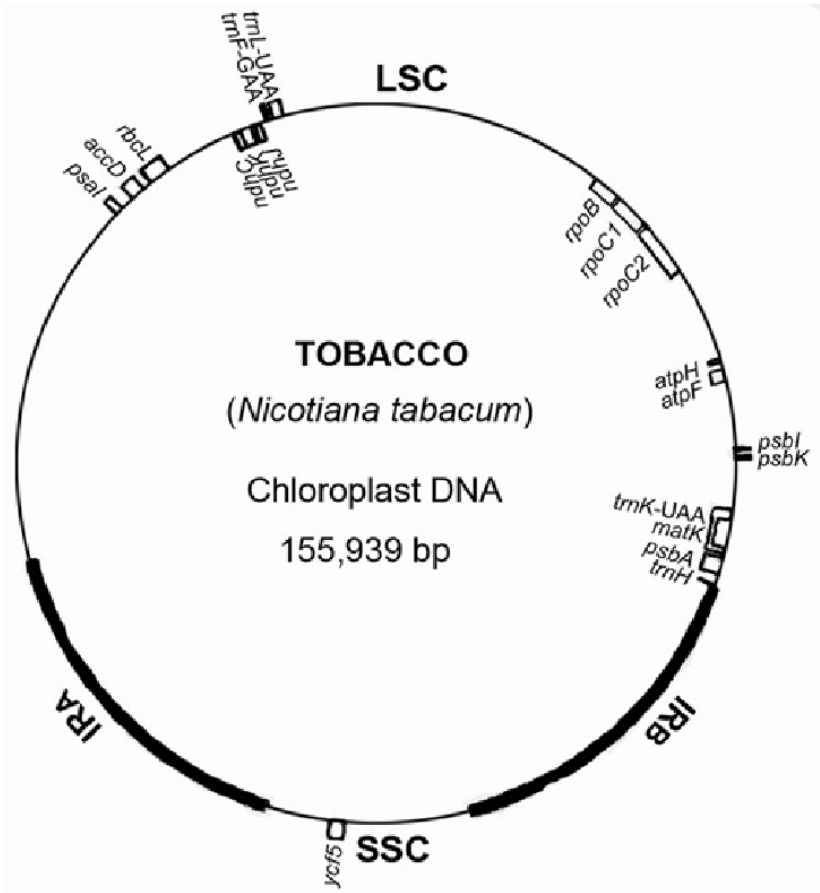
Phylogenetics: Used to trace evolutionary relationships among plant species.

Barcoding & Taxonomy: Helps in plant species identification.

Genetic Engineering: Used in transplastomic plants for improving traits.

Climate Adaptation Research: Studying photosynthetic efficiency under environmental stress.





Introduction to the Chloroplast Genome

What is the Chloroplast Genome?

The chloroplast genome (cpDNA) is the circular DNA found in chloroplasts, the organelles responsible for photosynthesis.
It is independent of the nuclear genome and mitochondria, with its own machinery for gene expression.
The size varies between **120-160 kb** and contains **100-250 genes**.

Key Features of the Chloroplast Genome

Structure

- Circular, double-stranded DNA.
- Highly conserved across plant species.
- Composed of **two inverted repeats (IRs), a large single-copy region (LSC), and a small single-copy region (SSC)**.

Gene Content

- Contains genes related to photosynthesis, transcription, and translation.
- Codes for **rRNAs, tRNAs, and essential photosynthetic proteins (e.g., psbA, rbcL).**

Inheritance Pattern

- **Mostly maternally inherited** in angiosperms.
- **Paternally inherited** in some gymnosperms.

Advantages of Chloroplast Genome for Genetic Studies

- **Highly conserved** sequence makes it useful for phylogenetics and species identification.
- **Low recombination rate** ensures stability in inheritance.
- **Efficient transformation system** for transgenic plants.

Functions, Applications, and Engineering of the Chloroplast Genome

Functions of the Chloroplast Genome

Photosynthesis: Encodes proteins required for the light-dependent and light-independent reactions.

Protein Synthesis: Contains genes for **ribosomal RNA (rRNA)** and **transfer RNA (tRNA)** essential for translation.

Regulation of Metabolism: Plays a role in **starch biosynthesis, pigment production, and stress responses**.

Applications of Chloroplast Genome Studies

Phylogenetics & Evolution:

Used to determine evolutionary relationships due to its conserved nature.

Barcoding & Taxonomy:

Used in DNA barcoding to identify plant species (e.g., **rbcL, matK genes**).

Crop Improvement:

Genetic engineering of chloroplasts can enhance photosynthesis and increase resistance to stress.

Environmental Adaptation Research:

Studying chloroplast genomes helps understand how plants adapt to extreme climates.

Chloroplast Genetic Engineering (Transplastomics)

Definition: Genetic modification of the chloroplast genome to express foreign genes.

Advantages of Transplastomics:

High-level expression of transgenes.

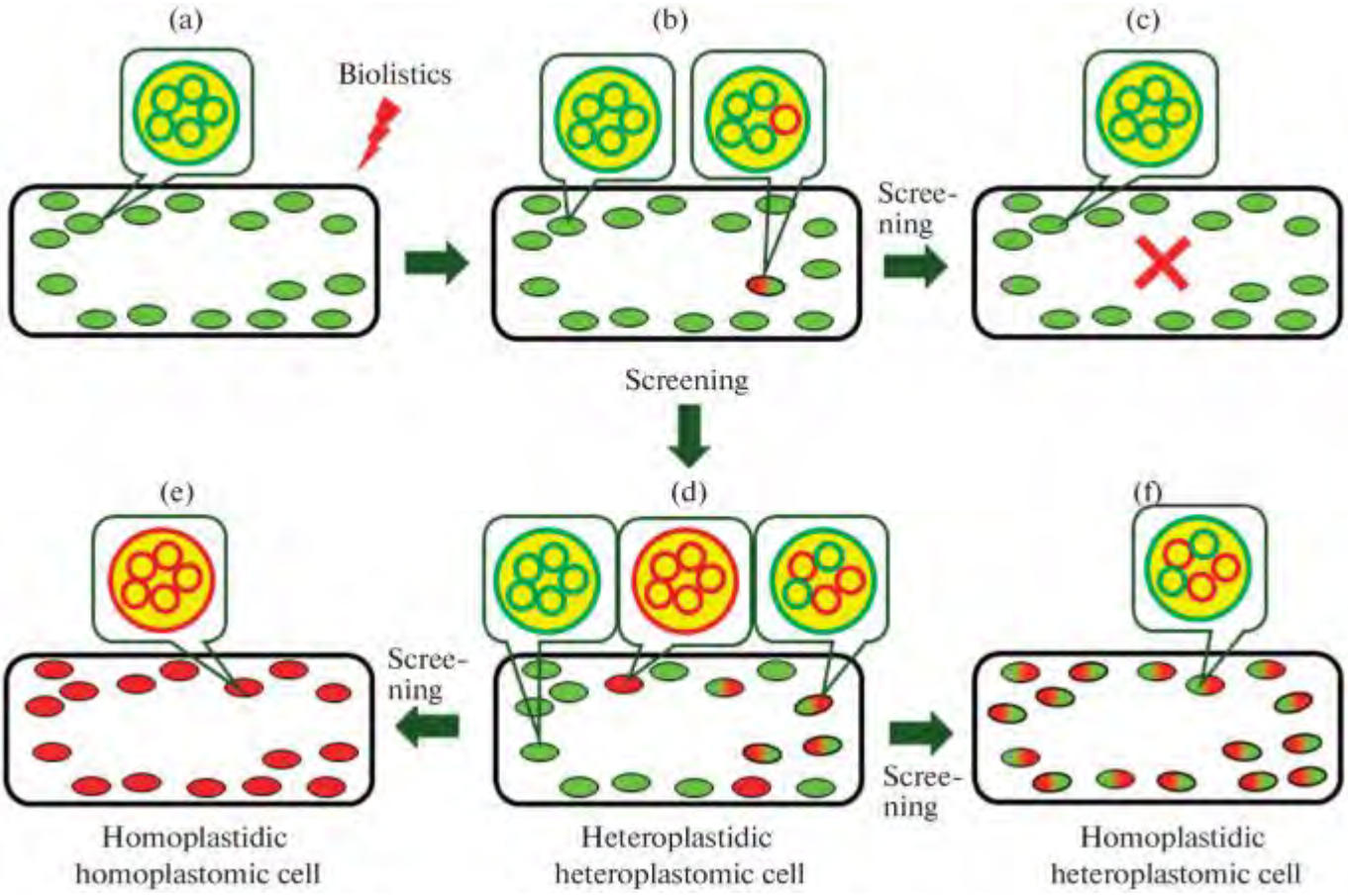
No gene silencing or positional effects.

Maternal inheritance reduces gene escape risks into wild relatives.

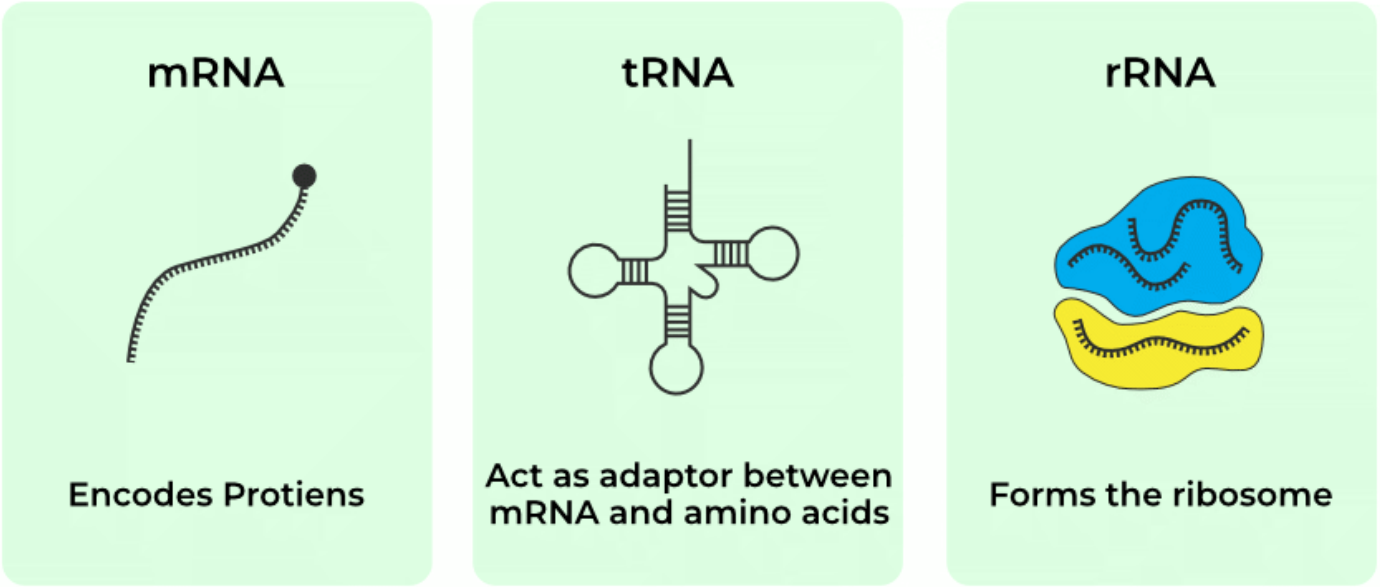
Applications in Biotechnology:

Development of **disease-resistant and stress-tolerant crops**.

Production of **pharmaceutical proteins and vaccines in plants**.



Transplastomic Plants



Introduction to rbcL Gene

What is the rbcL Gene?

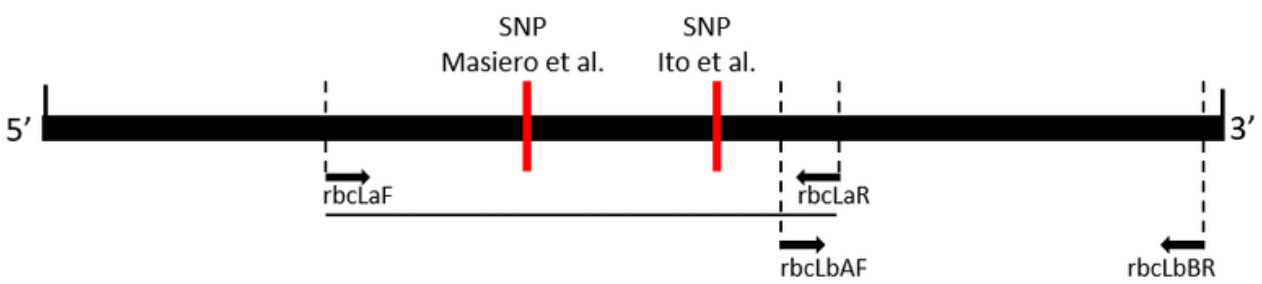
rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) is a gene found in the **chloroplast genome (cpDNA)**.
It encodes the **large subunit of RuBisCO**, the key enzyme in the **Calvin cycle**, responsible for carbon fixation.
One of the most **highly conserved genes** in plants, making it useful for phylogenetic and evolutionary studies.

Structure and Location of rbcL

Located in the **large single-copy (LSC) region** of the chloroplast genome.
Encodes a **protein of approximately 475 amino acids**.
Works in coordination with a **small subunit (rbcS), which is encoded by nuclear DNA**.

Function of rbcL in Photosynthesis

RuBisCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase) is responsible for CO_2 fixation in the Calvin cycle.
The rbcL gene encodes the **catalytic site of RuBisCO**, which binds CO_2 and initiates the process of carbon assimilation.
Plays a key role in **regulating photosynthetic efficiency and biomass production**.



Applications and Importance of rbcL

1. rbcL in Phylogenetics and DNA Barcoding

Highly conserved sequence across plant species makes it ideal for studying evolutionary relationships.

Used in **plant DNA barcoding**, along with the **matK gene**, to identify and classify plant species.

Helps in **tracing plant evolution and understanding biodiversity**.

2. rbcL in Molecular Ecology and Conservation Biology

Used to study **genetic diversity and population structure** of plant species.

Helps in **conservation genetics** by identifying endangered plant species.

Provides insights into how plants adapt to **climate change and environmental stress**.

3. rbcL in Genetic Engineering and Crop Improvement

Targeted for **RuBisCO engineering** to improve photosynthetic efficiency.

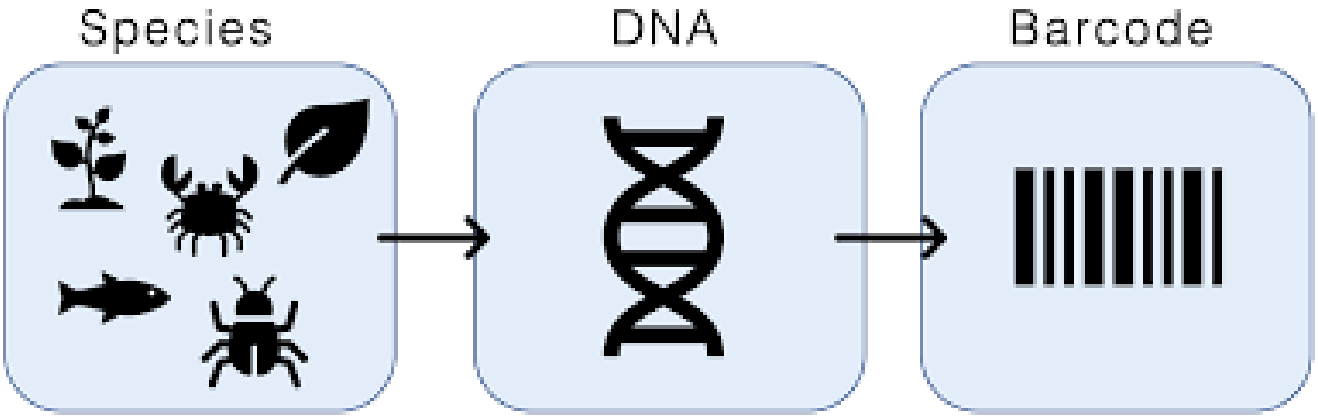
Used in **synthetic biology approaches** to enhance carbon fixation and increase crop yield.

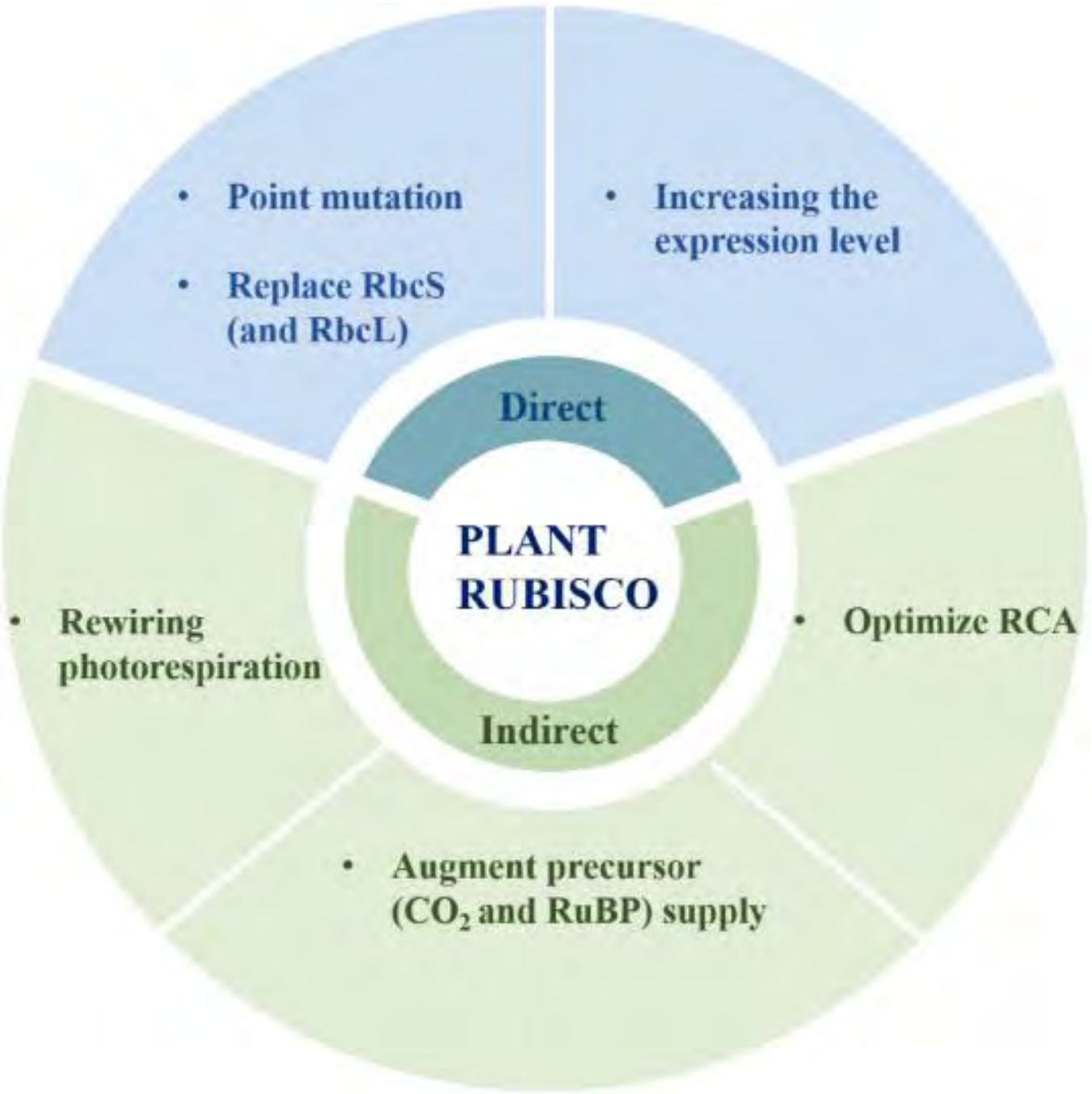
Genetic modifications of **rbcL can lead to improved drought and temperature tolerance** in plants.

4. rbcL in Paleobotany and Fossil Studies

Ancient DNA (aDNA) studies use **rbcL sequences** to identify plant remains from fossils.

Helps reconstruct **past ecosystems and climate conditions** through plant phylogeny.


Species
DNA
Barcode


• Point mutation
• Replace RbcS (and RbcL)
• Increasing the expression level
Direct
PLANT RUBISCO
Indirect
• Rewiring photorespiration
• Optimize RCA
• Augment precursor (CO_2 and RuBP) supply

Introduction to the matK Gene

What is the matK Gene?

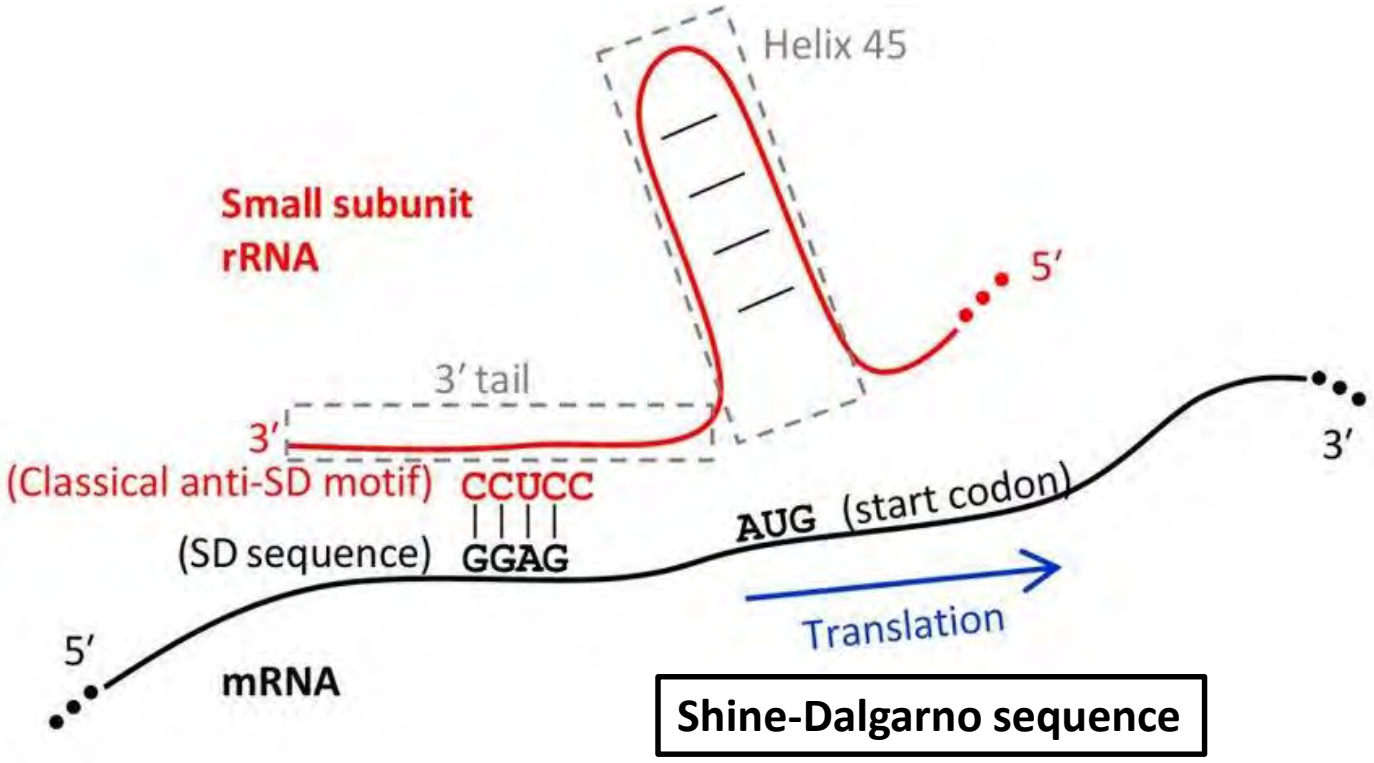
matK (maturase K) is a **protein-coding gene** located in the **chloroplast genome (cpDNA)** of plants.
It encodes a **maturase enzyme** that helps in the splicing of group II introns in chloroplast transcripts.
Found within the **intron of the trnK (tRNA-Lys) gene** in the **large single-copy (LSC) region** of cpDNA.
One of the **fastest-evolving genes in the chloroplast genome**, making it valuable for phylogenetic and evolutionary studies.

Structure and Location of matK

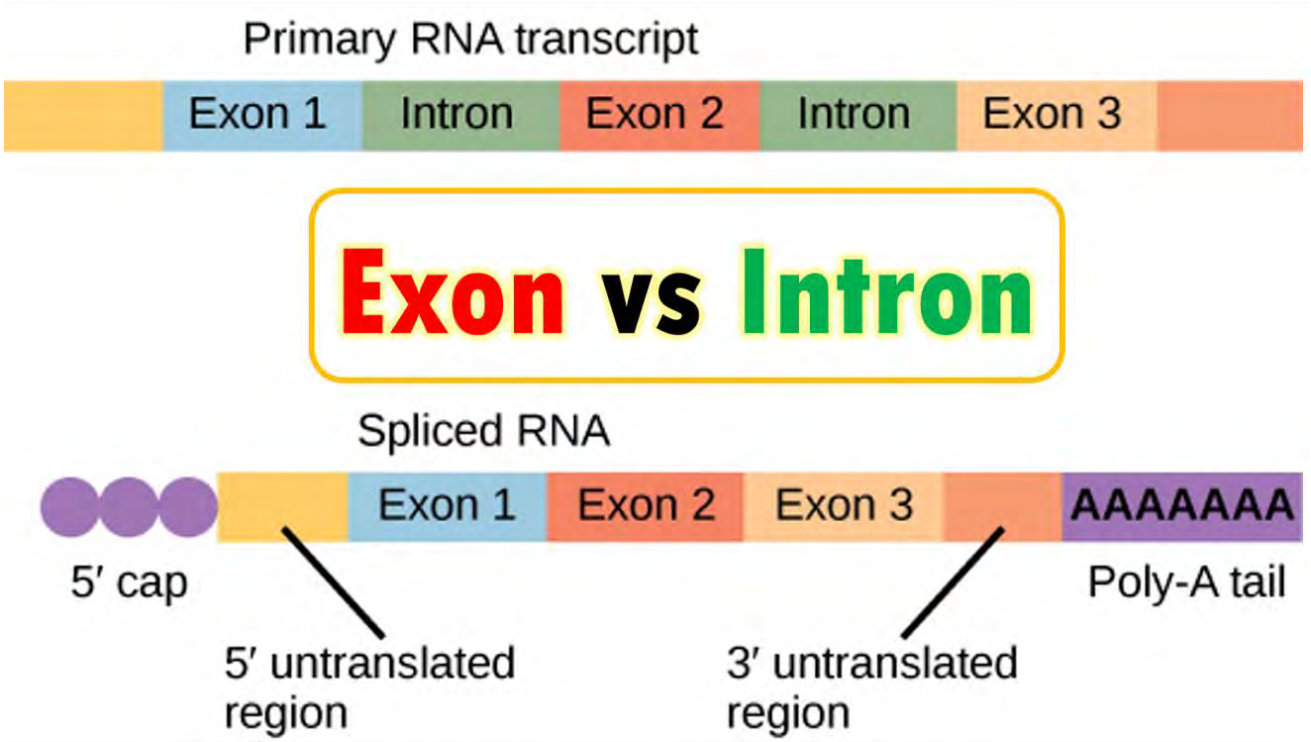
Size: Approximately **1500 base pairs (bp)**.
Function: Plays a key role in **RNA splicing and gene expression regulation** in the chloroplast.
Unlike most chloroplast genes, **matK lacks a Shine-Dalgarno sequence**, which is usually required for translation initiation.

Function of matK in Chloroplast Biology

Encodes a **maturase enzyme** that facilitates the **splicing of group II introns** in chloroplast RNA.
Helps in the proper expression of essential chloroplast genes, including those involved in **photosynthesis and energy metabolism**.
Plays a critical role in maintaining **chloroplast gene stability and function**.



Shine-Dalgarno sequence



Applications and Importance of matK

1. matK in Phylogenetics and DNA Barcoding

Due to its **rapid evolution** and **high sequence variability**, matK is widely used in **plant phylogenetic studies**.

Used in **DNA barcoding** along with **rbcL** for species identification in plants.

Helps in resolving **evolutionary relationships among angiosperms, gymnosperms, and ferns**.

2. matK in Taxonomy and Biodiversity Studies

Used to **differentiate closely related plant species** that are difficult to distinguish morphologically.

Plays a key role in **documenting plant biodiversity and species conservation**.

The Consortium for the Barcode of Life (CBOL) recognizes **matK and rbcL as standard markers** for plant DNA barcoding.

3. matK in Molecular Ecology and Conservation Biology

Useful for studying **plant population genetics and adaptation to environmental changes**.

Helps in **identifying endangered plant species** and developing conservation strategies.

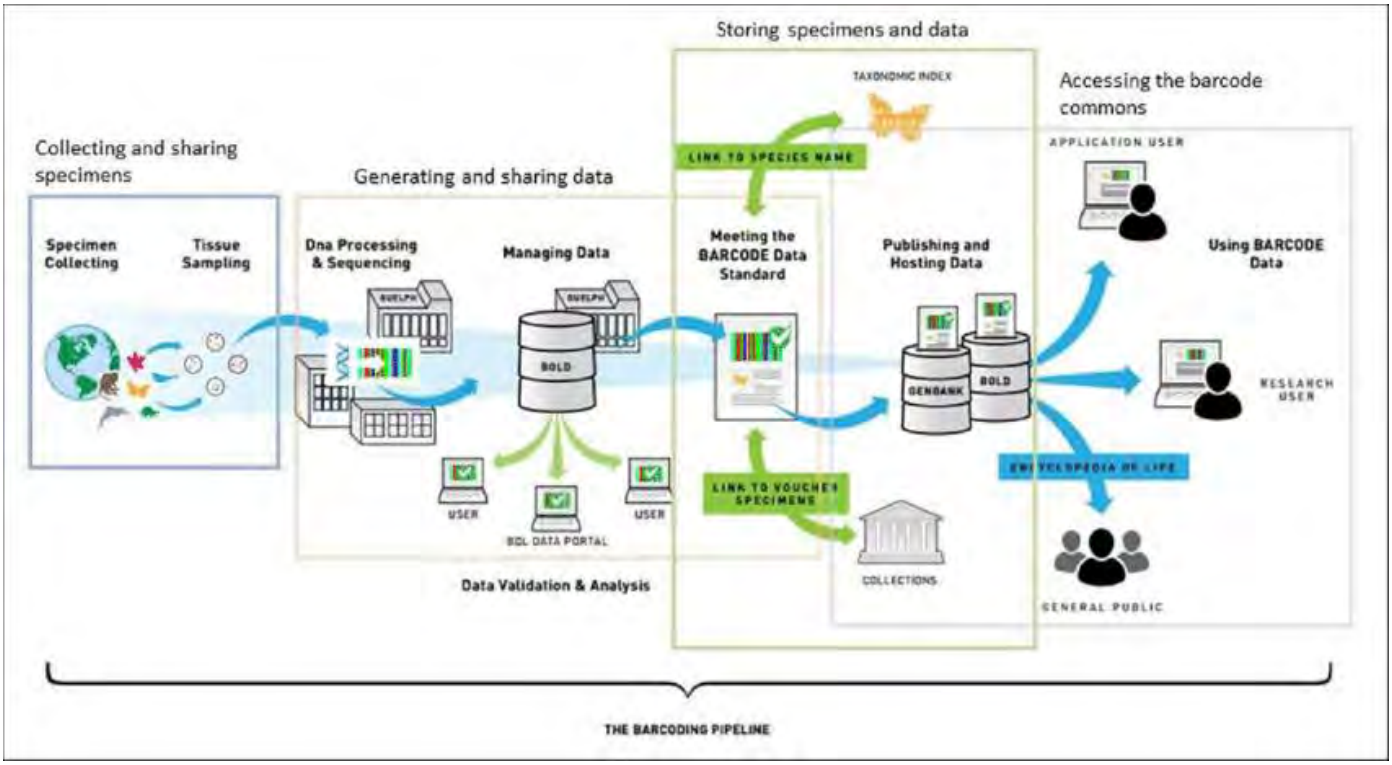
Applied in **forensic botany** to track illegal plant trade and biodiversity loss.

4. matK in Evolutionary and Comparative Genomics

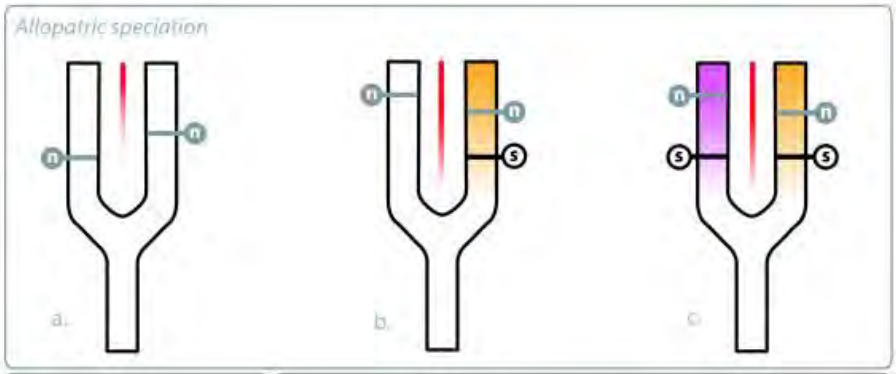
Used in **phylogenomic studies** to understand the evolution of plant genomes.

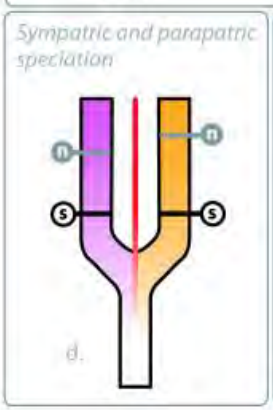
Helps trace **gene duplication, divergence, and functional adaptation** in plant lineages.

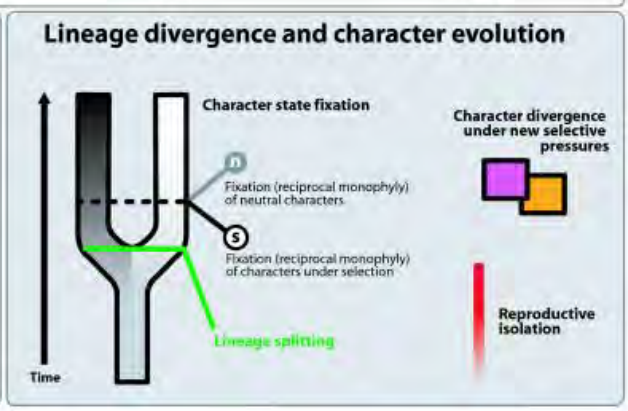
Comparative analyses of **matK sequences provide insights into the genetic basis of plant diversity**.


Storing specimens and data
TAXONOMIC INDEX
LINK TO SPECIES NAME
Accessing the barcode commons
APPLICATION USER
Collecting and sharing specimens
Generating and sharing data
Specimen Collecting
Tissue Sampling
Dna Processing & Sequencing
Managing Data
Meeting the BARCODE Data Standard
Publishing and Hosting Data
Using BARCODE Data
GENBANK
BOLD
RESEARCH USER
USER
USER
BOL DATA PORTAL
LINK TO VOUCHER SPECIMENS
COLLECTIONS
GENERAL PUBLIC
Data Validation & Analysis
THE BARCODING PIPELINE


A
Gene Duplication
A
A'
Gain of Function
Original Gene
Original Gene
Duplicated Gene
Loss of Function


Allopatric speciation
a.
b.
c.


Sympatric and parapatric speciation
d.


Lineage divergence and character evolution
Character state fixation
Character divergence under new selective pressures
Fixation (reciprocal monophyly) of neutral characters
Fixation (reciprocal monophyly) of characters under selection
Lineage splitting
Reproductive isolation
Time

Introduction to the ndhF Gene

What is the ndhF Gene?

ndhF (NADH dehydrogenase subunit F) is a **protein-coding gene** found in the **chloroplast genome (cpDNA)** of plants.

It encodes a **subunit of NADH dehydrogenase (NDH) complex**, which is involved in the **chlororespiratory electron transport chain**.

Located in the **large single-copy (LSC) region** of the chloroplast genome.

Plays an essential role in **cyclic electron transport around Photosystem I (PSI)**, helping in energy conservation.

Structure and Location of ndhF

Size: Varies between **2,100 – 2,300 base pairs (bp)** in most plants.

Part of the **NADH dehydrogenase (NDH) complex**, which is homologous to Complex I in mitochondria.

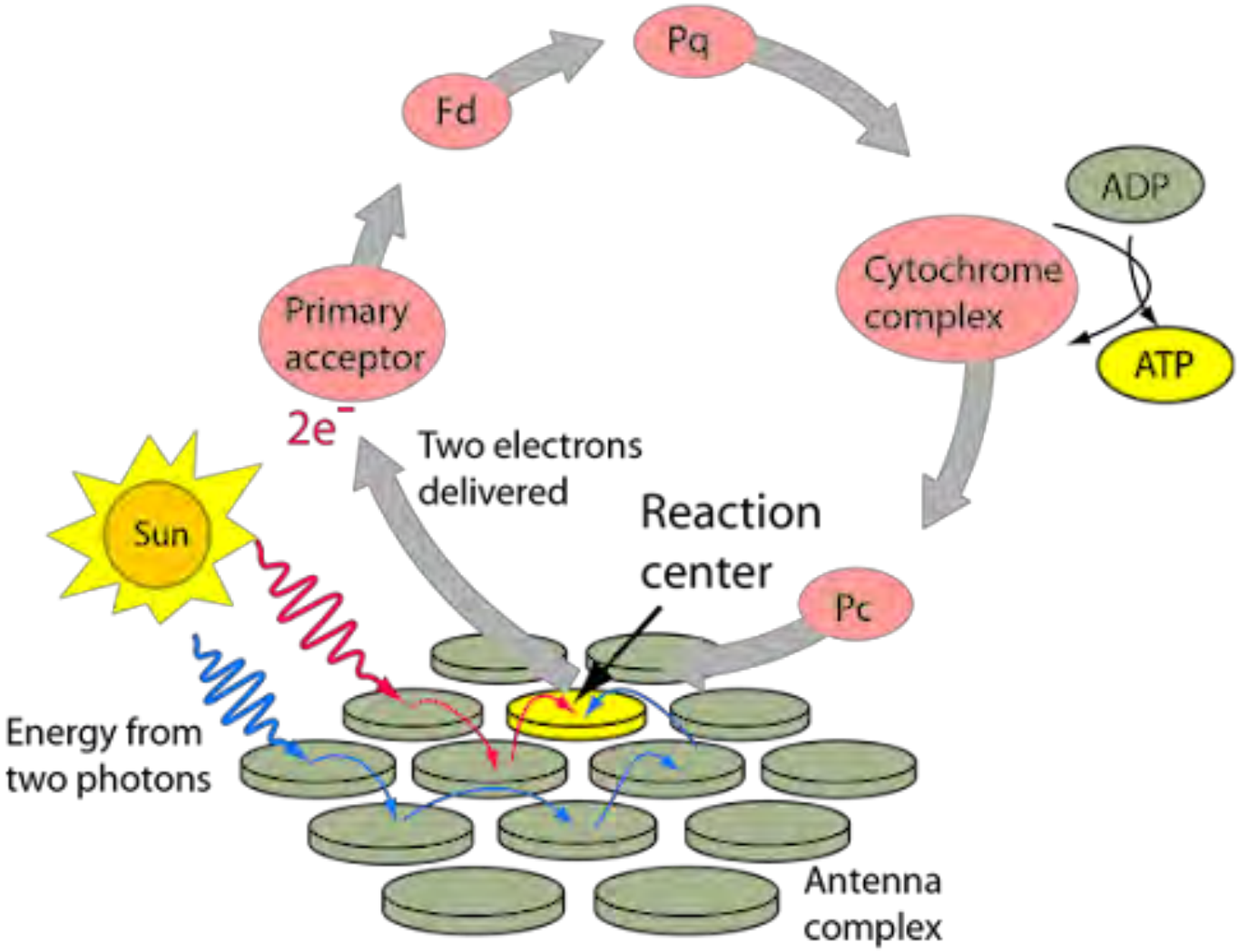
Works in coordination with other **NDH subunits (ndhA-ndhK)** for proper electron transport function.

Function of ndhF in Photosynthesis and Energy Metabolism

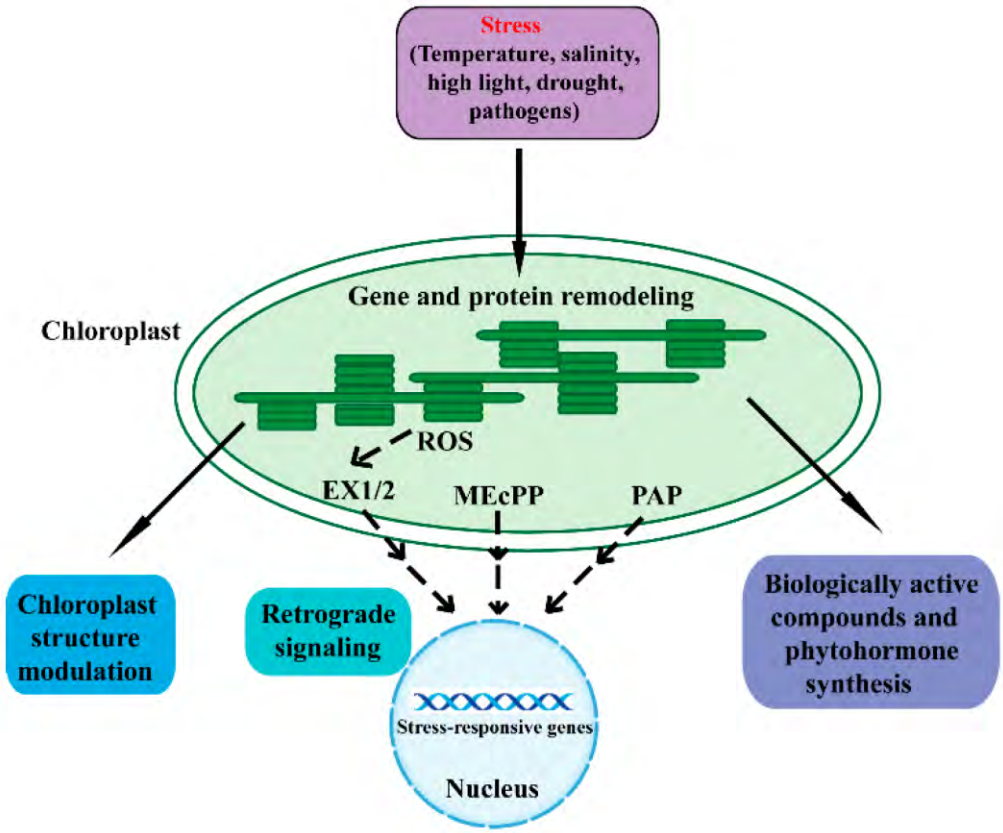
Involved in **cyclic electron flow (CEF) around Photosystem I**, which enhances ATP production without producing NADPH.

Helps plants **adapt to fluctuating light conditions** by regulating electron transport efficiency.

Plays a role in **photoprotection**, protecting chloroplasts from oxidative stress during extreme light exposure.



Photosystem I



Applications and Importance of ndhF

1. ndhF in Phylogenetics and Evolutionary Studies

The **ndhF gene evolves at a moderate rate**, making it useful for **studying deep evolutionary relationships** in plants.

Used in **angiosperm phylogenetics**, especially in resolving relationships among complex plant families.

Frequently used alongside **rbcL, matK, and trnL-F** sequences for **comprehensive plant phylogenies**.

2. ndhF in Taxonomy and Biodiversity Research

Helps in distinguishing between **closely related plant species** based on molecular sequence variation.

Widely used in **identifying and classifying flowering plants (angiosperms)**.

Provides insights into the **evolutionary adaptation of plants to different ecological niches**.

3. ndhF in Plant Physiology and Stress Adaptation

Plants with **defective ndhF genes show impaired cyclic electron flow**, affecting ATP production.

Involved in **regulating plant responses to drought, high light, and temperature stress**.

Mutations in **ndhF and related NDH genes** can lead to **reduced photosynthetic efficiency and growth defects**.

4. ndhF in Comparative Genomics and Crop Improvement

Used in **comparative chloroplast genomics** to understand evolutionary divergence among plant species.

Genetic modification of **ndhF and NDH complex components** is being explored to **enhance photosynthetic efficiency**.

Potential applications in **crop improvement and climate-resilient agriculture**.

Introduction to the atpB Gene

What is the atpB Gene?

atpB (ATP synthase β-subunit gene) is a **protein-coding gene** located in the **chloroplast genome (cpDNA)** of plants.

Encodes the **β-subunit of ATP synthase (CF1-β)**, which is a key enzyme in **photosynthetic ATP production**.

Found in the **large single-copy (LSC) region** of the chloroplast genome.

Essential for the **conversion of light energy into chemical energy (ATP) during photosynthesis**.

Structure and Location of atpB

Size: Approximately **1,500–1,700 base pairs (bp)** in most plants.

Located next to **atpE**, which encodes the **ε-subunit of ATP synthase**.

Part of the **chloroplast ATP synthase complex (CF0-CF1 complex)**, which is homologous to mitochondrial ATP synthase.

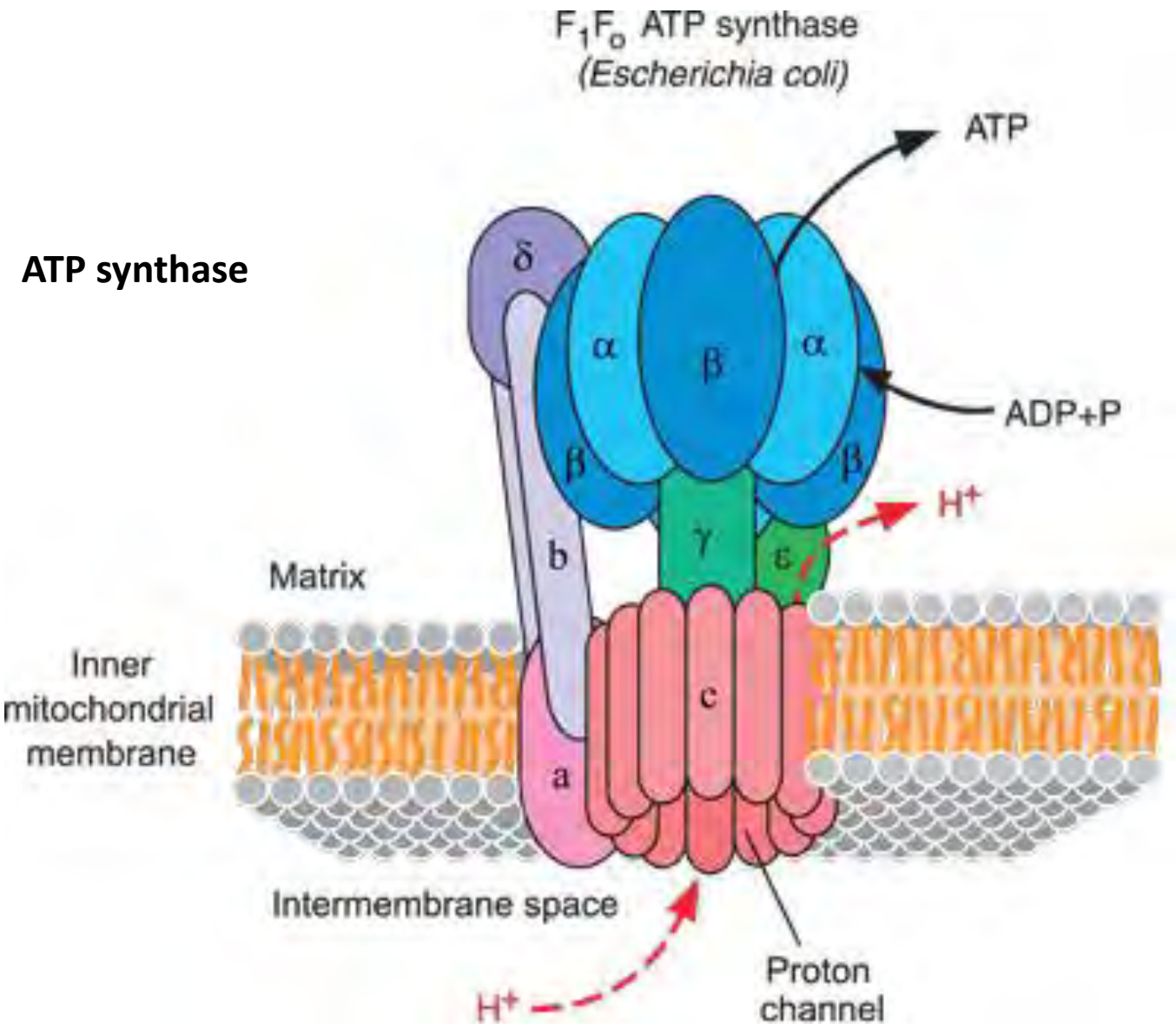
Function of atpB in Photosynthesis

The **β-subunit of ATP synthase** is responsible for **catalyzing ATP synthesis** in chloroplasts.

Works in coordination with **other subunits (atpA, atpE, atpF, atpH, atpI)** to form the **ATP synthase complex**.

Plays a crucial role in **proton-driven ATP production**, where protons (H^+) flow through the enzyme to generate ATP from ADP and Pi.

ATP synthase



Applications and Importance of atpB

1. atpB in Phylogenetics and Evolutionary Studies

The **atpB gene evolves at a moderate rate**, making it valuable for **plant phylogenetics and evolutionary research**.

Often used alongside **rbcL, matK, and ndhF** in molecular systematics to **study plant relationships**.

Helps resolve evolutionary relationships in **angiosperms, gymnosperms, and ferns**.

2. atpB in Plant Taxonomy and Biodiversity Studies

Used in **molecular classification** of plants, aiding in the identification of species and genera.

Provides insights into the **evolutionary divergence of different plant lineages**.

Helps in the study of **hybridization and speciation** events in plants.

3. atpB in Photosynthetic Efficiency and Crop Improvement

Mutations in **atpB can disrupt ATP synthesis**, leading to **reduced photosynthetic efficiency**.

Research on **atpB and ATP synthase modifications** aims to improve **energy efficiency in crops**.

Potential applications in **enhancing crop productivity under varying environmental conditions**.

4. atpB in Comparative Genomics and Functional Studies

Used in **comparative chloroplast genomics** to analyze structural variations in plant genomes.

Helps in understanding the **functional adaptation of ATP synthase across different plant species**.

Genetic studies on **atpB mutations** provide insights into **photosynthetic disorders and plant development**.

Introduction to the rpl16 Gene

What is the rpl16 Gene?

rpl16 (Ribosomal Protein L16 gene) is a **protein-coding gene** found in the **chloroplast genome (cpDNA)** of plants.

It encodes the **L16 ribosomal protein**, which is a component of the **50S large ribosomal subunit** in chloroplasts.

Plays a crucial role in **chloroplast protein synthesis** by participating in **ribosome assembly and translation**.

Structure and Location of rpl16

Size: Varies between **450 – 600 base pairs (bp)** in most plants.

Located in the **large single-copy (LSC) region** of the chloroplast genome.

Often contains an **intron**, making it a useful marker for studying plant evolution and genetic variation.

Function of rpl16 in Chloroplast Protein Synthesis

Essential for **ribosomal stability and function**, aiding in the assembly of the **chloroplast ribosome**.

Plays a role in the **translation of chloroplast-encoded proteins**, including those involved in **photosynthesis**.

Mutations or loss of function in **rpl16** can lead to **defective ribosome assembly and impaired chloroplast function**.

Applications and Importance of rpl16

1. rpl16 in Phylogenetics and Evolutionary Studies

Moderately conserved but contains an intron, making it useful for **plant phylogenetics**.

Helps in **resolving relationships among angiosperms, gymnosperms, and lower plants**.

Often used alongside other **ribosomal genes (rpl2, rps4, rps16) and plastid genes (matK, ndhF, rbcL)** for **molecular systematics**.

2. rpl16 in Plant Taxonomy and Genetic Diversity Studies

The **presence and variation of the intron** in rpl16 can be used for **species identification and classification**.

Useful for studying **population genetics and evolutionary divergence** among plant lineages.

Helps differentiate between **closely related species and subspecies** based on **sequence variation**.

3. rpl16 in Chloroplast Genome Evolution and RNA Processing

Introns in **rpl16** provide insights into the **evolution of plastid genomes** in different plant groups.

Studying **intron retention and splicing efficiency** helps in understanding **chloroplast RNA processing mechanisms**.

Structural variations in rpl16 introns can indicate **adaptive evolution in response to environmental changes**.

4. rpl16 in Functional Genomics and Biotechnology

Mutations in rpl16 can be linked to **chloroplast ribosome defects**, affecting **photosynthesis and plant growth**.

Potential use in **genetic engineering** to enhance **chloroplast translation efficiency and protein synthesis**.

Can serve as a **target for studying ribosome biogenesis and gene regulation** in plant chloroplasts.

Introduction to the trnH-psbA Intergenic Spacer

What is the trnH-psbA Intergenic Spacer?

The **trnH-psbA intergenic spacer** is a **non-coding region** of the **chloroplast genome (cpDNA)** found between:

- **trnH (tRNA-Histidine) gene**, which encodes a transfer RNA.
- **psbA (Photosystem II protein D1) gene**, which is involved in **photosynthesis**.

This spacer region is widely used in **plant molecular studies** due to its **high sequence variation**.

Unlike protein-coding genes, intergenic spacers evolve **more rapidly**, making them **useful for DNA barcoding and phylogenetics**.

Structure and Characteristics of trnH-psbA

Length: Ranges from **200 – 600 base pairs (bp)** depending on plant species.

Location: Found in the **large single-copy (LSC) region** of the chloroplast genome.

Sequence variation: Exhibits **insertions, deletions, and mutations**, useful for **species differentiation**.

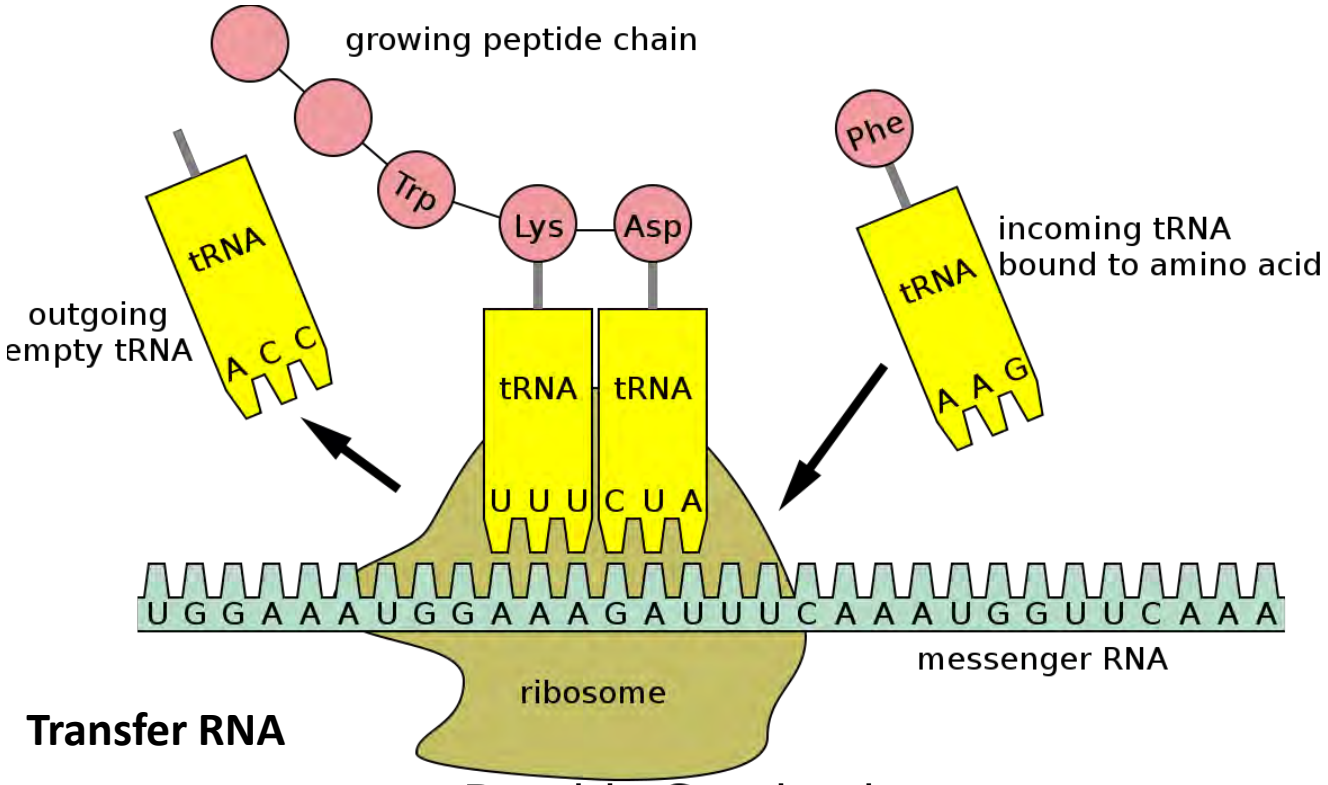
Non-coding but transcribed: May play a role in **chloroplast RNA processing and regulation**.

Function and Significance in Chloroplast Biology

Though **non-coding**, it may contribute to **gene expression regulation** in chloroplasts.

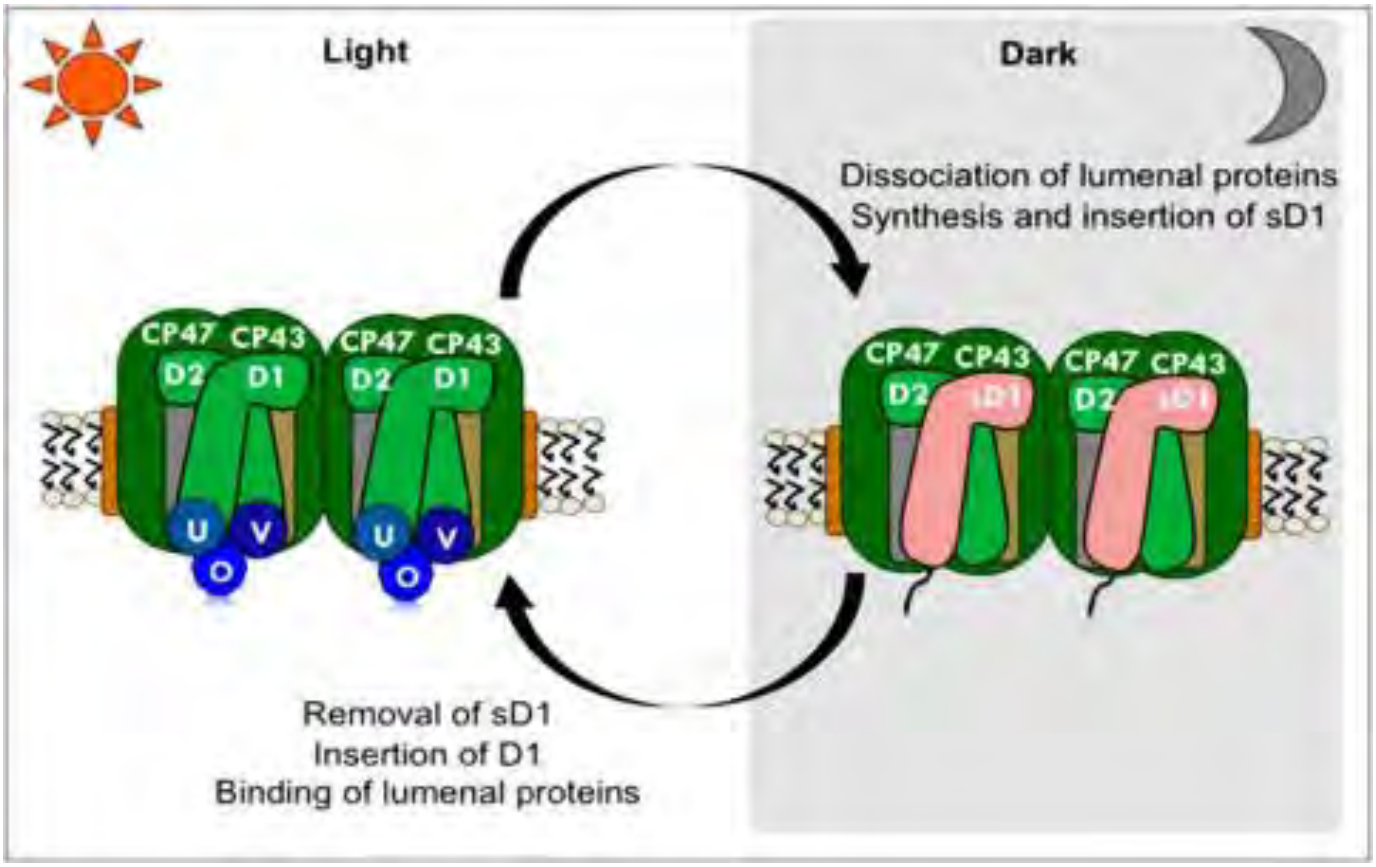
Serves as a **spacer between functional genes**, allowing proper transcription and translation.

Used extensively in **plant molecular systematics and evolutionary studies** due to its variability.



Transfer RNA

Peptide Synthesis



psbA (Photosystem II protein D1) gene

Applications and Importance of trnH-psbA

1. trnH-psbA in DNA Barcoding and Species Identification

One of the most **variable plastid regions**, making it useful for **species discrimination**.

Used in combination with other barcode markers (**rbcL, matK, ITS**) for plant identification.

Recommended by the **Consortium for the Barcode of Life (CBOL)** for plant DNA barcoding.

2. trnH-psbA in Phylogenetics and Evolutionary Studies

Due to its **high mutation rate**, it helps resolve **closely related species and subspecies**.

Used to study **plant lineage diversification and evolutionary relationships**.

Frequently applied in **angiosperms, ferns, and bryophytes** for **phylogenetic reconstruction**.

3. trnH-psbA in Plant Taxonomy and Genetic Diversity Studies

Provides insights into **genetic variation within and between species**.

Useful for **identifying hybrid species and detecting introgression**.

Variability in sequence length and GC content can reflect **adaptive evolution** in plants.

4. trnH-psbA in Conservation Genetics and Ecological Research

Helps in studying **endangered species and conservation priorities**.

Used to analyze **plant population structure and gene flow in different environments**.

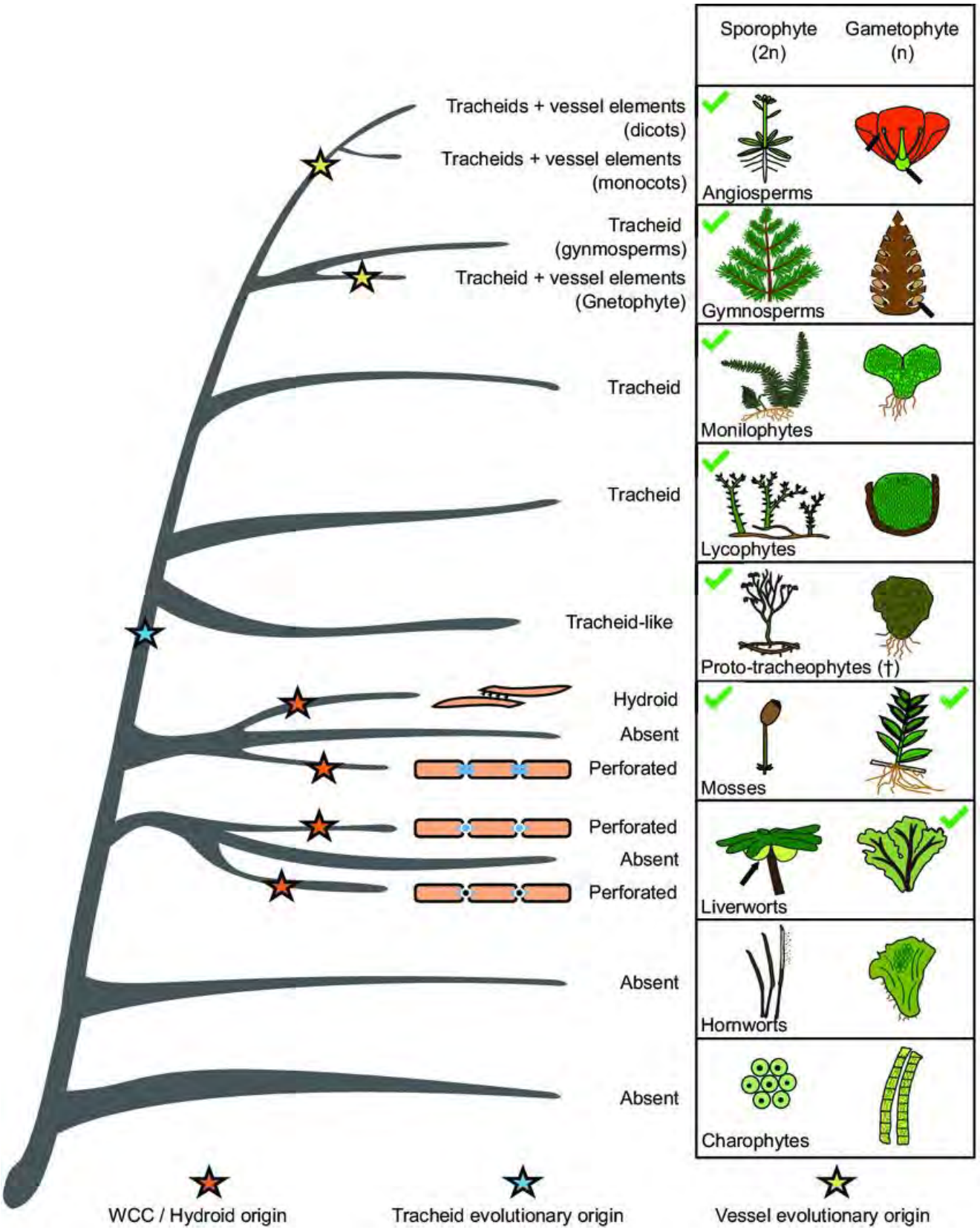
Assists in identifying **cryptic species and plant phylogeography**.

5. trnH-psbA in Comparative Chloroplast Genomics

Used to compare **plastid genome evolution** across different plant taxa.

Structural changes (insertions, deletions, rearrangements) help in **genome evolution studies**.

Provides information on **chloroplast genome stability and mutation hotspots**.



plant lineage diversification

Introduction to the trnL-trnF Region

What is the trnL-trnF Region?

The **trnL-trnF region** is a **non-coding segment** in the **chloroplast genome (cpDNA)** located between:

trnL (tRNA-Leucine) gene, which contains an **intron**.

trnF (tRNA-Phenylalanine) gene, which is an **exon**.

This region includes two distinct parts:

trnL Intron – A non-coding sequence within the **trnL gene**.

trnL-trnF Intergenic Spacer – A non-coding sequence **between trnL and trnF genes**.

Structure and Characteristics of trnL-trnF

Size: Varies between **250 – 1,000 base pairs (bp)** depending on species.

Location: Found in the **large single-copy (LSC) region** of the chloroplast genome.

Sequence variation: Due to **indels (insertions/deletions), mutations, and rearrangements**, it is widely used in evolutionary studies.

Highly conserved in plants, but non-coding, allowing for **higher rates of mutation** compared to protein-coding regions.

Function and Biological Role

While **non-coding**, the **trnL intron is essential for proper tRNA processing** and splicing.

The **intergenic spacer plays a regulatory role** in chloroplast gene expression.

It serves as an **important genetic marker for plant molecular studies**.

Applications and Importance of trnL-trnF

1. trnL-trnF in Phylogenetics and Evolutionary Studies

Used to **determine relationships among plant species, genera, and families**.

Helps in reconstructing **plant evolutionary histories and divergence times**.

Particularly useful for studying **angiosperms, gymnosperms, and ferns**.

2. trnL-trnF as a DNA Barcoding Marker

Due to **high sequence variation**, it is used for **species identification and classification**.

Works well for **differentiating closely related plant species**.

Often used with **rbcL, matK, and trnH-psbA** in plant barcoding studies.

3. trnL-trnF in Population Genetics and Biodiversity Studies

Assists in analyzing **genetic variation within and between populations**.

Useful for studying **gene flow, genetic drift, and plant adaptation**.

Applied in identifying **hybridization events and cryptic species**.

4. trnL-trnF in Conservation Genetics and Ecology

Used to study **endangered plant species and their genetic diversity**.

Helps in **understanding plant dispersal, migration patterns, and habitat adaptation**.

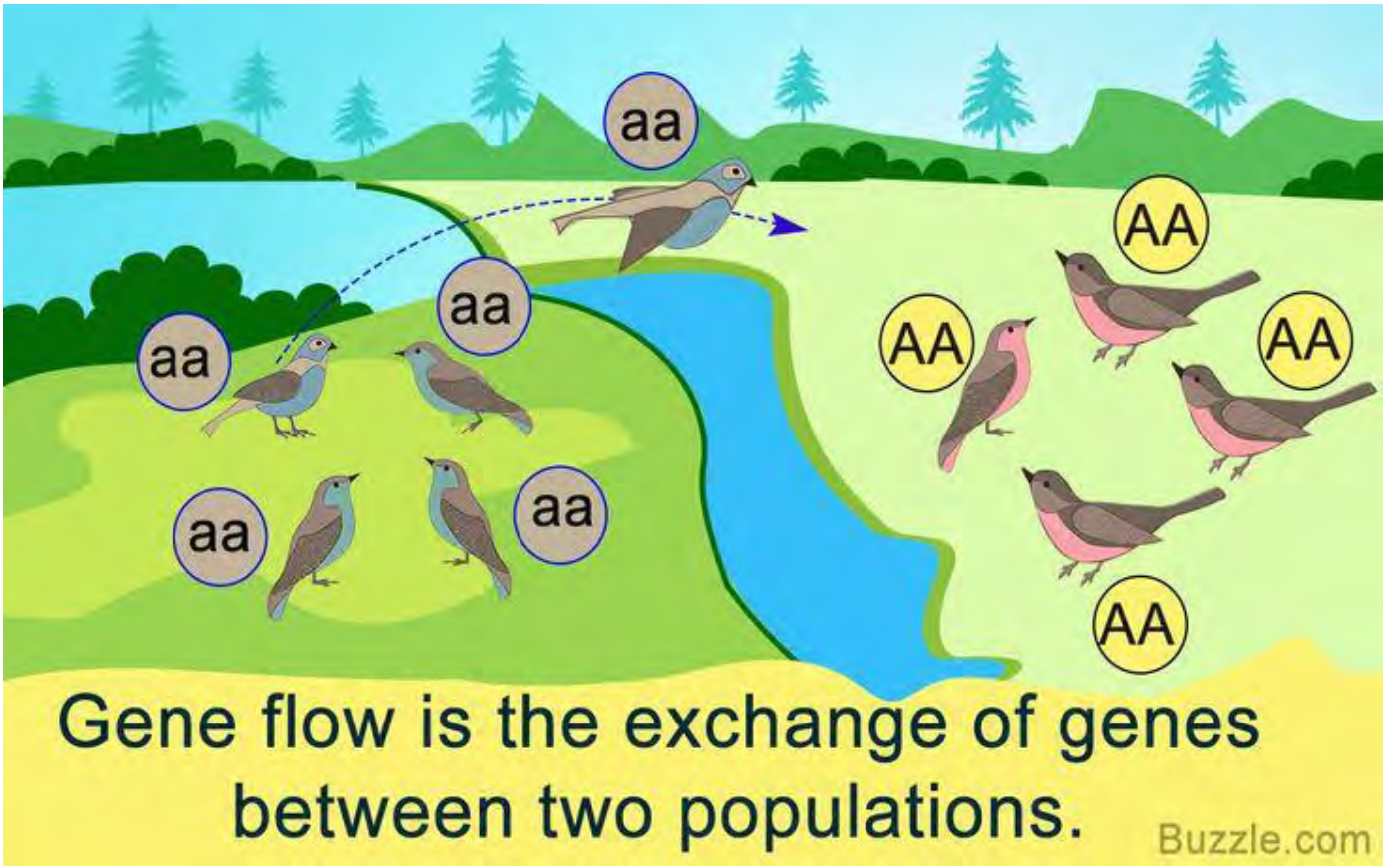
Commonly used in **paleoecological studies** to reconstruct **ancient plant communities**.

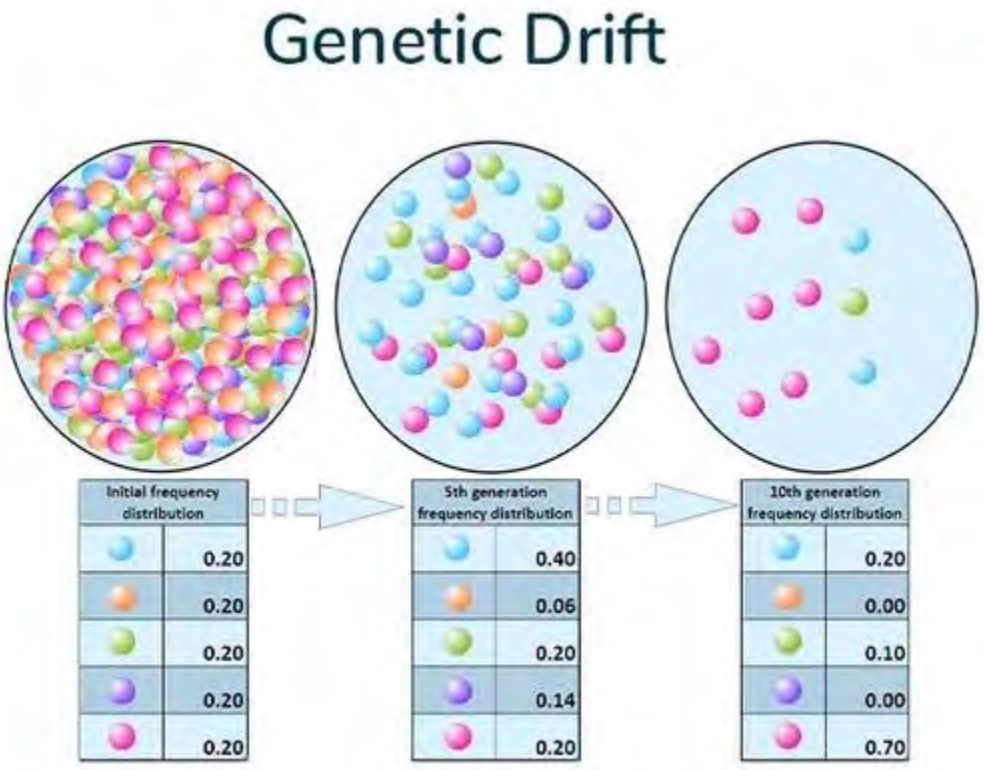
5. trnL-trnF in Comparative Chloroplast Genomics

The presence of **insertions, deletions, and rearrangements** makes it valuable for **studying chloroplast genome evolution**.

Helps detect **mutational hotspots and plastid genome stability**.

Provides insights into **chloroplast RNA processing and intron splicing mechanisms**.


aa
aa
aa
aa
aa
AA
AA
AA
AA
Gene flow is the exchange of genes between two populations.
Buzzle.com


Genetic Drift
Initial frequency distribution
0.20
0.20
0.20
0.20
0.20
5th generation frequency distribution
0.40
0.06
0.20
0.14
0.20
10th generation frequency distribution
0.20
0.00
0.10
0.00
0.70

Advantages of Chloroplast DNA (cpDNA) for Plant Systematics

Maternal Inheritance

In most angiosperms, cpDNA is **maternally inherited**, reducing genetic recombination.

This makes it easier to track **lineage relationships** without the complexities of biparental inheritance.

Highly Conserved Structure

The **gene content and organization** of cpDNA remain largely unchanged across plant species.

This allows for **comparative studies** across wide evolutionary distances.

Moderate Mutation Rate

cpDNA evolves more **slowly than nuclear DNA**, providing a stable molecular clock.

It is **faster than rDNA**, making it useful for **resolving species-level relationships**.

Presence of Both Coding and Non-Coding Regions

Coding regions (**rbcL, matK, ndhF**) are highly conserved and useful for **deep phylogenetic studies**.

Non-coding regions (**trnH-psbA, trnL-trnF**) show **high variability**, making them useful for **species-level discrimination**.

Circular and Non-Recombinant Genome

cpDNA is a **circular molecule**, reducing the risk of **chromosomal rearrangements** found in nuclear DNA.

Lack of recombination helps in maintaining **consistent phylogenetic signals**.

Uniparental Transmission in Hybrids

In hybrid plants, cpDNA is inherited from **only one parent**, allowing for clear **maternal lineage tracing**.

This helps in identifying **introgression events and hybrid speciation**.

High Copy Number in Cells

Each plant cell contains **multiple copies** of cpDNA, making it easy to extract and analyze.

cpDNA can be retrieved from **degraded or ancient samples**, useful in **paleobotany and herbarium studies**.

Universal and Well-Characterized Regions

Many cpDNA markers (**rbcL, matK, trnL-trnF, trnH-psbA**) are **standardized for phylogenetic analysis**.

These markers have been extensively used in **DNA barcoding and plant classification**.

Useful for Biogeographical and Evolutionary Studies

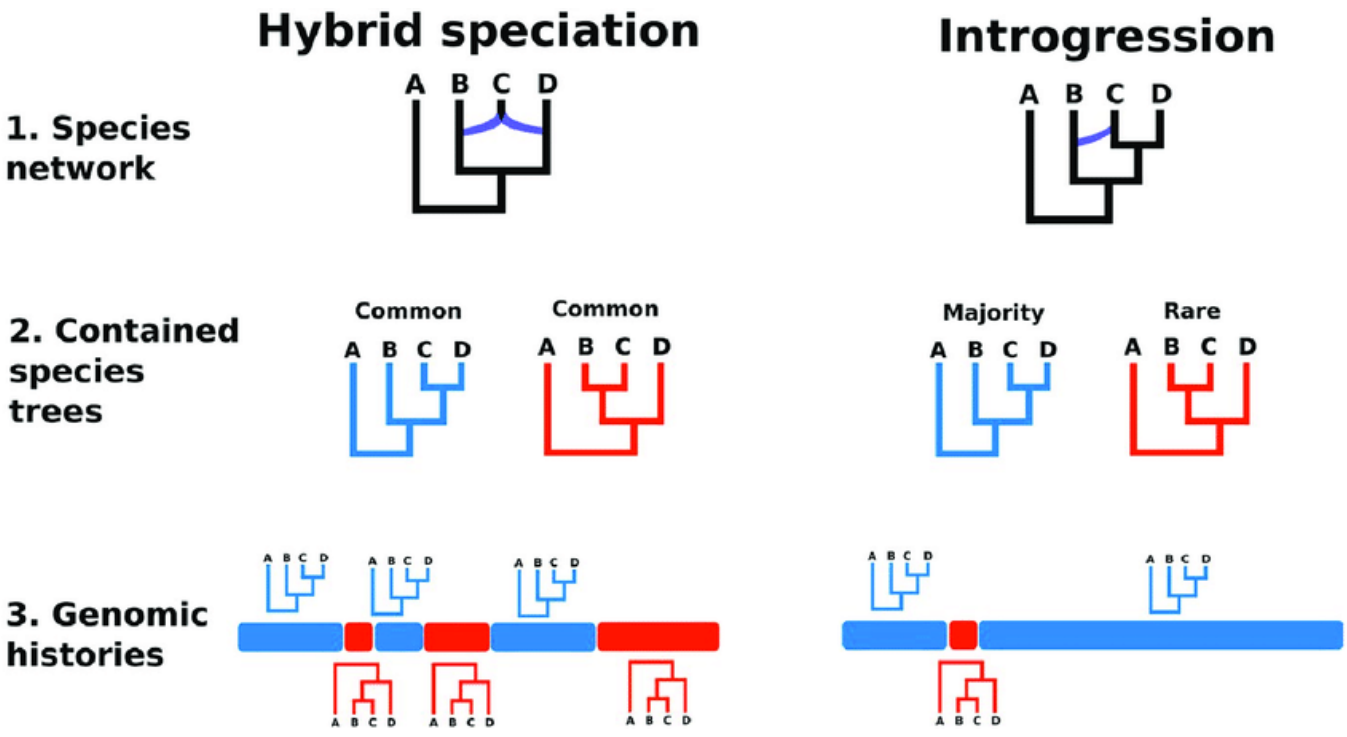
cpDNA markers help in tracing **plant migration patterns** and **historical biogeography**.

They assist in understanding **chloroplast evolution and plant adaptation to different environments**.

Wide Application in Conservation and Ecology

cpDNA is used for **assessing genetic diversity in endangered species**.

It helps in understanding **plant population structure, speciation, and ecosystem dynamics**.



Mitochondrial Genome in Plants

1. Introduction to the Mitochondrial Genome (mtDNA)

The **mitochondrial genome** is a circular or linear DNA molecule located within the **mitochondria** of plant cells.

It plays a crucial role in **cellular respiration** and **energy production** through oxidative phosphorylation.

Unlike chloroplast and nuclear genomes, plant **mtDNA exhibits a highly variable size and structure**.

2. Unique Characteristics of Plant mtDNA

Large and Variable Size: Ranges from **200 kb to 11.3 Mb**, much larger than animal mtDNA.

Low Mutation Rate: Evolves more slowly than nuclear DNA but can undergo **extensive rearrangements**.

High Recombination: Contains **repetitive sequences** that promote frequent recombination events.

Cytoplasmic Inheritance: Inherited **maternally** in most plants (though exceptions exist).

Introns and Foreign DNA: Contains **self-splicing introns** and sequences acquired from nuclear and chloroplast genomes.

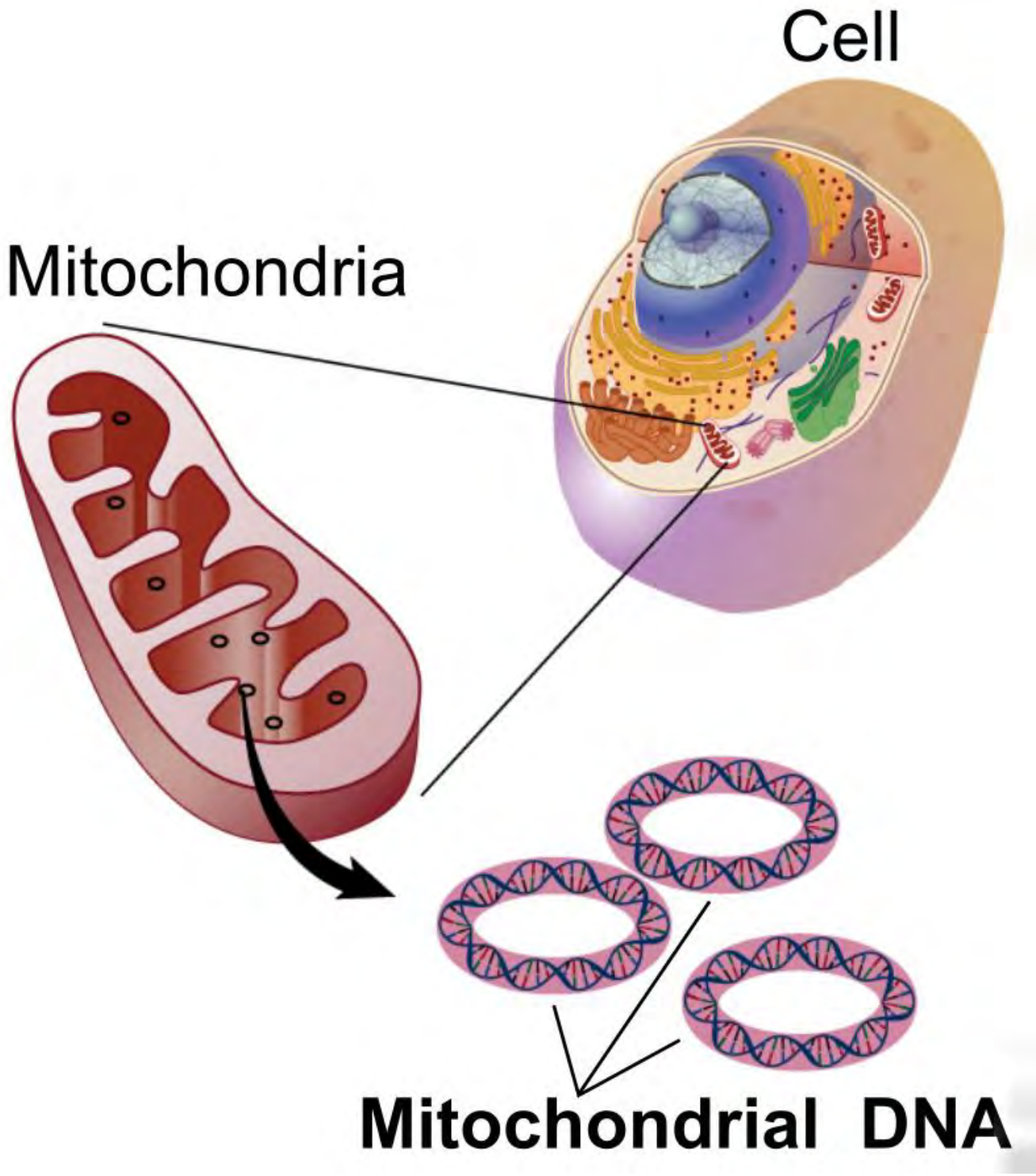
3. Structure and Organization of Plant mtDNA

Unlike the small and stable animal mtDNA, plant mtDNA is often **complex, multipartite, and highly recombinogenic**.

Contains **essential genes for oxidative phosphorylation**, ribosomal RNAs (**rRNA**), and transfer RNAs (**tRNA**).

Includes **non-coding sequences** that contribute to genome expansion.

Exhibits **RNA editing**, where specific bases are modified post-transcriptionally.


Cell
Mitochondria
Mitochondrial DNA

4. Functional Genes in Plant mtDNA

Genes for Electron Transport and ATP Synthesis:

- **nad (NADH dehydrogenase complex subunits)**
- **cob (cytochrome b)**
- **cox (cytochrome c oxidase subunits)**
- **atp (ATP synthase subunits)**

Genes for rRNA and tRNA: Required for mitochondrial protein synthesis.

Chimeric and Recombination-Associated Open Reading Frames (ORFs): Often linked to cytoplasmic male sterility (**CMS**) in plants.

5. Role of Plant mtDNA in Cytoplasmic Male Sterility (CMS)

Cytoplasmic male sterility is a **mitochondrially controlled trait** that prevents pollen formation.

Caused by **rearrangements or mutations** in mitochondrial genes.

Exploited in **hybrid crop breeding** to produce seedless or high-yielding varieties.

6. Evolution and Horizontal Gene Transfer in mtDNA

Plant mtDNA has undergone **horizontal gene transfer** from nuclear and chloroplast genomes.

Comparative genomics reveals **extensive rearrangements and gene duplications**.

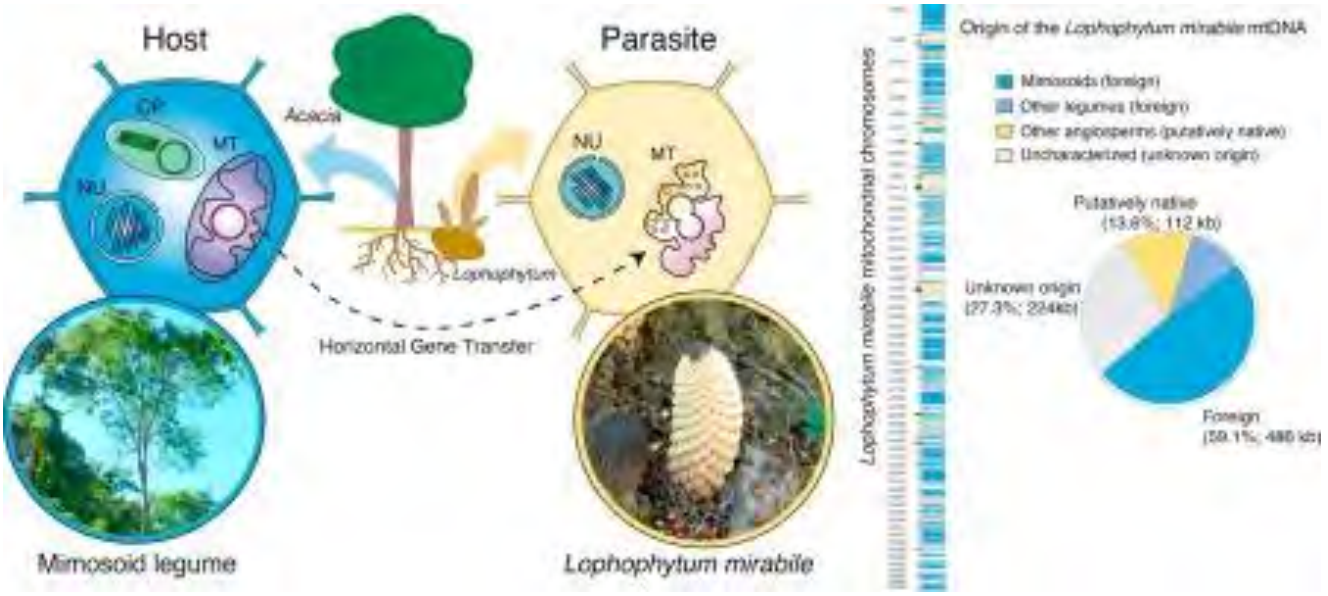
Plays a role in **plant adaptation and evolution**.

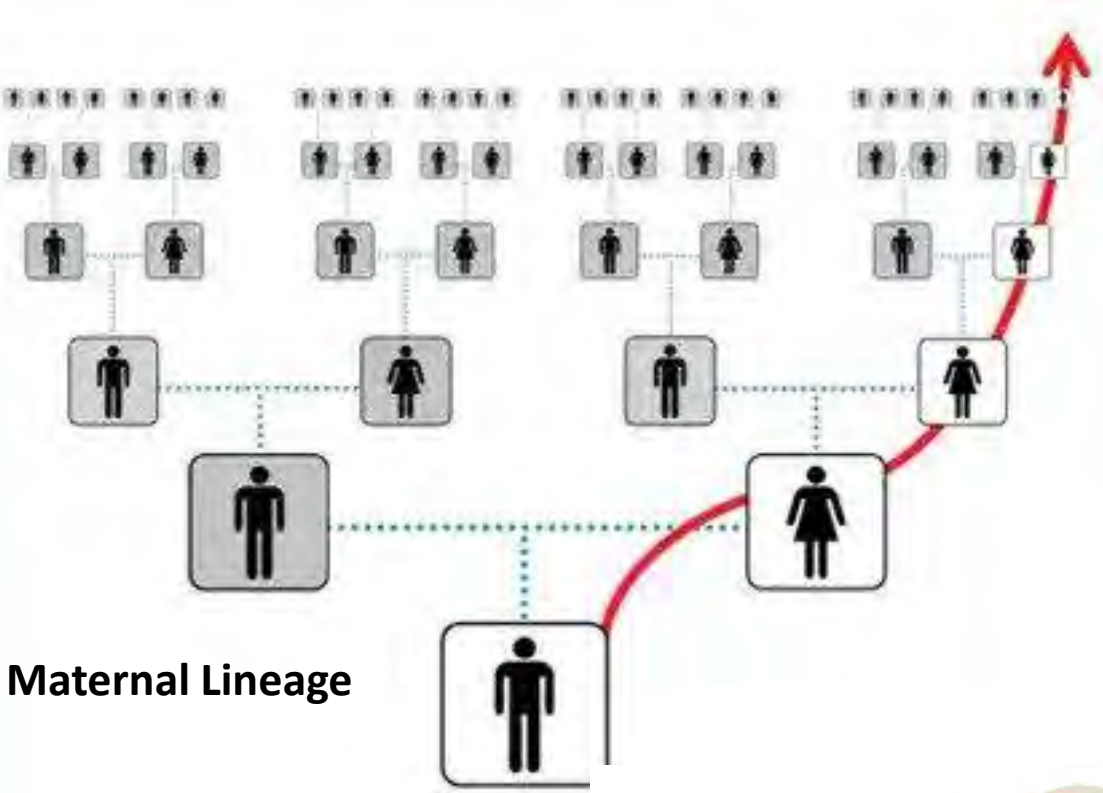
7. Applications of Plant mtDNA in Research

Phylogenetics and Evolutionary Studies: Used to study deep evolutionary relationships due to its slow mutation rate.

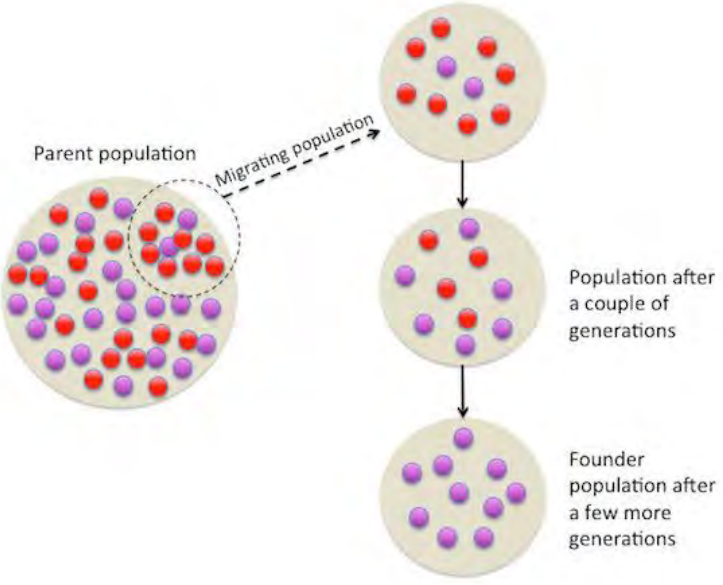
Hybrid Identification and Breeding: Helps in tracking **maternal lineage and CMS traits**.

Molecular Ecology and Conservation: Useful for studying **population genetics and plant dispersal patterns**.





Maternal Lineage



population genetics

Nuclear Genome in Plants

1. Introduction to the Nuclear Genome

The **nuclear genome (nDNA)** is the **largest and most complex genome** in plant cells, housed within the **nucleus**. It contains the **majority of genetic material** responsible for plant growth, development, and reproduction.

Organized into **chromosomes**, it follows **Mendelian inheritance** (biparental).

2. Characteristics of the Plant Nuclear Genome

Large and Highly Variable in Size: Ranges from **small genomes (~100 Mb, e.g., Arabidopsis)** to **gigantic genomes (>100 Gb, e.g., Paris japonica)**.

Linear, Chromosomal Structure: Unlike **circular** cpDNA or mtDNA, the nuclear genome consists of **multiple linear chromosomes**.

Contains Both Coding and Non-Coding DNA:

Coding regions (**genes**) constitute a small percentage.

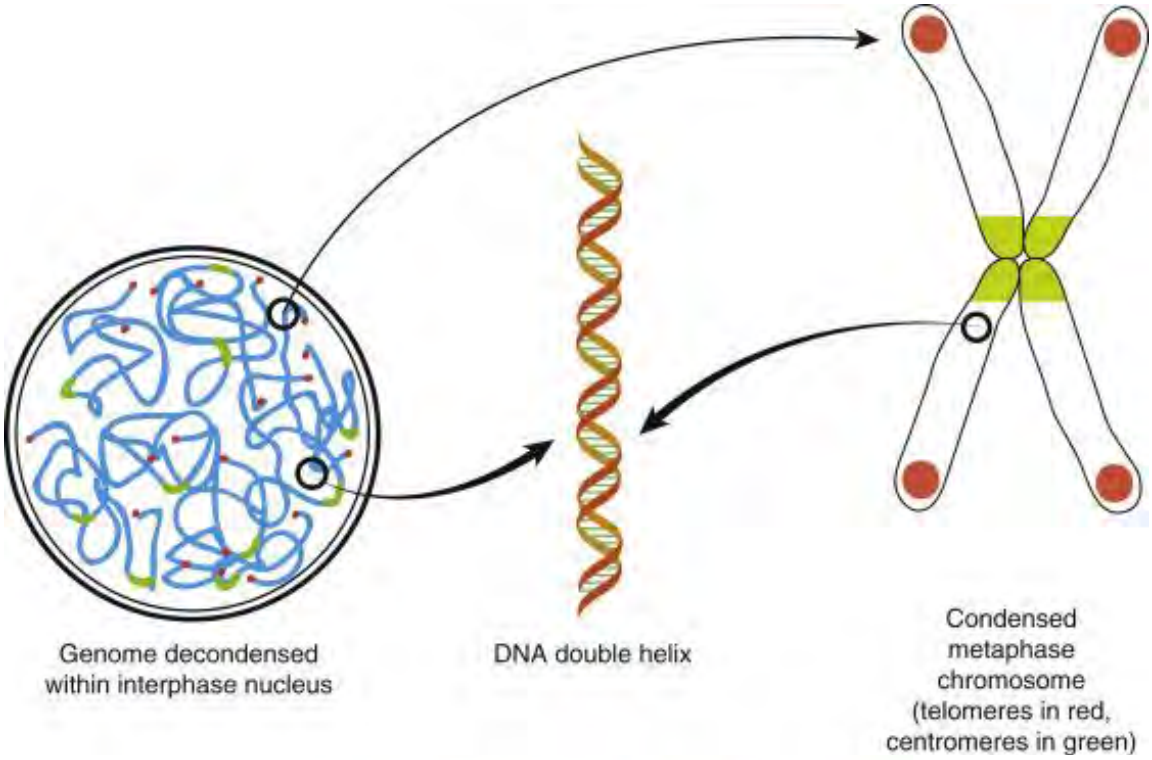
Majority consists of **non-coding DNA**, including introns, regulatory sequences, and transposable elements.

Highly Repetitive Sequences:

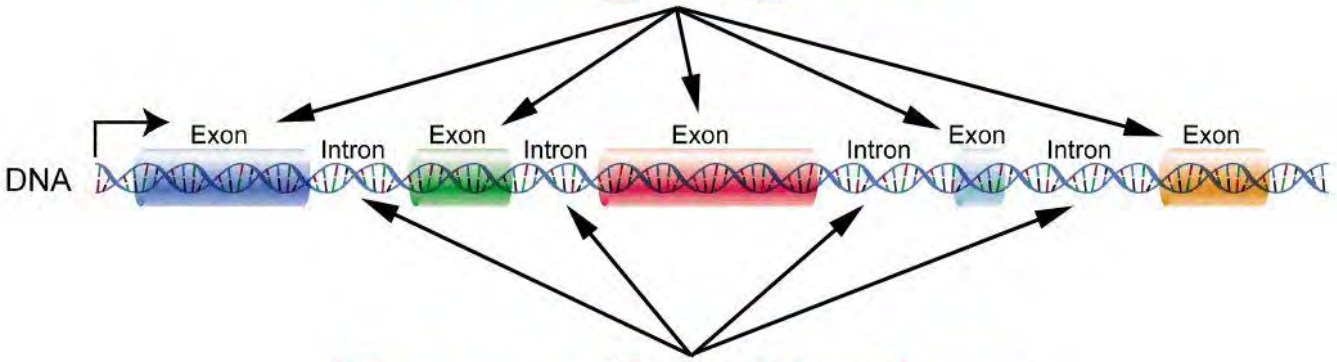
Includes **transposons and satellite DNA**, which contribute to genome size expansion.

Undergoes Recombination:

Unlike cpDNA and mtDNA, nuclear DNA **undergoes genetic recombination** during **meiosis**, promoting genetic diversity.

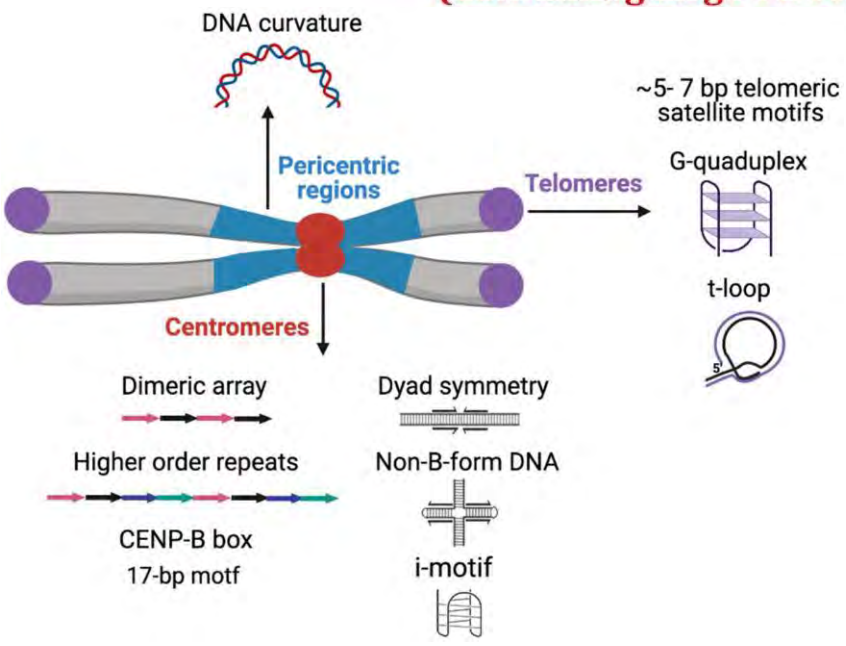


Coding Regions



Non-coding Regions

(Containing large TE content)



3. Organization of the Plant Nuclear Genome

Chromosomes:

Plants have a **varying number of chromosomes**, often with **polyploidy** (multiple chromosome sets).
Example: **Wheat (Triticum aestivum) is hexaploid (6n = 42 chromosomes)**.

Genes and Coding Regions:

Encodes **structural proteins, enzymes, transcription factors, and regulatory RNAs**.
Genes contain **exons (coding regions) and introns (non-coding regions)**.

Regulatory Elements:

Includes **promoters, enhancers, and silencers** that regulate gene expression.

4. Polyploidy in Plants

Common in plants, leading to genome expansion and evolution.

Types:

Autopolyploidy (duplication within a species).
Allopolyploidy (hybridization between species).

Examples:

Wheat (Hexaploid), Cotton (Tetraploid), Strawberry (Octoploid).

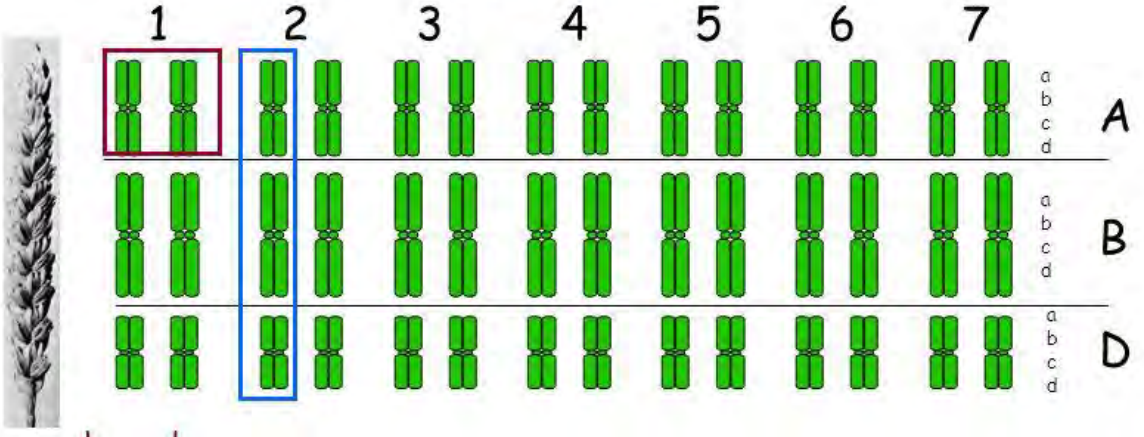
Polyploidy contributes to **genetic diversity, adaptation, and speciation**.

To be fertile, true homologues must pair at meiosis

Hexaploid (Bread) wheat

Triticum aestivum

2n = 6x =42



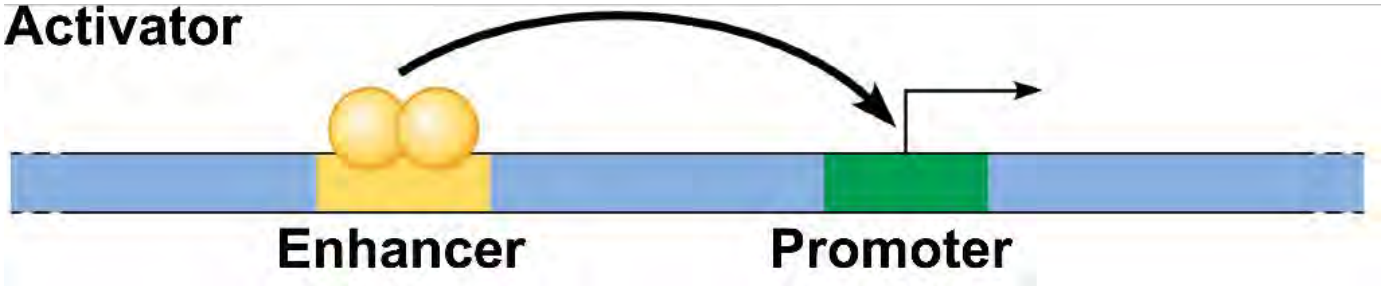
homologues

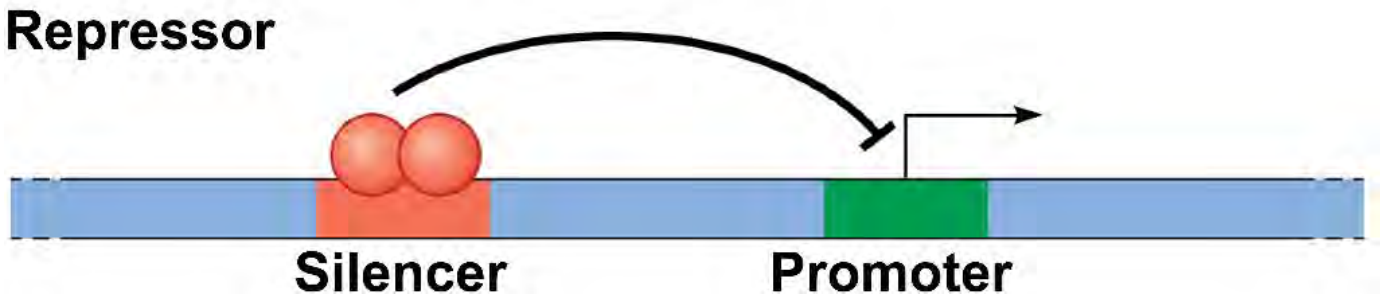
homoeologues

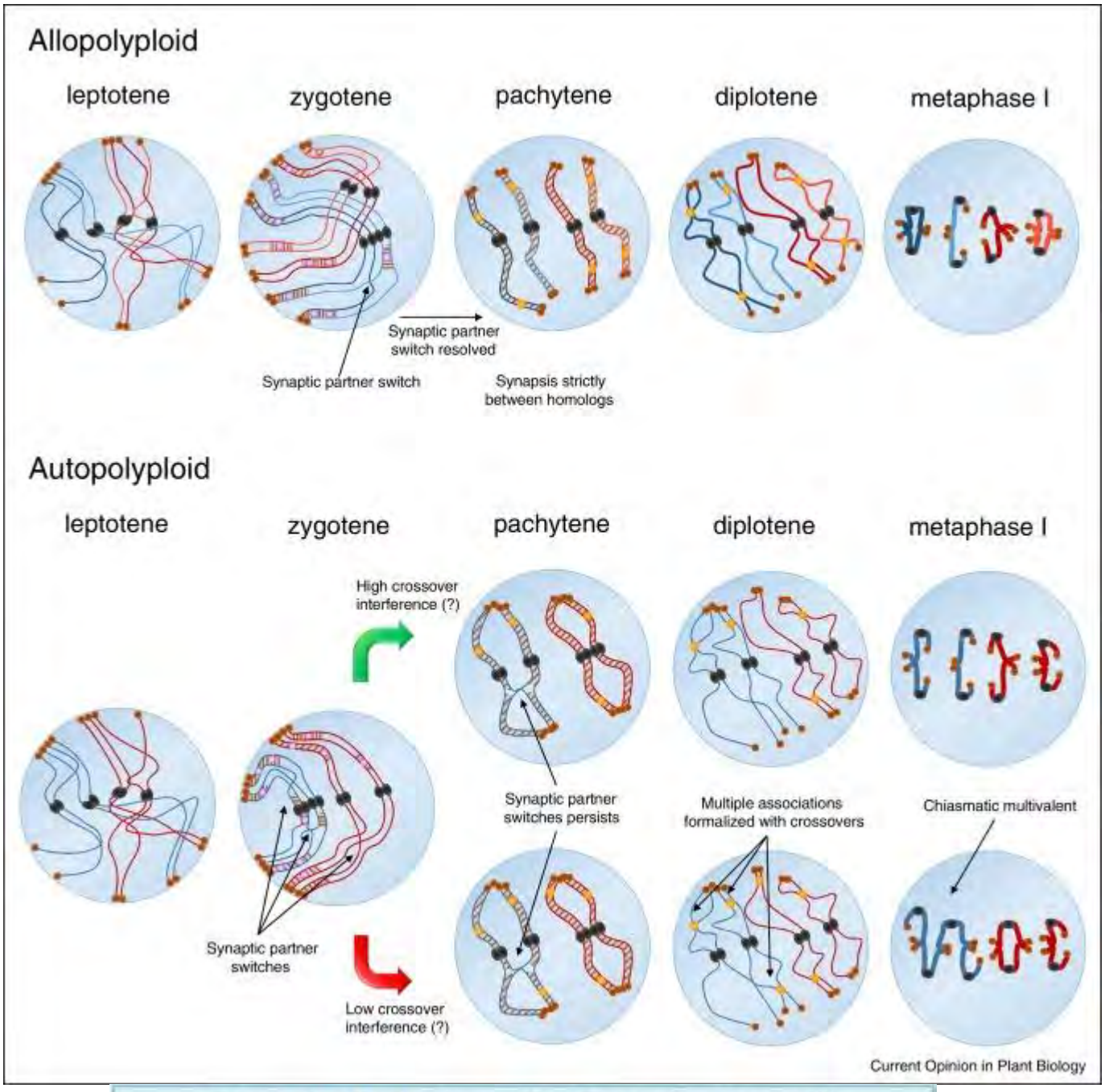
Diploid - Homologues distinguished from non-homologous chromosomes

Wheat - Homologues, homoeologues and non-homologous chromosomes

Enhancer and Silencer DNA elements.







Gossypium herbaceum **X** ***Gossypium ramondii***

Old world cotton **American upland cotton**

2n= 26 **2n=26**

(small chromosomes) **(Large chromosomes)**

F1 = 2n = 13 small + 13 Large (26 univalents)

Chromosome doubling

2n = 52 (26 bivalents = 13 small+13 large)

Gossypium hirsutum **(New world Cotton)**

5. Transposable Elements (TEs) in the Nuclear Genome

"Jumping genes" discovered by **Barbara McClintock**.

Two types:

Retrotransposons (copy-and-paste mechanism).

DNA transposons (cut-and-paste mechanism).

Account for **a large fraction of nuclear DNA**, influencing genome evolution and function.

6. Gene Expression in the Nuclear Genome

Transcription:

DNA → RNA (by RNA polymerase).

Post-Transcriptional Modifications:

mRNA splicing (removal of introns).

Addition of **5' cap and poly-A tail**.

Translation:

mRNA → Proteins (in ribosomes).

Epigenetic Regulation:

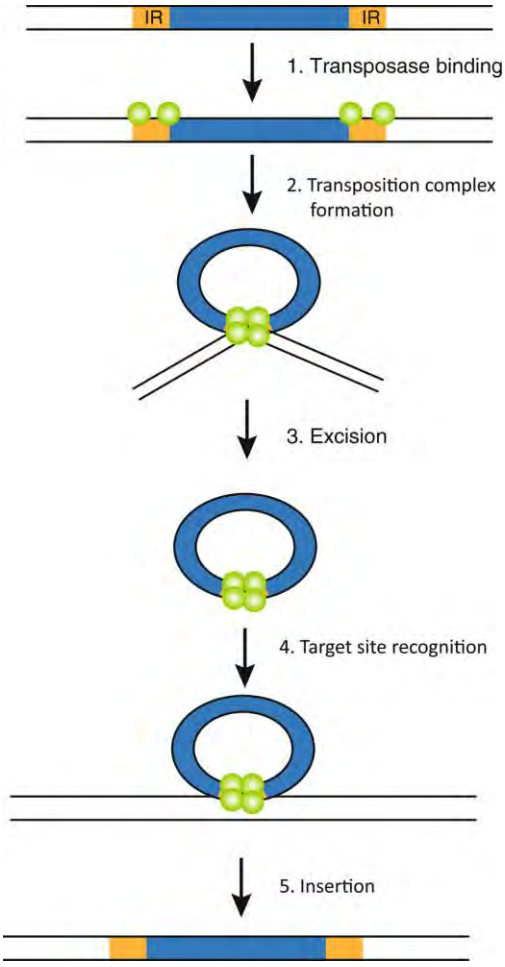
DNA methylation, histone modification, and **RNA interference (RNAi)** control gene expression.

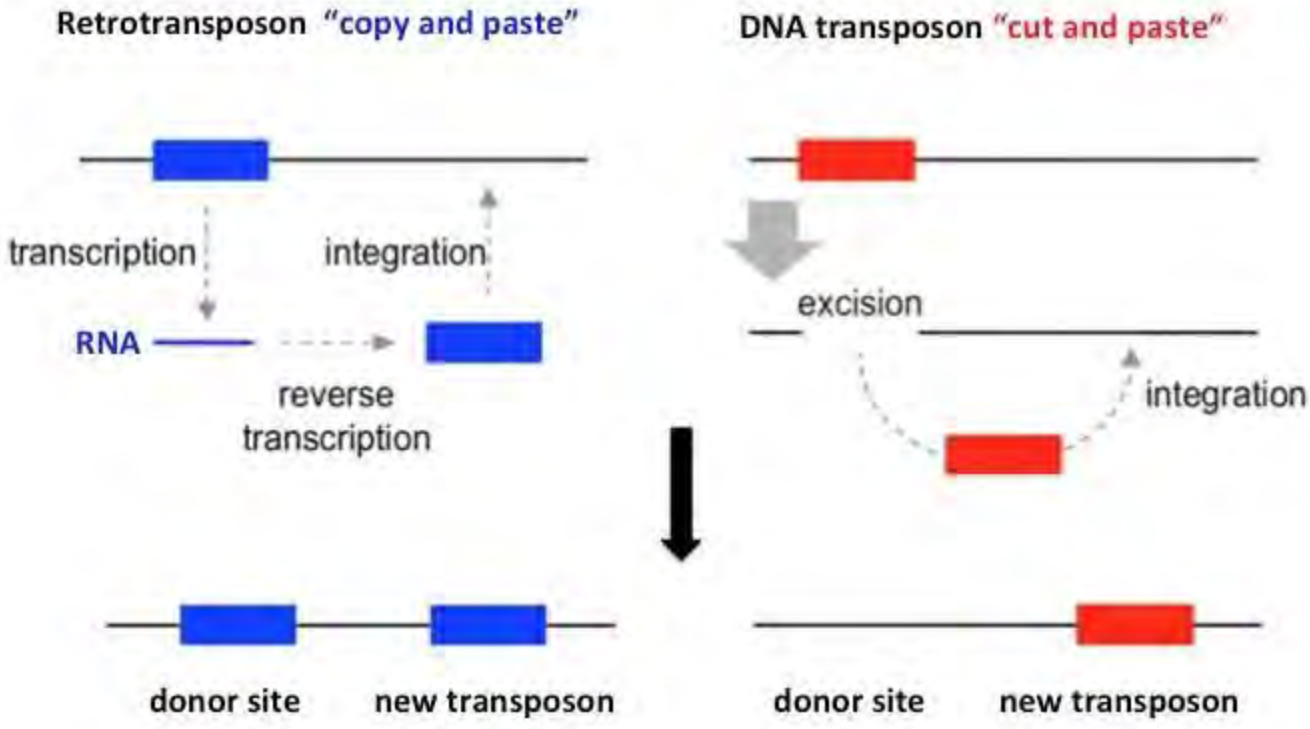
7. Role of the Nuclear Genome in Evolution and Adaptation

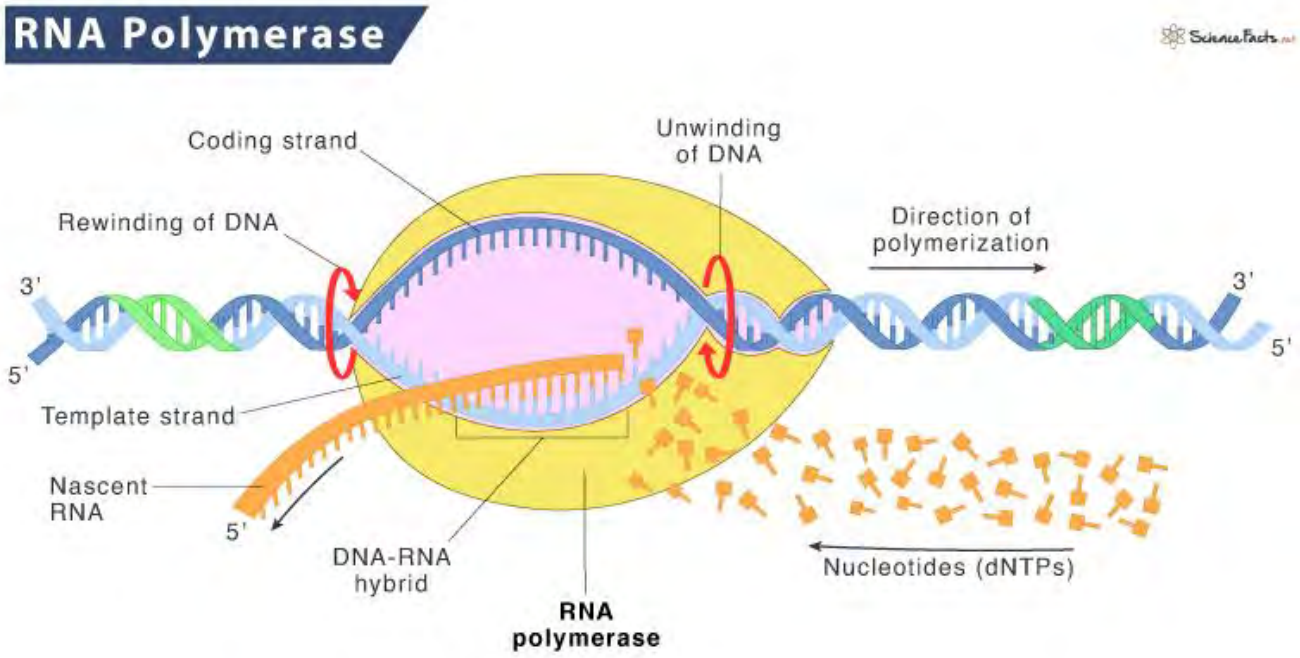
Mutations, genetic recombination, and polyploidy drive **plant diversity and adaptation**.

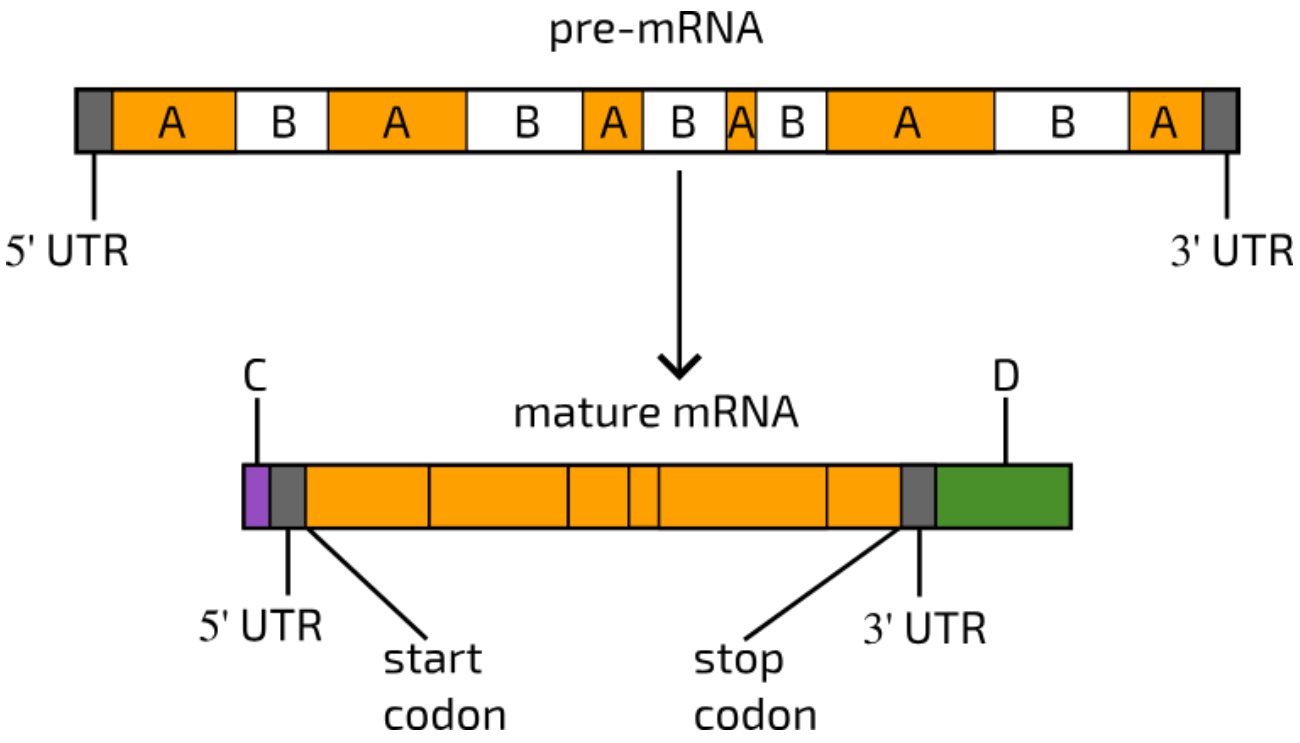
Facilitates **domestication of crop plants** through selective breeding and hybridization.

Comparative genomics helps in understanding plant phylogeny and evolution.


IR
IR
1. Transposase binding
2. Transposition complex formation
3. Excision
4. Target site recognition
5. Insertion

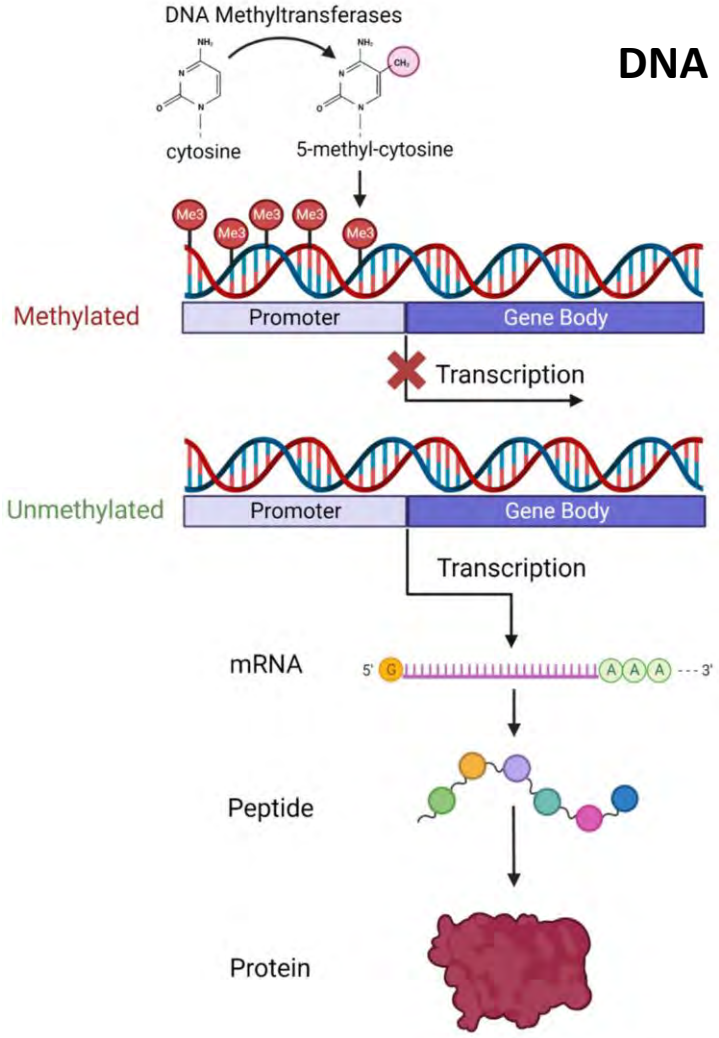

Retrotransposon "copy and paste"
transcription
integration
RNA
reverse transcription
donor site
new transposon
DNA transposon "cut and paste"
excision
integration
donor site
new transposon

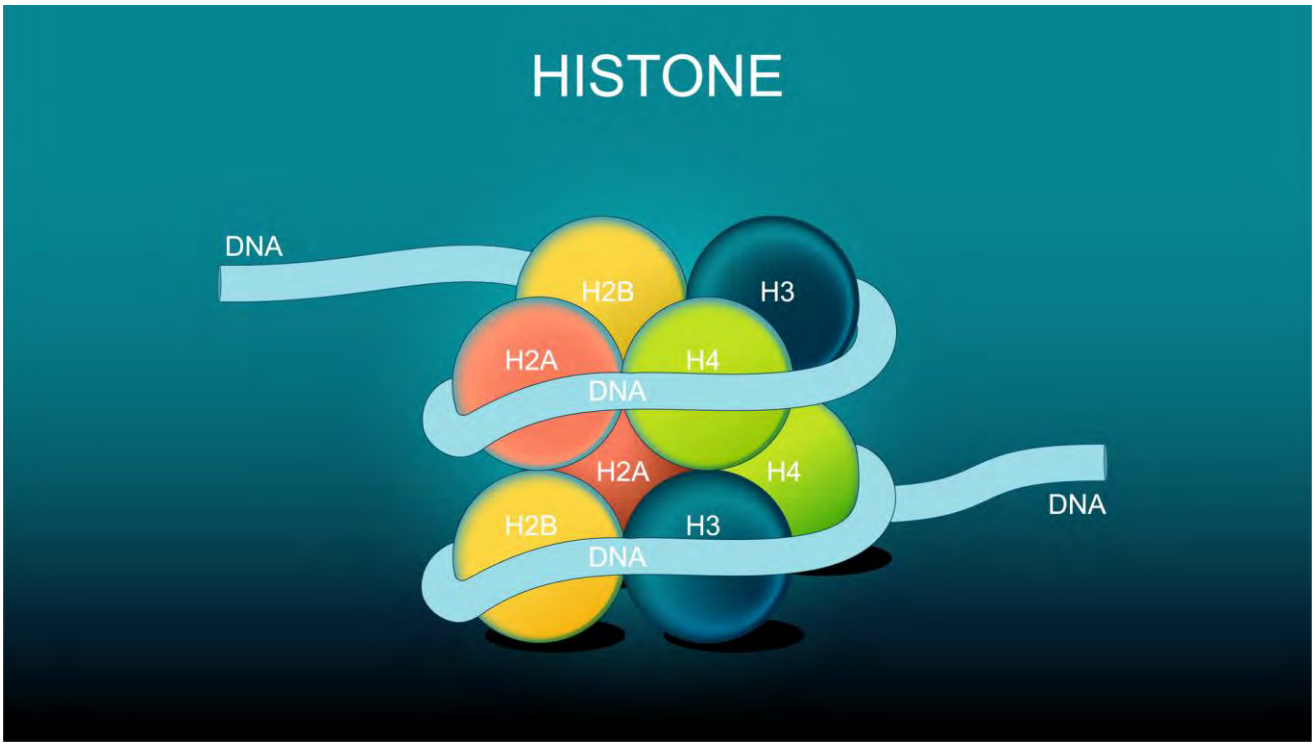




Post-Transcriptional Modifications

DNA methylation





8. Applications of Nuclear Genome Studies

Crop Improvement: Genomic selection for high-yield, disease resistance, and stress tolerance.

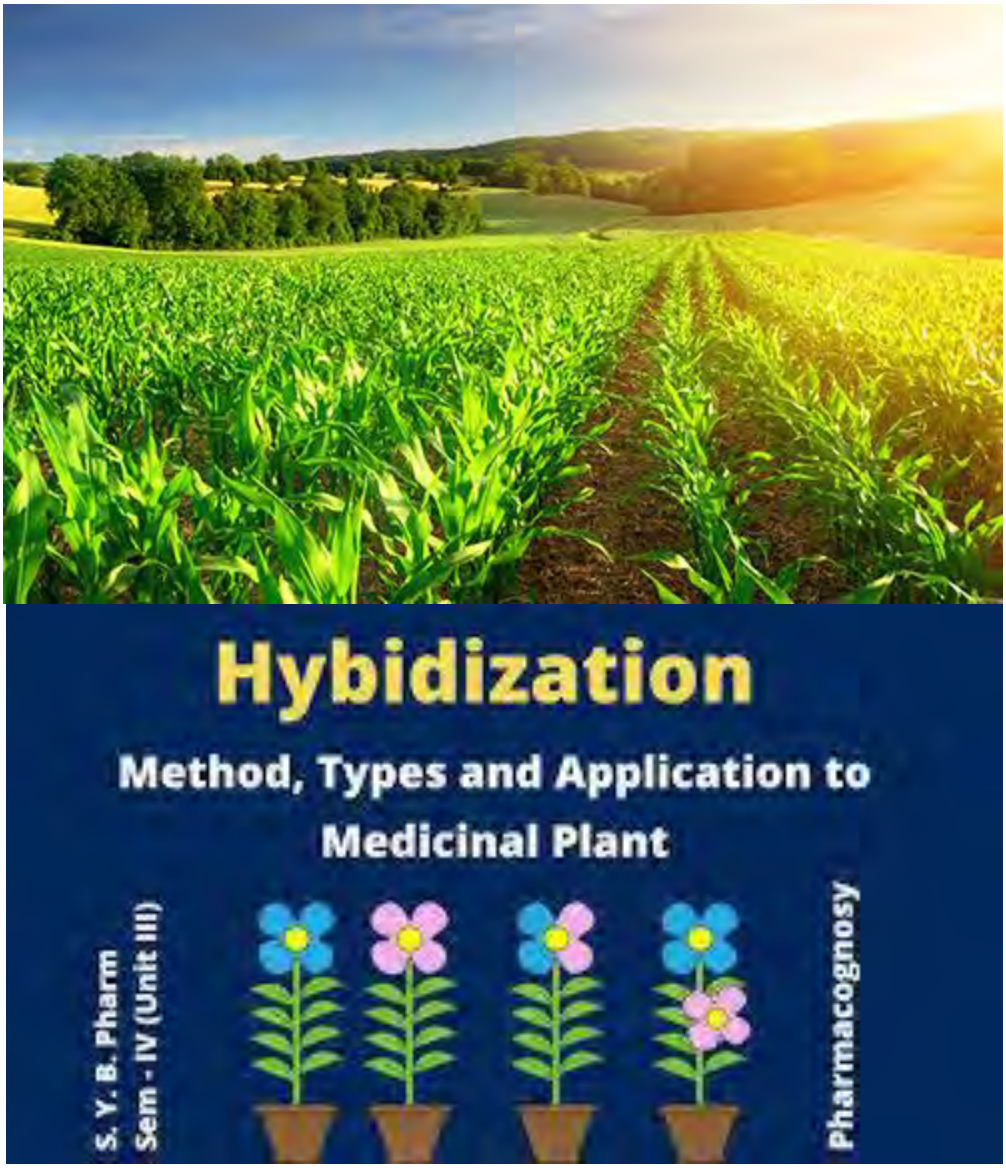
Plant Breeding & Hybridization: Understanding nuclear DNA helps in developing better cultivars.

Phylogenetics & Evolution: Helps in studying plant relationships at species and family levels.

Biotechnology & Genetic Engineering:

CRISPR-based genome editing for trait improvement.

Gene transfer for creating genetically modified (GM) crops.



Comparison of Mitochondrial, Nuclear, and Chloroplast Genomes

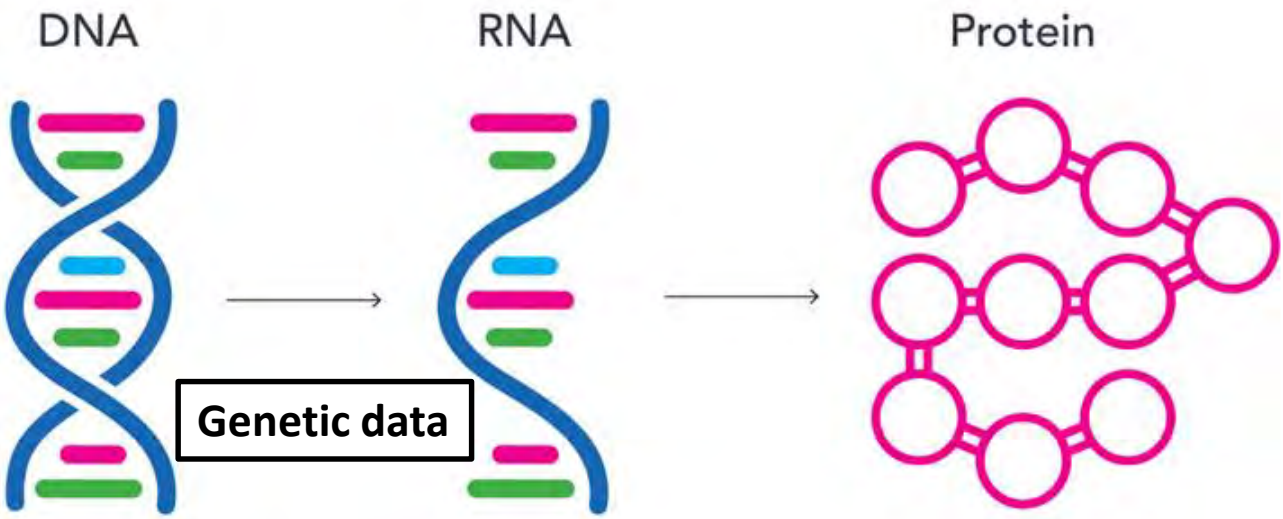
Feature	Mitochondrial Genome (mtDNA)	Nuclear Genome (nDNA)	Chloroplast Genome (cpDNA)
Location	Mitochondria	Nucleus	Chloroplast
Inheritance	Mostly maternal	Biparental	Mostly maternal
Size	Large (200 kb – 11.3 Mb)	Largest (100s-1000s Mb)	Small (~120-160 kb)
Recombination	High	Moderate	Low
Mutation Rate	Low	Moderate	High
RNA Editing	Common	Rare	Present but less frequent
Main Function	Energy production	Encodes structural & regulatory proteins	Photosynthesis

Advance Experimental Taxonomy

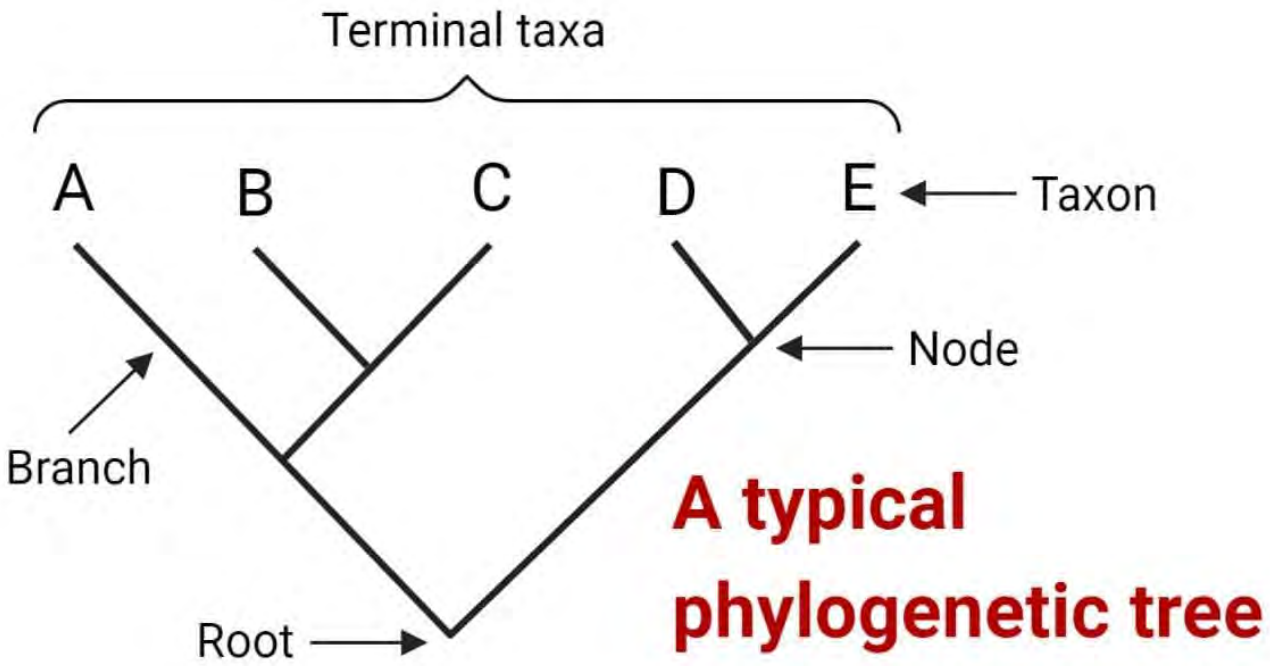
Chapter 5- Phylogenetic analyses: DNA sequences

Introduction to Phylogenetic Analysis

- **Study of evolutionary relationships**
  - Phylogenetics examines how different organisms, genes, or proteins are related through common ancestry.
  - It helps scientists understand the evolutionary history of life on Earth.
- **Uses genetic, morphological, or biochemical data**
  - **Genetic data:** DNA, RNA, or protein sequences provide molecular evidence of relationships.
  - **Morphological data:** Physical characteristics, such as bone structure or body shape, help compare extinct and living species.
  - **Biochemical data:** Enzyme activity, metabolic pathways, and other molecular functions can also indicate evolutionary links.
- **Constructs phylogenetic trees**
  - Phylogenetic trees visually represent relationships between species or genes.
  - These trees can be rooted (with a common ancestor) or unrooted (showing relative relationships).
  - Trees help classify organisms, track disease evolution, and study genetic variations.



Biochemical data



Importance of Phylogenetics

Understands species and gene evolution

- Helps trace the evolutionary history of organisms, showing how species have evolved over time.
- Identifies common ancestors and evolutionary divergence between species.
- Assists in studying gene evolution, gene duplication, and horizontal gene transfer.

Helps in disease research and vaccine development

- Tracks the evolution of viruses, bacteria, and other pathogens.
- Helps understand mutation rates and the emergence of new strains (e.g., COVID-19 variants).
- Guides vaccine design by identifying conserved genetic regions of viruses.
- Assists in tracing outbreaks and controlling disease spread.

Supports conservation efforts

- Identifies evolutionary significant units (ESUs) for conservation planning.
- Helps prioritize species for protection based on genetic diversity and evolutionary history.
- Used in studying endangered species' genetic variability to improve breeding programs.

Used in agriculture and biotechnology

- Helps in crop improvement by identifying related species with desirable traits.
- Assists in developing disease-resistant and high-yielding plant varieties.
- Supports livestock breeding programs by studying genetic relationships.
- Aids in biotechnology applications like genetic modification and synthetic biology.

Key Terms in Phylogenetics

Taxon (plural: Taxa)

A taxon is any group of organisms that are classified together based on shared characteristics.

Examples: A species (e.g., *Homo sapiens*), a genus (e.g., *Felis*), or a family (e.g., Felidae).

Node

A node represents a point of divergence in a phylogenetic tree.

Internal nodes indicate common ancestors of multiple taxa.

Terminal nodes (leaves/tips) represent existing species, genes, or sequences.

Branch

A branch connects nodes and represents evolutionary relationships.

The length of a branch can indicate evolutionary time or genetic change, depending on the tree type.

Clade

A clade includes a common ancestor and all of its descendants.

Also called a **monophyletic group**.

Example: Birds and reptiles belong to the same clade because they share a common ancestor.

Root

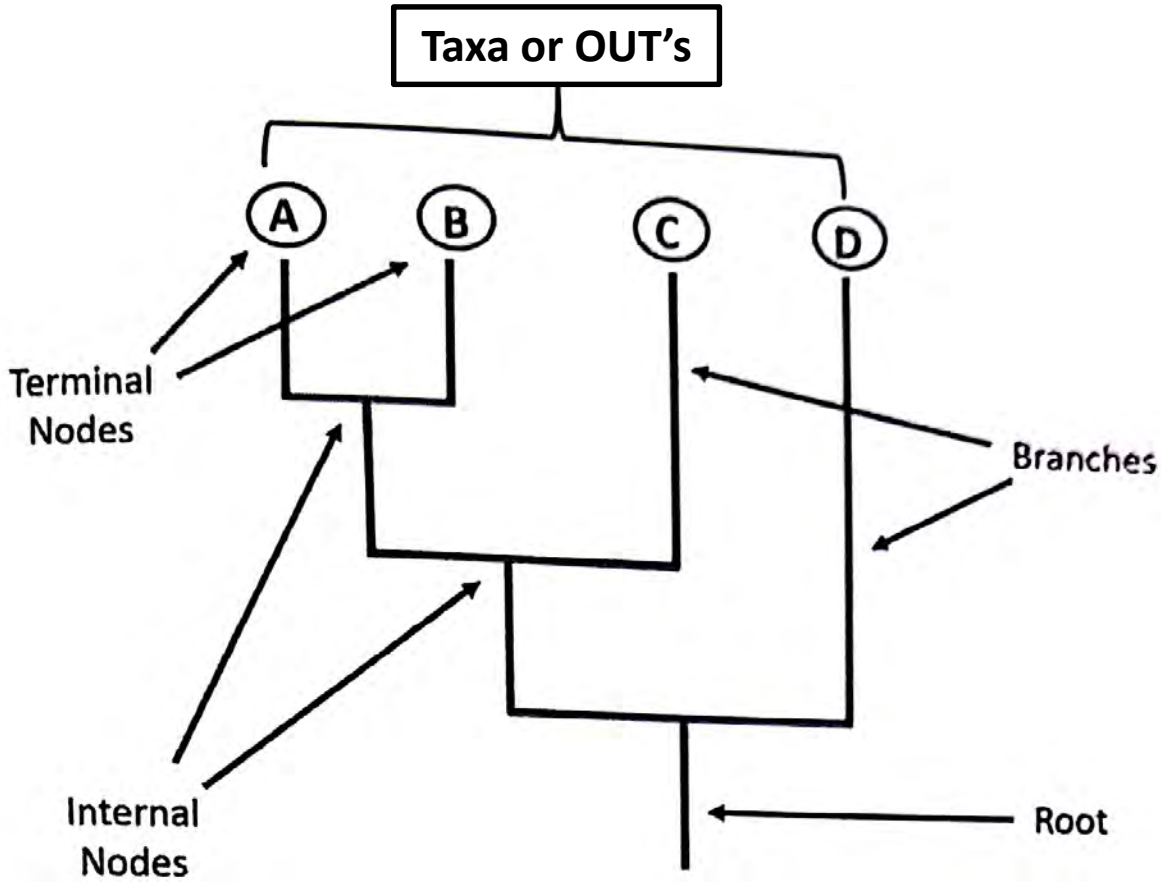
The root is the most ancestral node in a **rooted** phylogenetic tree.

It represents the oldest common ancestor of all taxa in the tree.

Outgroup

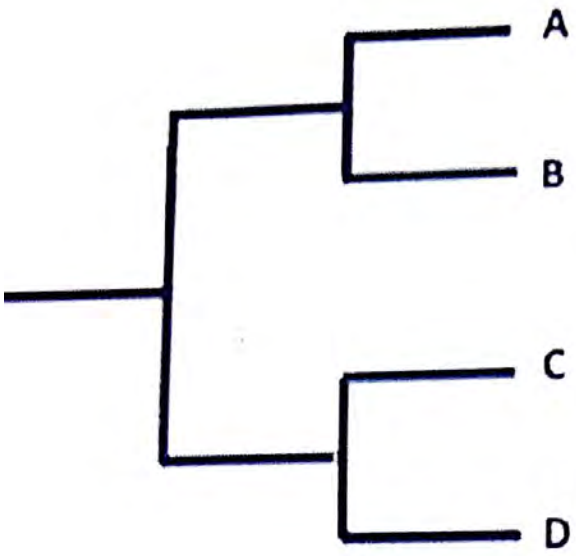
An outgroup is a taxon that is distantly related to the group being studied.

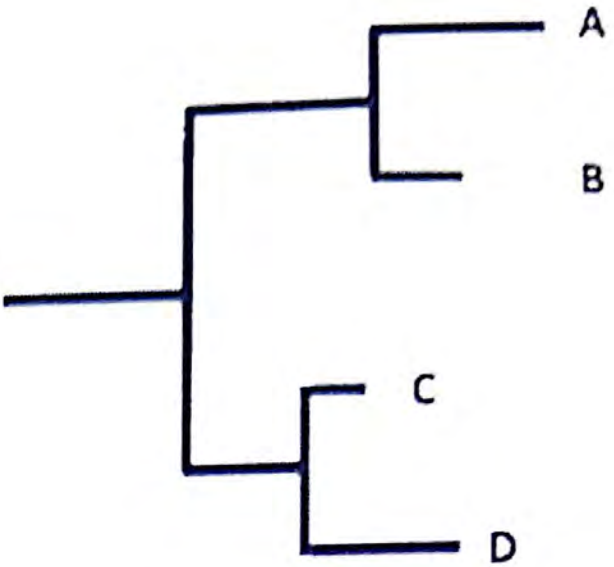
Helps in determining the evolutionary direction and rooting the tree.



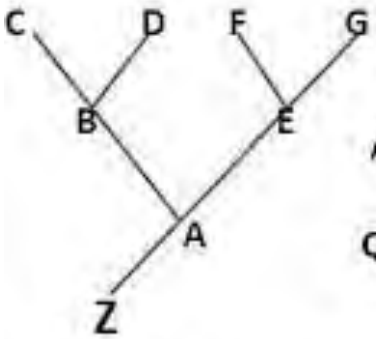
A typical phylogenetic tree showing root, nodes and branches

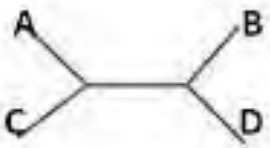
A phylogenetic tree in rectangular and slanted form (the trees can be drawn as angled form (right) or squared form (left)





A phylogenetic tree drawn as cladogram (left) and phylogram (right). The branch lengths are unscled in the cladogram and scaled in the phylogram

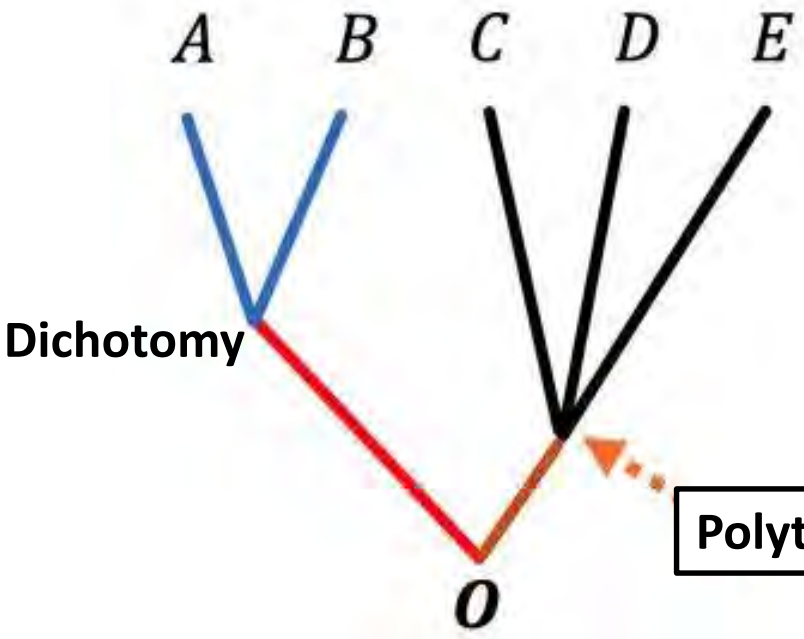




Rooted tree

Unrooted tree.

Partially resolved tree



Sister Taxa

Two taxa that share an immediate common ancestor.

They are each other's closest relatives in the tree.

Monophyletic, Paraphyletic, and Polyphyletic Groups

Monophyletic group (Clade): A group that includes an ancestor and all its descendants.

Paraphyletic group: Includes an ancestor but not all of its descendants.

Polyphyletic group: Includes species from different ancestors, ignoring true evolutionary relationships.

Homology and Analogy

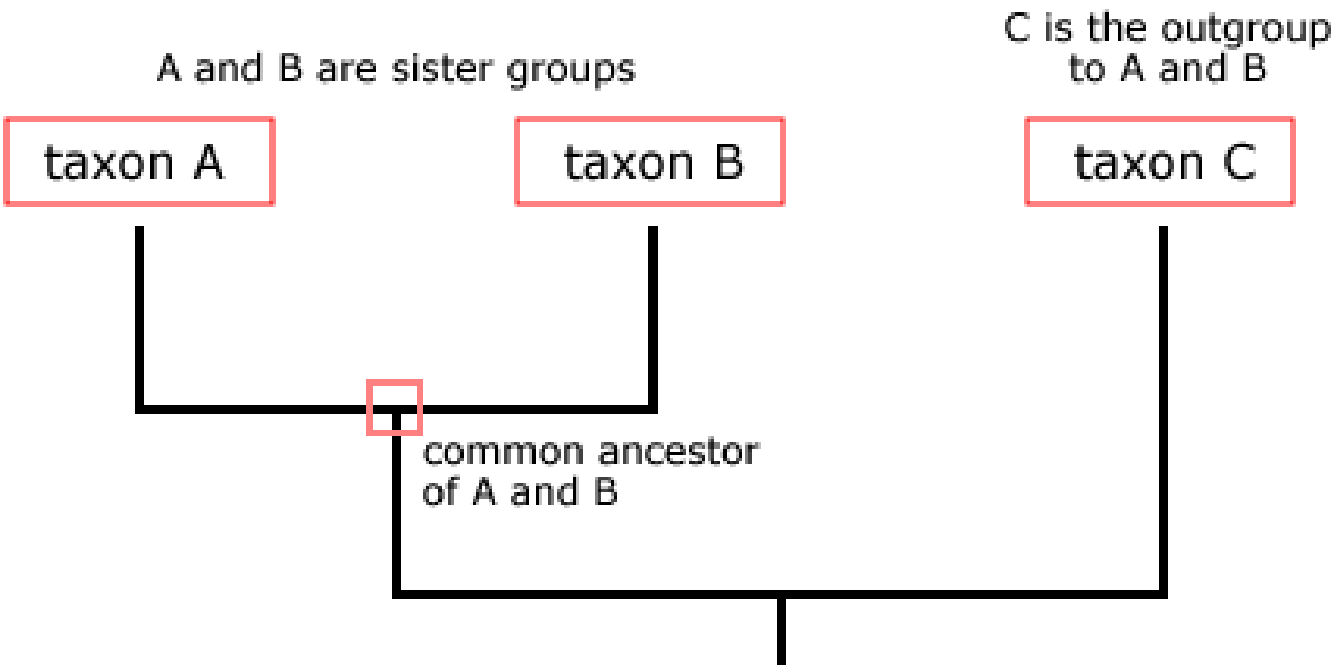
Homologous traits: Features inherited from a common ancestor (e.g., vertebrate limbs).

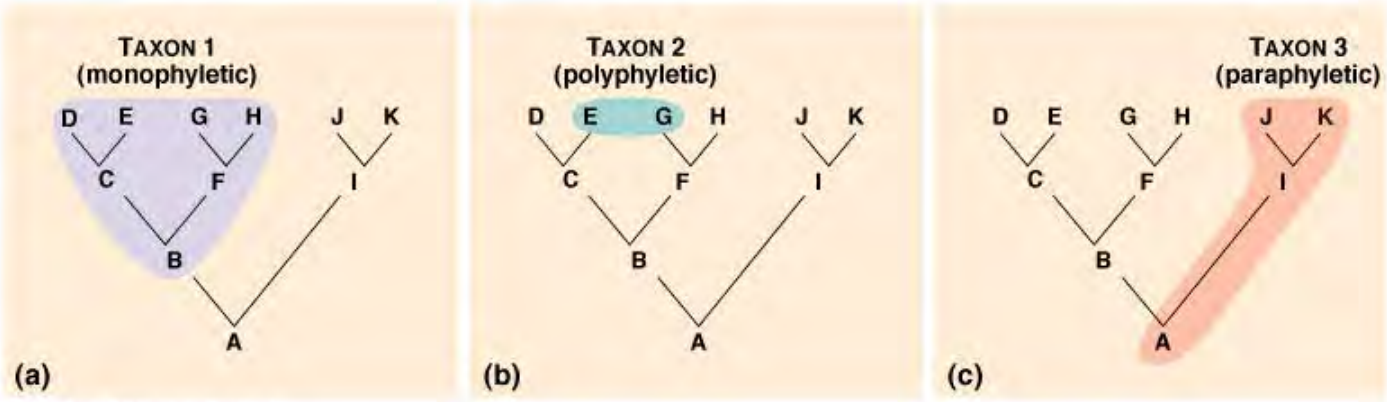
Analogous traits (Convergent evolution): Similar traits evolved independently in different lineages (e.g., wings in birds and bats).

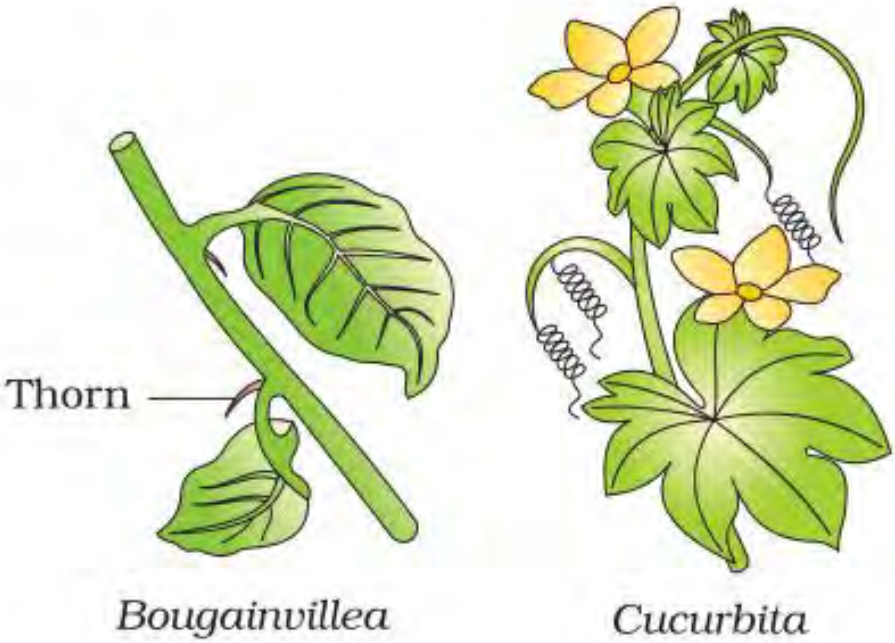
Bootstrap Value

A statistical method used to measure the reliability of a phylogenetic tree.

Higher bootstrap values (e.g., 90% or above) indicate strong support for the grouping.







Homologous traits

Convergent evolution

Types of Phylogenetic Trees

Rooted Tree

Represents evolutionary relationships with a single common ancestor at the base.

Shows the direction of evolution from the ancestor to present-day taxa.

Example: A tree of vertebrates with a common ancestor at the root.

Unrooted Tree

Displays relationships among taxa but does not indicate a common ancestor.

Does not show the direction of evolutionary change.

Useful when the root is unknown or uncertain.

Cladogram

Represents evolutionary relationships but does not show the amount of change.

Branches have equal lengths, emphasizing only the order of divergence.

Example: A simple tree showing how mammals, birds, and reptiles are related.

Phylogram

Shows evolutionary relationships and the **amount of genetic change** through branch lengths.

Longer branches indicate more evolutionary changes.

Used in molecular phylogenetics with DNA sequence data.

Polytomy

A node where more than two branches emerge, indicating an uncertain relationship.

Represents an unresolved evolutionary relationship due to lack of data or rapid diversification.

Outgroup

A taxon included in the tree that is **distantly related** to the rest of the taxa.

Helps in rooting the tree and determining the direction of evolution.

Example: In a tree of mammals, a reptile can be used as an outgroup.

Sister Group

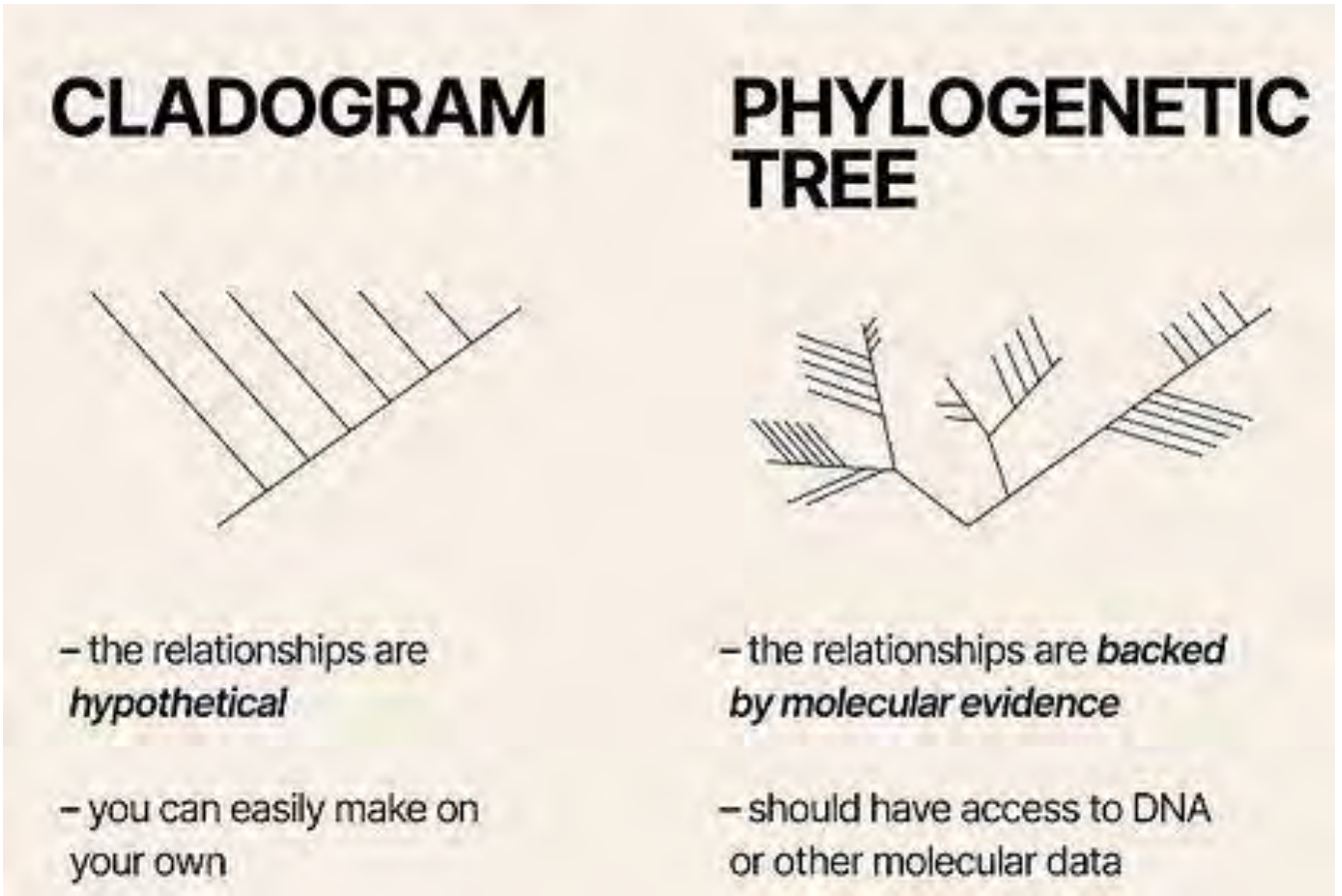
Two groups that share a **recent common ancestor** and are each other's closest relatives.

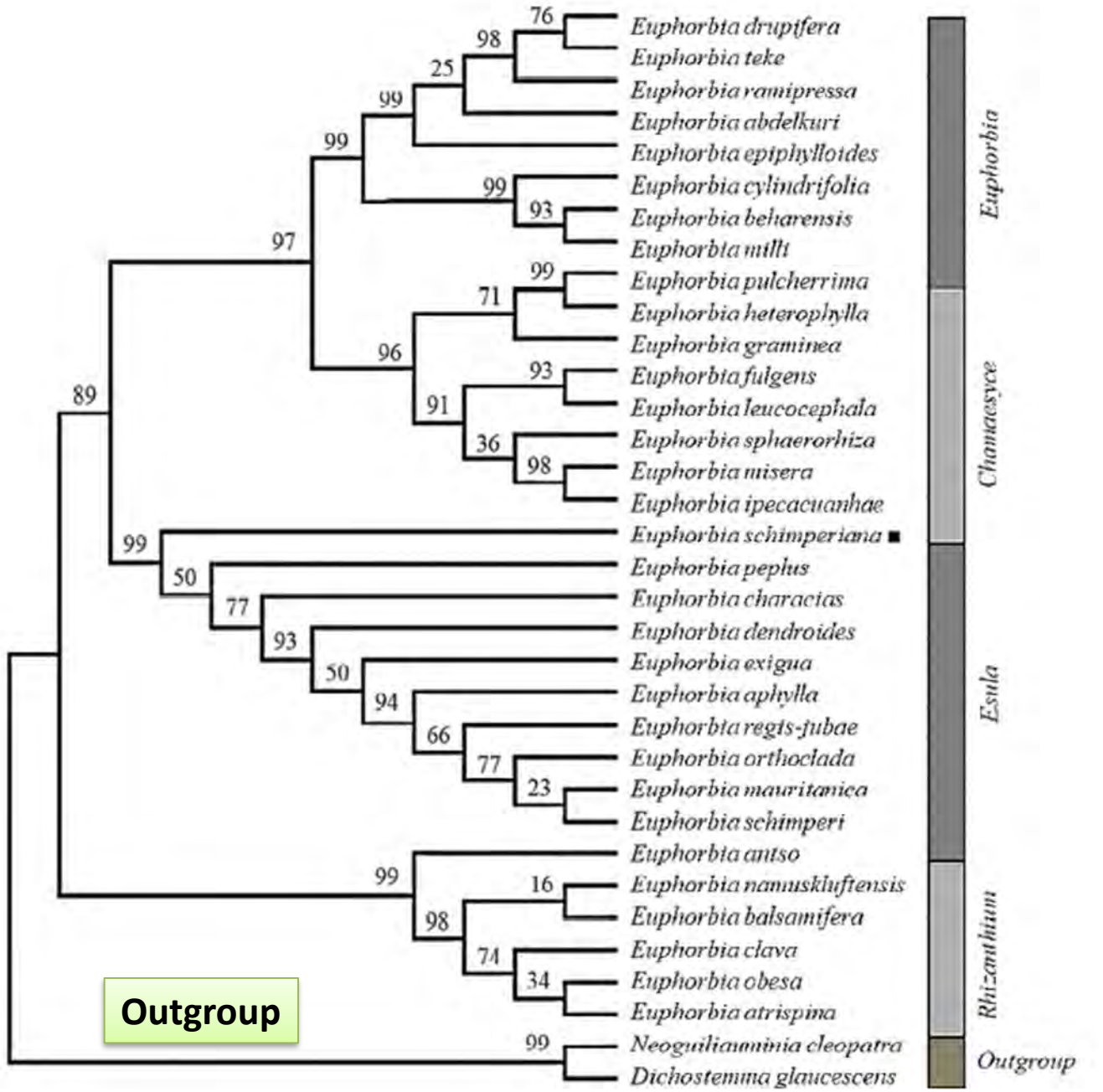
Example: Humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*) are sister taxa.

Rotation of Trees

Phylogenetic trees can be rotated around any node **without changing evolutionary relationships**.

Tree rotation does not affect the relationships but can make trees easier to read.





Molecular phylogenetics of *Euphorbia schimperiana* inferred from nrDNA ITS sequences using the Maximum Parsimony method.

Tree Terminology in Phylogenetics

1. Character Trees and Sequence Trees

Character Trees:

Constructed using morphological or molecular traits (characters).

Each trait is coded as a data point (e.g., presence or absence of a feature).

Example: Comparing bone structures in vertebrates.

Sequence Trees:

Built using DNA, RNA, or protein sequences.

Compares genetic similarity to infer evolutionary relationships.

Example: Tracking virus evolution using genome sequences.

2. Gene Trees vs. Species Trees

Gene Tree:

Represents the evolutionary history of a single gene.

May differ from the species tree due to gene duplication, horizontal gene transfer, or incomplete lineage sorting.

Example: Hemoglobin gene evolution across mammals.

Species Tree:

Shows the evolutionary relationships among species.

Based on multiple genes or whole-genome data.

Represents the actual divergence history of species.

3. Consensus Tree

A tree constructed by combining multiple phylogenetic trees.

Used when different methods give slightly different results.

Types:

Majority-rule consensus tree: Includes groups that appear in most input trees.

Strict consensus tree: Includes only groups that appear in all input trees.

4. Super Tree

A large tree assembled from multiple smaller phylogenetic trees.

Integrates data from different studies or datasets.

Useful in large-scale evolutionary studies.

5. Diagonal Tree

A visually distinct way of displaying phylogenetic trees.

Uses diagonal or slanted branches instead of traditional vertical or horizontal formats.

Helps in visualizing relationships more compactly.

6. Phylogenetic Network

A network-based representation of evolutionary relationships.

Used when evolution is not strictly tree-like (e.g., hybridization, horizontal gene transfer).

Examples: Reticulate evolution in plants, bacterial gene transfer.

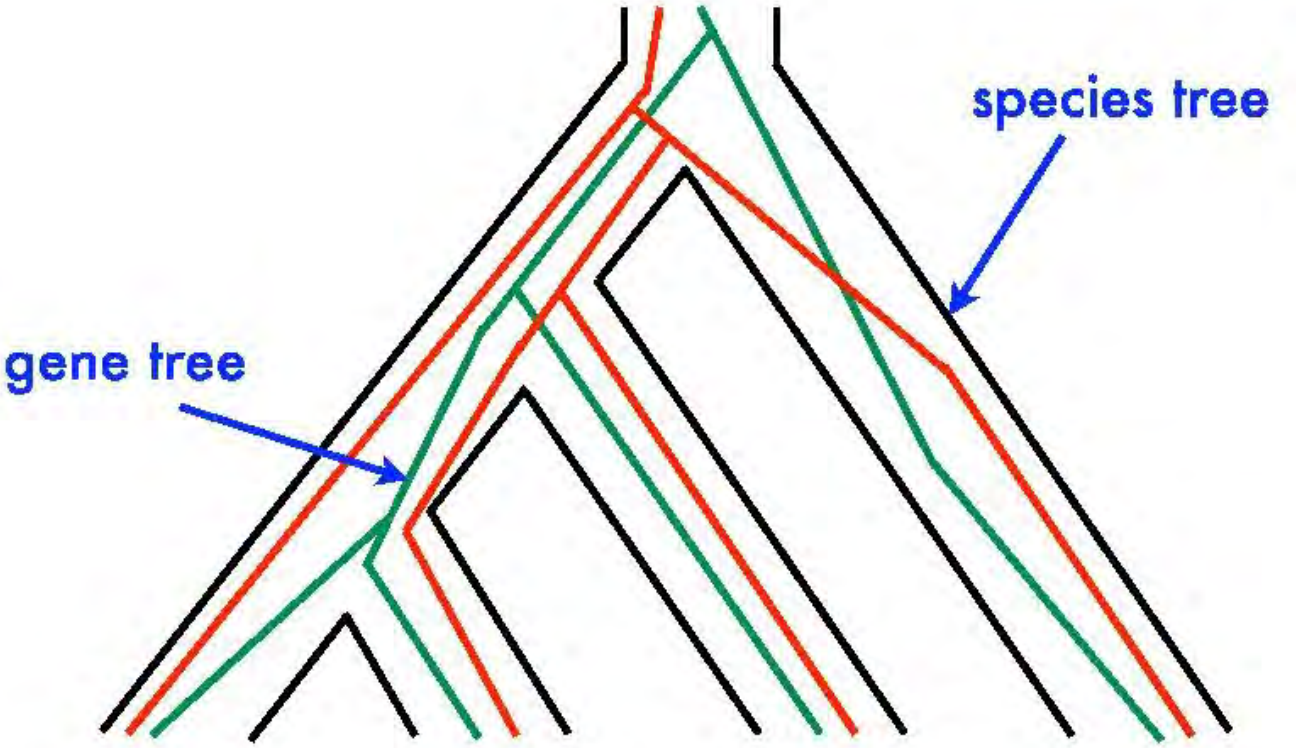
7. Evolutionary Models

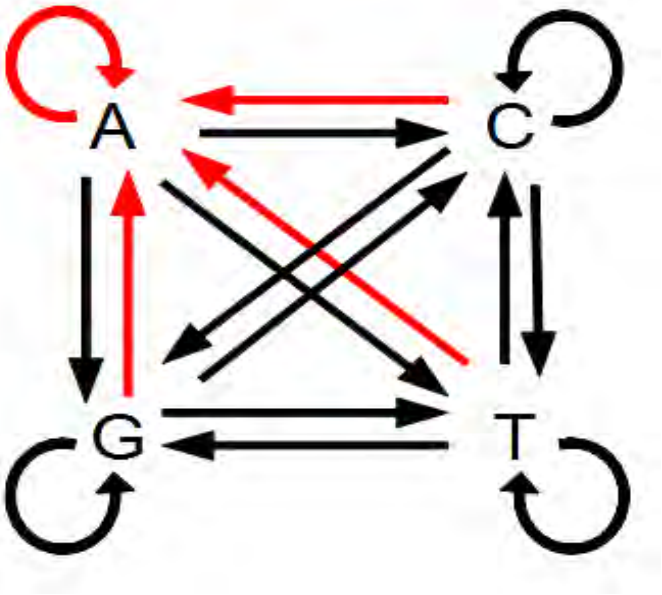
Mathematical models used to describe genetic changes over time.

Common models:

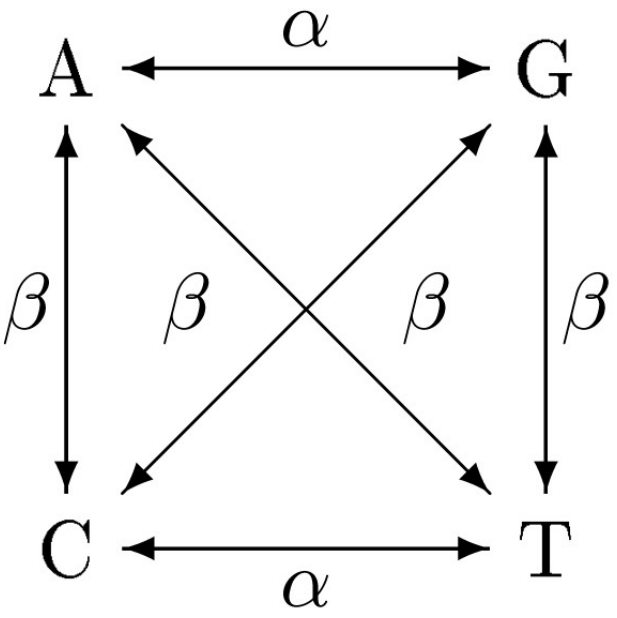
- **Jukes-Cantor (JC) model:** Assumes equal mutation probabilities for all nucleotides.
- **Kimura 2-Parameter (K2P) model:** Accounts for differences in transition and transversion mutations.
- **General Time Reversible (GTR) model:** The most flexible, allowing different mutation rates.

Helps in estimating branch lengths in phylogenetic trees.





Jukes-Cantor (JC) model



Kimura 2-Parameter (K2P) model

Major Methods for Estimating Phylogenetic Trees

1. Distance-Based Approach

Uses pairwise distances between sequences to construct trees.

Assumes that the greater the genetic difference, the more distantly related the taxa are.

Common methods:

- **UPGMA (Unweighted Pair Group Method with Arithmetic Mean)**
- **Neighbor-Joining (NJ) Method**

2. Tree Searching Approach

Instead of relying on pairwise distances, these methods explore different tree topologies to find the best one.

Typically use **character-based** data (e.g., DNA, protein sequences).

More computationally intensive but often more accurate.

Common methods:

- **Maximum Parsimony (MP)**
- **Maximum Likelihood (ML)**
- **Bayesian Inference (BI)**

1. Distance-Based Approach

UPGMA (Unweighted Pair Group Method with Arithmetic Mean)

Overview:

A distance-based method for constructing phylogenetic trees.
Assumes a **molecular clock** (i.e., evolution occurs at a constant rate).
Produces a **rooted tree**, where branch lengths reflect evolutionary time.

Steps in UPGMA Algorithm:

Compute Pairwise Distances:

Calculate the genetic distance between all sequences (e.g., based on nucleotide or protein differences).
Represent these distances in a matrix.

Find the Closest Pair:

Identify the two taxa with the smallest distance.
These are merged into a single cluster.

Recalculate Distances:

Compute the average distance between the new cluster and all remaining taxa.

Repeat Until All Taxa Are Clustered:

Continue merging the closest groups and updating the distance matrix.
The process continues until a single rooted tree is formed.

Advantages of UPGMA:

Simple and fast.
Suitable for **clock-like evolution** (constant mutation rates).
Useful for hierarchical clustering in bioinformatics.

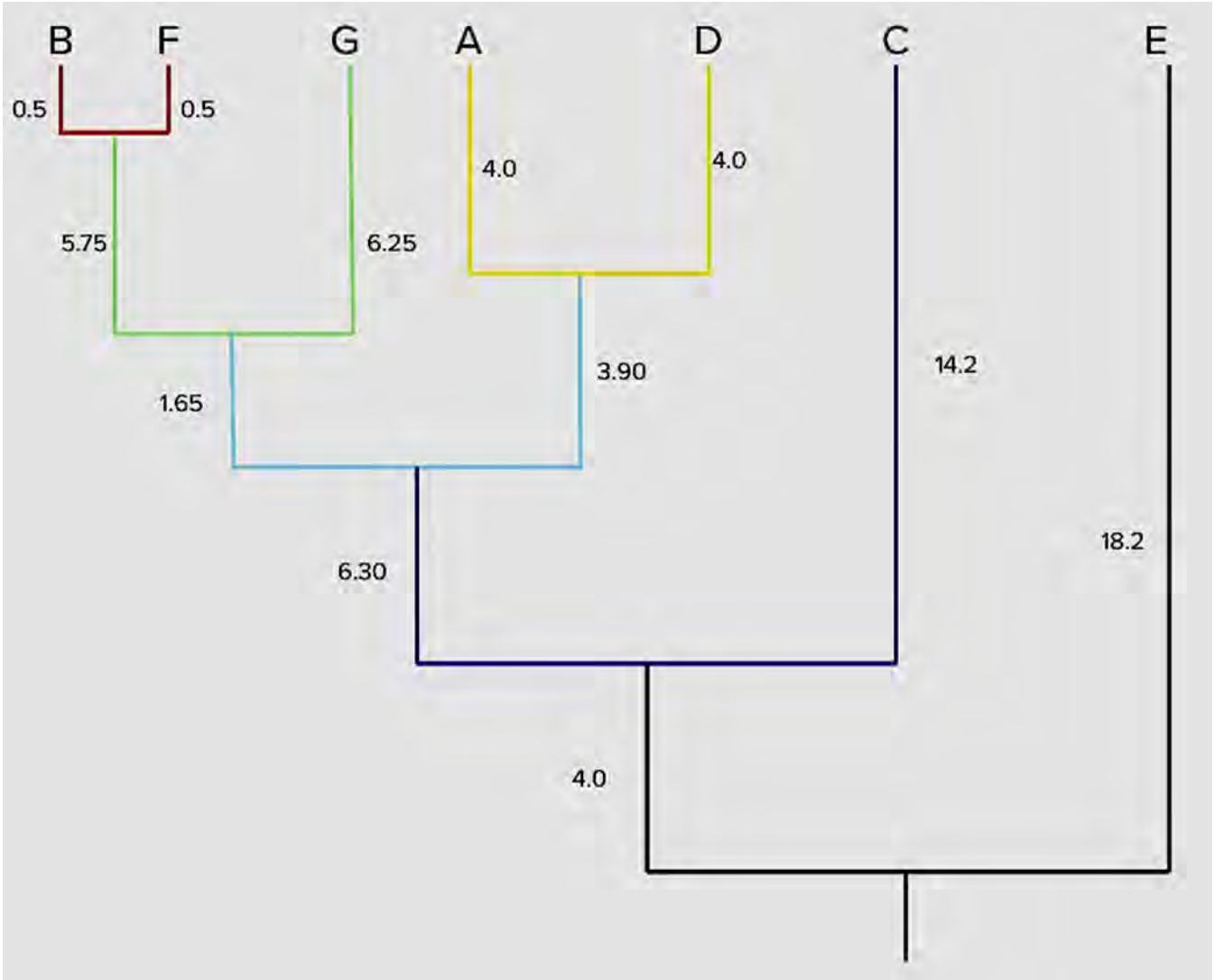
Limitations of UPGMA:

Assumes **equal rates of evolution** across lineages, which is often unrealistic.
Less accurate when evolutionary rates vary.
May produce incorrect tree topologies if the molecular clock assumption does not hold.

UPGMA (Unweighted Pair Group Method with Arithmetic Mean)

	A	B	C	D	E	F	G
A	0.0						
B	19.00	0.0					
C	27.00	31.00	0.0				
D	8.00	18.00	26.00	0.0			
E	33.00	36.00	41.00	31.00	0.0		
F	18.00	1.00	32.00	17.00	35.00	0.0	
G	13.00	13.00	29.00	14.00	28.00	12.00	0.0

Distance matrix



	S	T	U	V
S	0			
T	2	0		
U	5	5	0	
V	6	6	6	0

→

	ST	U	V
ST	0		
U	5	0	
V	6	6	0

→

	STU	V
STU	0	
V	6	0

→

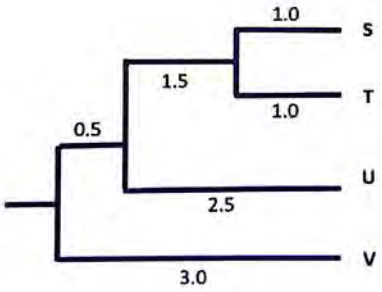
	STUV
STUV	0

S and T are the closest (distance is least, i.e. 2). Hence, S and T are clustered (ST) and matrix is recalculated

ST are closest to U (distance is least, i.e. 5). Hence, S, T, and U are clustered (STU) and matrix is recalculated

Clustering process is continued till the matrix contains only one cluster

Calculation of distance is done as: d(ST, U) = [d (S, U) + d(T, U)]/2 = (5+5)/2 = 5



Sequential clustering in the UPGMA method and the resultant phylogenetic tree

Neighbor-Joining (NJ) Method

Overview:

A distance-based method for constructing phylogenetic trees.

Does **not** assume a molecular clock (unlike UPGMA).

Produces an **unrooted tree**, meaning it does not assume a common ancestor unless an outgroup is provided.

Efficient for large datasets and widely used in molecular phylogenetics.

Steps in the Neighbor-Joining Algorithm:

Compute Pairwise Distances:

Create a distance matrix based on genetic differences between sequences.

Calculate a Neighbor-Joining (NJ) Matrix:

Adjust the distances to correct for varying evolutionary rates.

Helps identify the most closely related pair.

Find the Closest Pair of Taxa:

The two taxa with the smallest adjusted distance are joined together as a new node.

Recalculate Distances:

Compute new distances between the newly formed cluster and the remaining taxa.

Repeat Until a Single Tree is Formed:

The process continues until all taxa are connected into an unrooted tree.

Advantages of NJ Method:

Faster than other tree-building methods (e.g., Maximum Likelihood or Bayesian Inference).

Does **not require a molecular clock assumption**, making it more flexible.

Works well for **large datasets**.

Often used for preliminary phylogenetic analyses before more computationally intensive methods.

Limitations of NJ Method:

May produce incorrect trees if distance estimates are inaccurate.

Sensitive to **errors in distance calculation**.

Does not directly model evolutionary processes (relies on distance, not character states).

2. Tree Searching Approach

Maximum Parsimony (MP) Method

Overview:

A character-based method for constructing phylogenetic trees.

Goal: Find the tree that requires the **fewest evolutionary changes** (simplest explanation).

Based on **Occam's Razor** – the simplest explanation is preferred.

Works well for small datasets with low mutation rates.

Search Methods in Maximum Parsimony

Since there are many possible tree topologies, different strategies are used to find the most parsimonious tree:

1. Exhaustive Search

Evaluates all possible trees to find the one with the least number of changes.

Best for small datasets (up to ~10 taxa) since the number of trees grows exponentially.

Guarantees the **most accurate** (optimal) tree.

Limitation: Computationally expensive for large datasets.

2. Heuristic Search

Used when the number of possible trees is too large for an exhaustive search.

Finds an approximation of the best tree without checking all possibilities.

Common heuristic methods:

- **Stepwise Addition:** Starts with a simple tree and adds taxa one by one.
- **Branch Swapping (Tree Rearrangement):** Moves branches to explore different tree topologies.

Common algorithms for swapping:

Nearest-Neighbor Interchange (NNI): Swaps small parts of the tree.

Subtree Pruning and Regrafting (SPR): Cuts and reattaches subtrees in different locations.

Tree Bisection and Reconnection (TBR): Splits the tree and reconnects branches in new ways.

Limitation: May not always find the optimal tree.

3. Branch and Bound Search

A mix between exhaustive and heuristic methods.

Systematically eliminates tree topologies that **cannot** be more parsimonious than the best found so far.

More efficient than exhaustive search but still guarantees the optimal tree.

Works well for **moderate-sized datasets** (~15–20 taxa).

Advantages of Maximum Parsimony

Simple and intuitive.

Works well when evolutionary changes are rare.

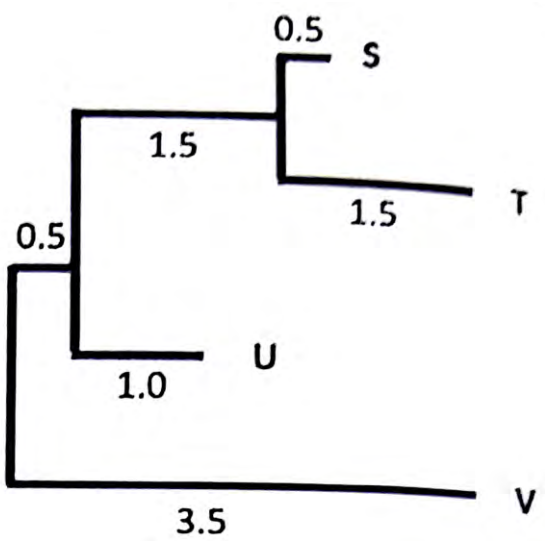
Does not require a specific evolutionary model.

Limitations of Maximum Parsimony

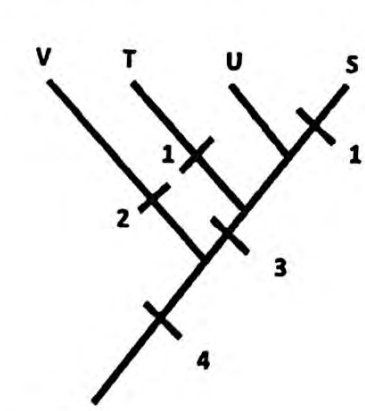
Long-branch attraction bias – may incorrectly group distantly related species if they have high mutation rates.

Computationally intensive for large datasets.

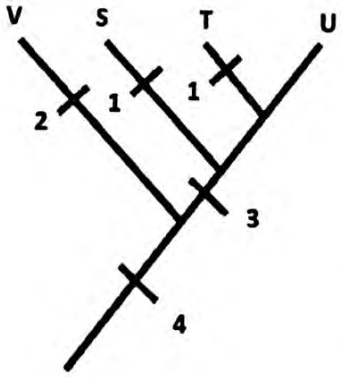
May not always give the correct tree when homoplasy (convergent evolution) is present.



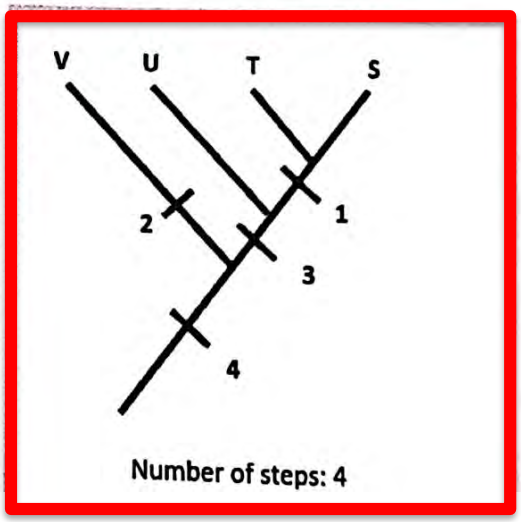
A Neighbor-joining tree showing evolutionary distances on the branches



Number of steps: 5



Number of steps: 5



Number of steps: 4

The most parsimonious tree from the three possible rooted trees of four taxa is the one having least number of evolutionary steps

Maximum Likelihood (ML) Method in Phylogenetics

Overview

A statistical method for constructing phylogenetic trees.

Finds the tree that has the **highest probability** of producing the observed data given an evolutionary model.

Uses **mathematical models** to account for different rates of evolution across sites and lineages.

Steps in Maximum Likelihood Analysis

Choose an Evolutionary Model

Defines how DNA, RNA, or protein sequences change over time.

Common models: Jukes-Cantor (JC), Kimura 2-Parameter (K2P), General Time Reversible (GTR).

Evaluate Possible Trees

Calculates the likelihood (probability) of each possible tree based on the given sequence data and model.

Tree with the **highest likelihood** is chosen.

Optimize Branch Lengths

Adjusts branch lengths to maximize the likelihood of the tree fitting the data.

Assess Tree Reliability

Uses **Bootstrap analysis** to check the confidence of the tree topology.

Advantages of Maximum Likelihood

More **accurate** than distance-based and parsimony methods.

Accounts for **different mutation rates** and complex evolutionary processes.

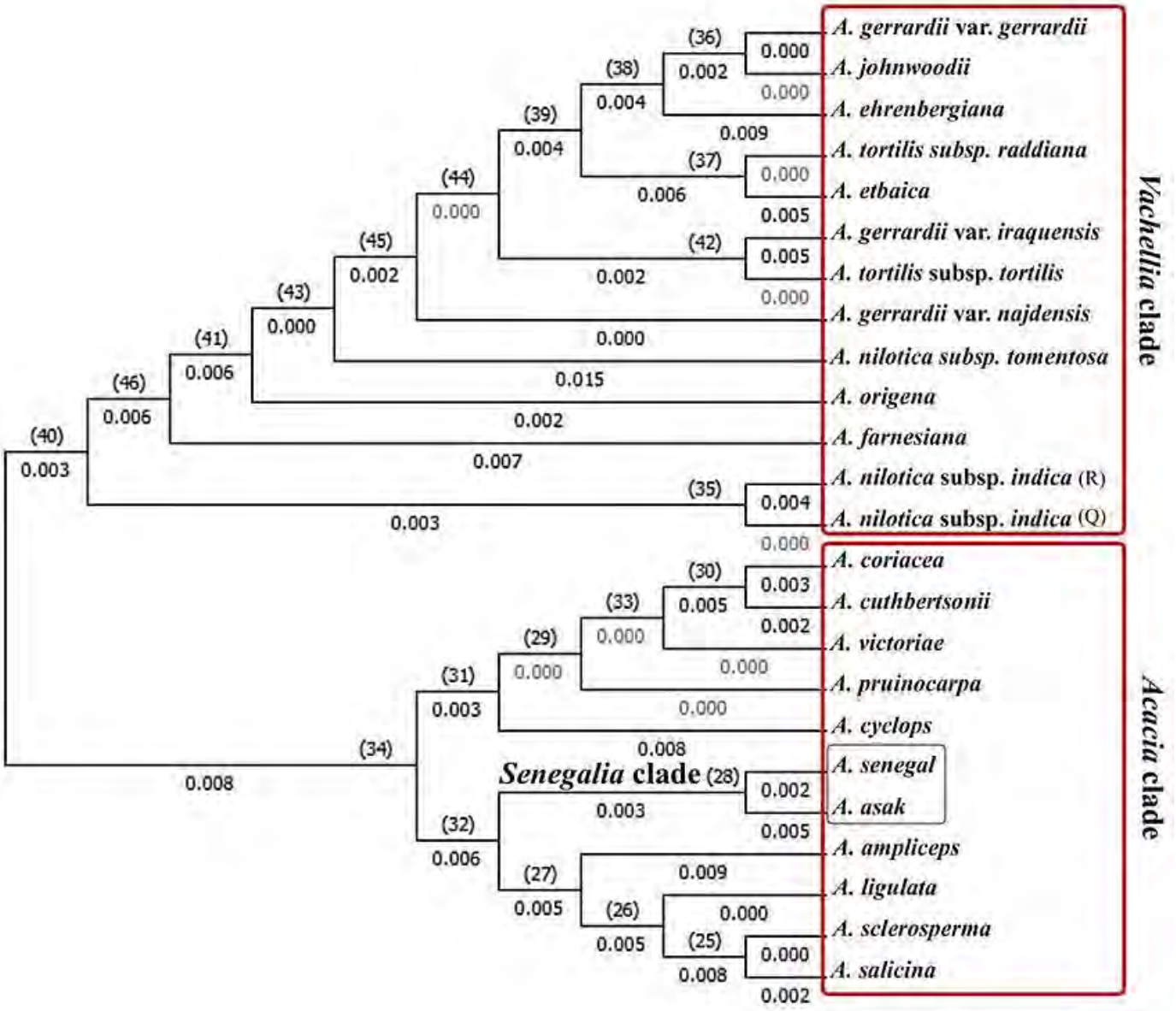
Provides **statistical support** for tree branches (e.g., bootstrap values).

Limitations of Maximum Likelihood

Computationally **slow** and intensive, especially for large datasets.

Requires selecting an **appropriate evolutionary model**.

More complex than simpler methods like **UPGMA** or **Neighbor-Joining**.



Phylogeny tree of 24 taxa of Acacia based on rbcL locus using Maximum Likelihood

Bayesian Inference (BI) in Phylogenetics

Overview

A statistical method for constructing phylogenetic trees.

Based on **Bayes' theorem**, which calculates the probability of a tree given the observed data and a prior model.

Uses **Markov Chain Monte Carlo (MCMC) sampling** to explore possible tree topologies.

Steps in Bayesian Inference Analysis

Choose an Evolutionary Model

Defines how DNA, RNA, or protein sequences change over time.

Common models: Jukes-Cantor (JC), Kimura 2-Parameter (K2P), General Time Reversible (GTR).

Set Prior Probabilities

Assigns prior beliefs about tree topology, branch lengths, and mutation rates.

Priors can be based on previous knowledge or set as uniform (equal probability for all trees).

Run MCMC Simulation

Generates many possible trees by slightly modifying branch lengths and topology in each step.

Trees with higher likelihoods are sampled more often.

Construct the Consensus Tree

After many generations, the trees are summarized into a single **posterior probability tree**.

Each branch is assigned a probability value (confidence level).

Advantages of Bayesian Inference

Provides **posterior probabilities**, which indicate strong statistical support for tree branches.

More **accurate** than methods like **Maximum Parsimony**.

Can incorporate **prior knowledge** into the analysis.

Efficient in handling **complex evolutionary models**.

Limitations of Bayesian Inference

Computationally **slow** due to MCMC sampling.

Requires careful selection of **priors**, which can affect results.

Results may vary depending on the number of generations and burn-in period.

Generating DNA Sequence Data in Phylogenetic Analysis

Overview

DNA sequence data is the primary source for constructing phylogenetic trees.

Sequences are collected from species and compared to determine evolutionary relationships.

Data can be generated **experimentally** (e.g., DNA sequencing) or **simulated** for computational analysis.

1. Methods for Generating DNA Sequence Data

Experimental Methods (Lab-based)

PCR (Polymerase Chain Reaction) – Amplifies specific DNA regions.

Sanger Sequencing – Traditional method for obtaining DNA sequences.

Next-Generation Sequencing (NGS) – High-throughput sequencing for large datasets.

Computational Simulations

Used to generate artificial DNA sequences for testing phylogenetic methods.

Tools include **Seq-Gen**, **Evolver**, and **INDELible**.

Databases for Real DNA Sequences

NCBI GenBank – Stores publicly available DNA sequences.

EMBL-EBI – European sequence archive.

DDBJ – Japanese DNA database.

2. Key Steps in DNA Data Generation for Phylogenetics

Sequence Collection

Extract DNA from organisms.

Amplify and sequence specific genes (e.g., mitochondrial DNA, rRNA).

Sequence Alignment

Use tools like **ClustalW, MUSCLE, or MAFFT** to align sequences.

Ensures homologous sites are compared.

Model Selection

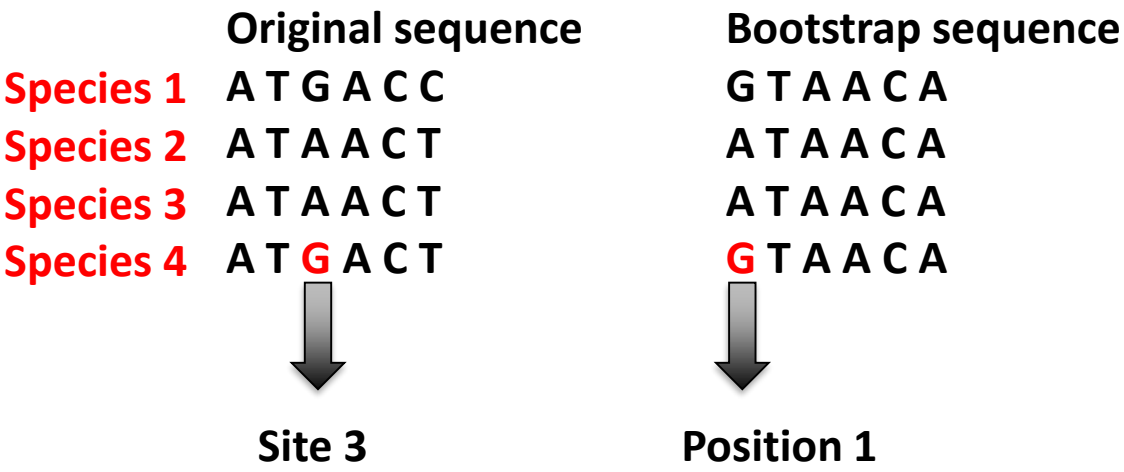
Choose the best **substitution model** (e.g., Jukes-Cantor, GTR).

Tree Construction

Use Maximum Parsimony (MP), Maximum Likelihood (ML), or Bayesian Inference (BI).

Bootstrap Analysis

Evaluates the statistical support of branches.



Site 3 is placed at position one in bootstrap sequence and next five randomly chosen sties; 2, 1, 1, 5 4 are placed in next five position

Steps in Acquiring DNA Sequence Data

Acquiring DNA sequence data involves **laboratory techniques** for extracting, amplifying, sequencing, and analyzing genetic material. Below are the key steps:

1. Sample Collection

Collect biological samples (e.g., blood, saliva, tissue, plant leaves).
Preserve samples using proper storage methods (e.g., freezing at -80°C).

2. DNA Extraction

Isolate DNA from cells using extraction kits or chemical methods.
Common methods include:

- **Phenol-Chloroform Extraction** – Organic solvent-based purification.
- **Silica Column-Based Extraction** – Used in commercial kits (e.g., Qiagen).
- **Magnetic Beads Method** – High-efficiency extraction using magnetic particles.

3. DNA Quantification & Quality Check (QC)

Assess DNA concentration and purity using:

- **Nanodrop Spectrophotometer** – Measures absorbance at 260/280 nm.
- **Qubit Fluorometer** – Provides more accurate DNA concentration.
- **Agarose Gel Electrophoresis** – Checks DNA integrity.

4. DNA Amplification (PCR - Polymerase Chain Reaction)

Amplify specific regions of DNA using **PCR primers**.
Commonly targeted genes:

- **Mitochondrial DNA (COI, Cyt-b)** – Used in species identification.
- **Ribosomal RNA (16S, 18S, ITS)** – Used in phylogenetic studies.
- **Nuclear Genes** – Used in genome-wide studies.

5. DNA Sequencing

Determines the **nucleotide order** (A, T, C, G) in the DNA fragment.

Sequencing technologies:

- **Sanger Sequencing** – Best for short DNA fragments (~800 bp).
- **Next-Generation Sequencing (NGS)** – High-throughput sequencing (e.g., Illumina, PacBio).
- **Third-Generation Sequencing (Oxford Nanopore, PacBio SMRT)** – Long-read sequencing.

6. Sequence Data Quality Control (QC)

Check raw sequences for errors and artifacts using:

- **Phred Score** – Measures sequencing quality.
- **FastQC Software** – Checks GC content, adapter contamination, and sequence quality.
- **Trimmomatic** – Removes low-quality reads and adapters.

7. Sequence Alignment & Processing

Align sequences using software like:

- **ClustalW**, **MUSCLE**, or **MAFFT** (for multiple sequence alignment).
- **BLAST (Basic Local Alignment Search Tool)** – Identifies sequences by comparison with known databases.

8. Phylogenetic Tree Construction

Select an evolutionary model (e.g., Jukes-Cantor, GTR).

Build trees using:

- **Maximum Parsimony (MP)** – Finds the simplest evolutionary pathway.
- **Maximum Likelihood (ML)** – Selects the most probable tree.
- **Bayesian Inference (BI)** – Estimates probability distributions for trees.

9. Data Storage & Submission

Store sequence data in online repositories:

- **NCBI GenBank**
- **EMBL-EBI**
- **DDBJ (DNA Data Bank of Japan)**

10. Interpretation & Publication

Analyze phylogenetic relationships.

Publish results in scientific journals and databases.

CTAB Method for DNA Extraction

The **CTAB (Cetyltrimethylammonium Bromide) method** is a widely used protocol for extracting high-quality DNA, especially from plant tissues rich in **polysaccharides and secondary metabolites**. It helps in obtaining **pure DNA** suitable for **PCR, sequencing, and phylogenetic analysis**.

Steps in the CTAB DNA Extraction Method

1. Sample Collection and Preparation

Collect **fresh or frozen plant tissue** (e.g., leaves, seeds, roots).
Freeze the sample in **liquid nitrogen** and grind into a fine powder using a **mortar and pestle**.
Transfer the powder to a **sterile microcentrifuge tube**.

2. Lysis (Cell Disruption) using CTAB Buffer

Add **CTAB extraction buffer** (preheated to **65°C**) to the ground tissue.

CTAB Buffer Composition:

- **2% CTAB** (detergent that breaks cell membranes).
- **100 mM Tris-HCl (pH 8.0)** (maintains pH stability).
- **1.4 M NaCl** (removes proteins and polysaccharides).
- **20 mM EDTA** (binds divalent ions to inhibit nucleases).
- **1% β-mercaptoethanol** (removes tannins and phenolics).

Incubate at **65°C for 30 minutes** to lyse cells and release DNA.

3. Removal of Proteins and Contaminants

Add an equal volume of **chloroform:isoamyl alcohol (24:1)** and mix by gentle inversion.
Centrifuge at **10,000–14,000 rpm for 10 minutes** to separate the layers.
Carefully transfer the **upper aqueous phase** (containing DNA) to a fresh tube.

4. DNA Precipitation

Add **0.6–1 volume of ice-cold isopropanol** to the aqueous phase.

Incubate at **-20°C for 30 minutes** to allow DNA precipitation.

Centrifuge at **10,000–14,000 rpm for 10 minutes** to pellet the DNA.

5. DNA Washing

Discard the supernatant and wash the DNA pellet with **70% ethanol**.

Centrifuge at **10,000 rpm for 5 minutes**, discard ethanol, and air-dry the pellet.

6. DNA Resuspension

Dissolve the dried DNA pellet in **TE buffer (Tris-EDTA)** or **nuclease-free water**.

Incubate at **37°C for 10 minutes** to ensure complete dissolution.

7. DNA Quality Check

Measure **DNA concentration** using **Nanodrop** or **Qubit**.

Check **DNA integrity** using **agarose gel electrophoresis**.

Advantages of the CTAB Method

Effective for plant tissues with **high polysaccharide content**.

Produces **high-yield and high-purity DNA**.

Suitable for **PCR, sequencing, and phylogenetic analysis**.

Limitations

Requires **hazardous chemicals** (chloroform, β-mercaptoethanol).

Time-consuming compared to commercial DNA extraction kits.

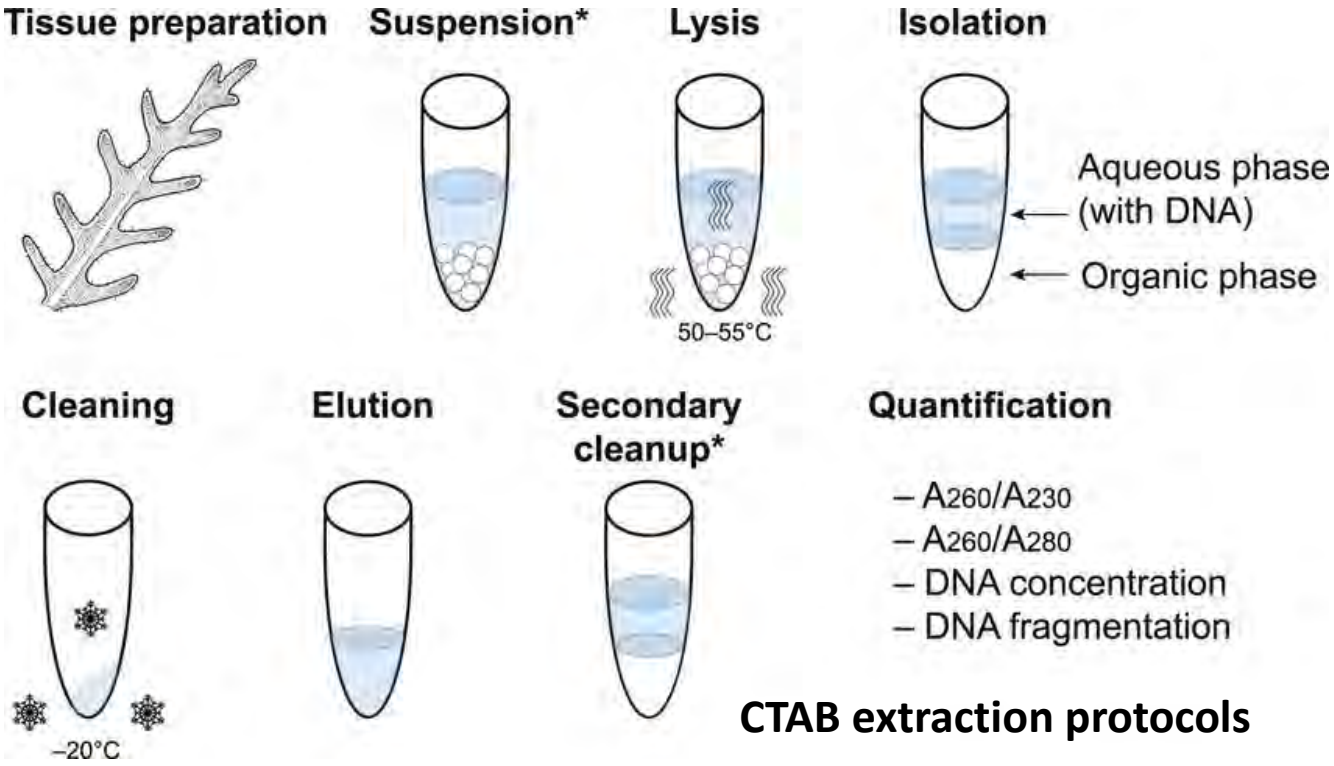
Applications of CTAB-Extracted DNA

Molecular phylogenetics (species evolution studies).

PCR and sequencing (Sanger, NGS).

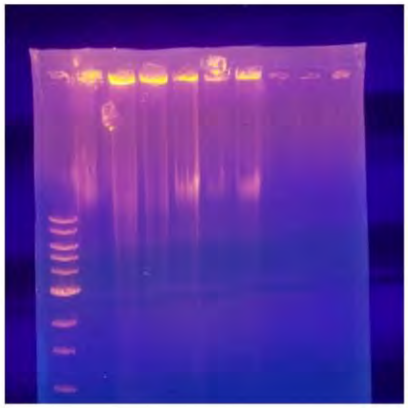
Genetic diversity studies (population genetics).

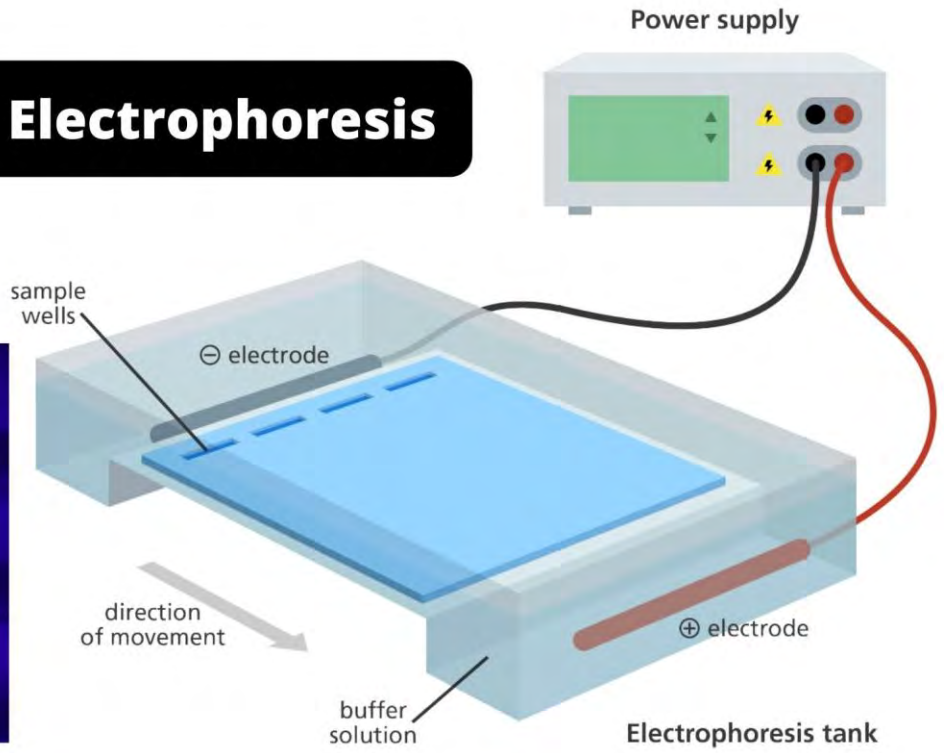
GMO and pathogen detection.



CTAB extraction protocols

Agarose Gel Electrophoresis





PCR (Polymerase Chain Reaction)

PCR is a technique used to **amplify specific DNA sequences** from a small DNA sample. It is widely used in **molecular biology, genetics, forensic science, and phylogenetics**.

Components of PCR

Template DNA – The DNA sample containing the target sequence.

Primers – Short DNA sequences that bind to specific regions of the template.

DNA Polymerase – Enzyme (e.g., Taq polymerase) that synthesizes new DNA strands.

dNTPs (Deoxynucleotide Triphosphates) – Building blocks for DNA synthesis (A, T, C, G).

Buffer Solution – Maintains the optimal pH and salt conditions for the reaction.

Mg^{2+} Ions – Cofactor required for DNA polymerase activity.

Steps in PCR

Denaturation (94–98°C, 30 sec)

DNA strands separate into single strands due to heat.

Annealing (50–65°C, 30 sec)

Primers bind to their complementary sequences on the template DNA.

Extension (72°C, 30–60 sec)

DNA polymerase extends the primers by adding dNTPs to synthesize new DNA strands.

Repeat for 25–40 Cycles

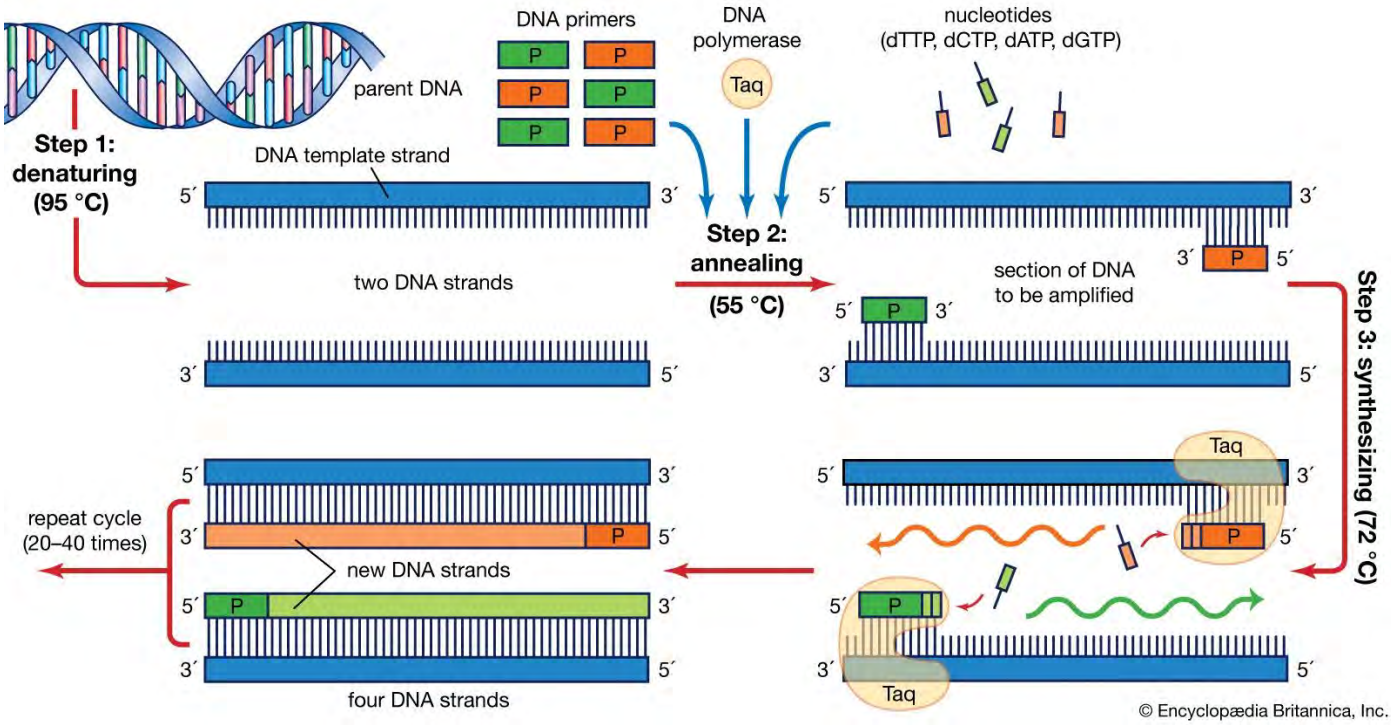
Each cycle **doubles** the amount of DNA, leading to exponential amplification.

Final Extension (72°C, 5–10 min)

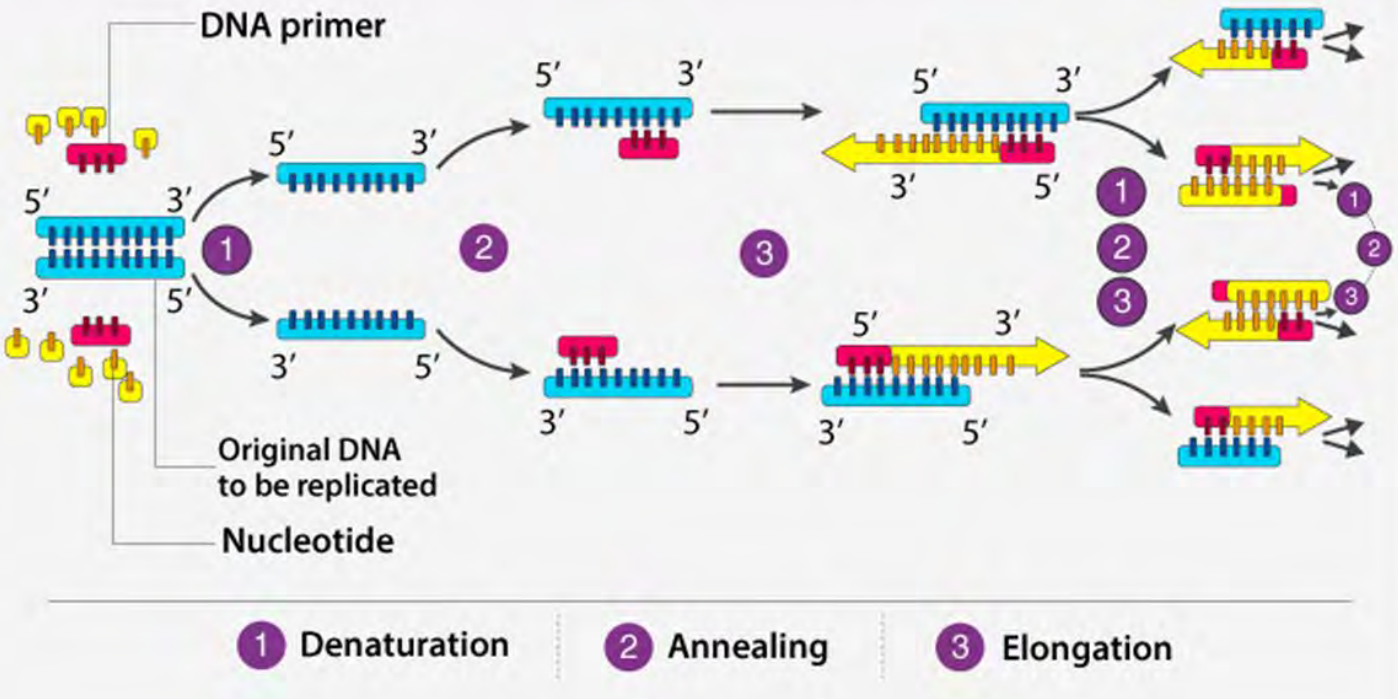
Ensures all DNA strands are fully extended.

Cooling (4°C, Hold)

Stops the reaction and preserves the amplified DNA.



POLYMERASE CHAIN REACTION (PCR)



Gel Electrophoresis

Gel electrophoresis is a technique used to **separate DNA, RNA, or proteins** based on their **size and charge** using an electric field. It is commonly used to **analyze PCR products, check DNA purity, and verify genetic material** in molecular biology studies.

Components of Gel Electrophoresis

Agarose Gel – A porous gel that acts as a molecular sieve for DNA movement.

Electrophoresis Buffer – Provides ions for conducting electricity (e.g., TAE or TBE buffer).

DNA Samples – DNA fragments mixed with a loading dye for visualization.

DNA Ladder (Marker) – A reference set of DNA fragments with known sizes.

Loading Dye – Helps visualize DNA movement and adds density for easy loading.

Ethidium Bromide (EtBr) or SYBR Green – DNA staining agents for UV visualization.

Power Supply and Gel Box – Generates an electric field to move DNA through the gel.

Steps in Gel Electrophoresis

Prepare Agarose Gel

Mix agarose powder with buffer (e.g., TAE/TBE) and heat until dissolved.

Pour into a gel tray and insert a comb to create wells.

Allow it to solidify.

Load DNA Samples

Mix DNA with **loading dye** and pipette into the gel wells.

Load a **DNA ladder** in one well for size comparison.

Run the Gel

Submerge gel in buffer inside an electrophoresis chamber.

Apply an **electric current (typically 80–150V)**.

DNA moves from **negative (-) to positive (+) electrode** since it is negatively charged.

Visualize DNA

Stain gel with **EtBr or SYBR Green**.

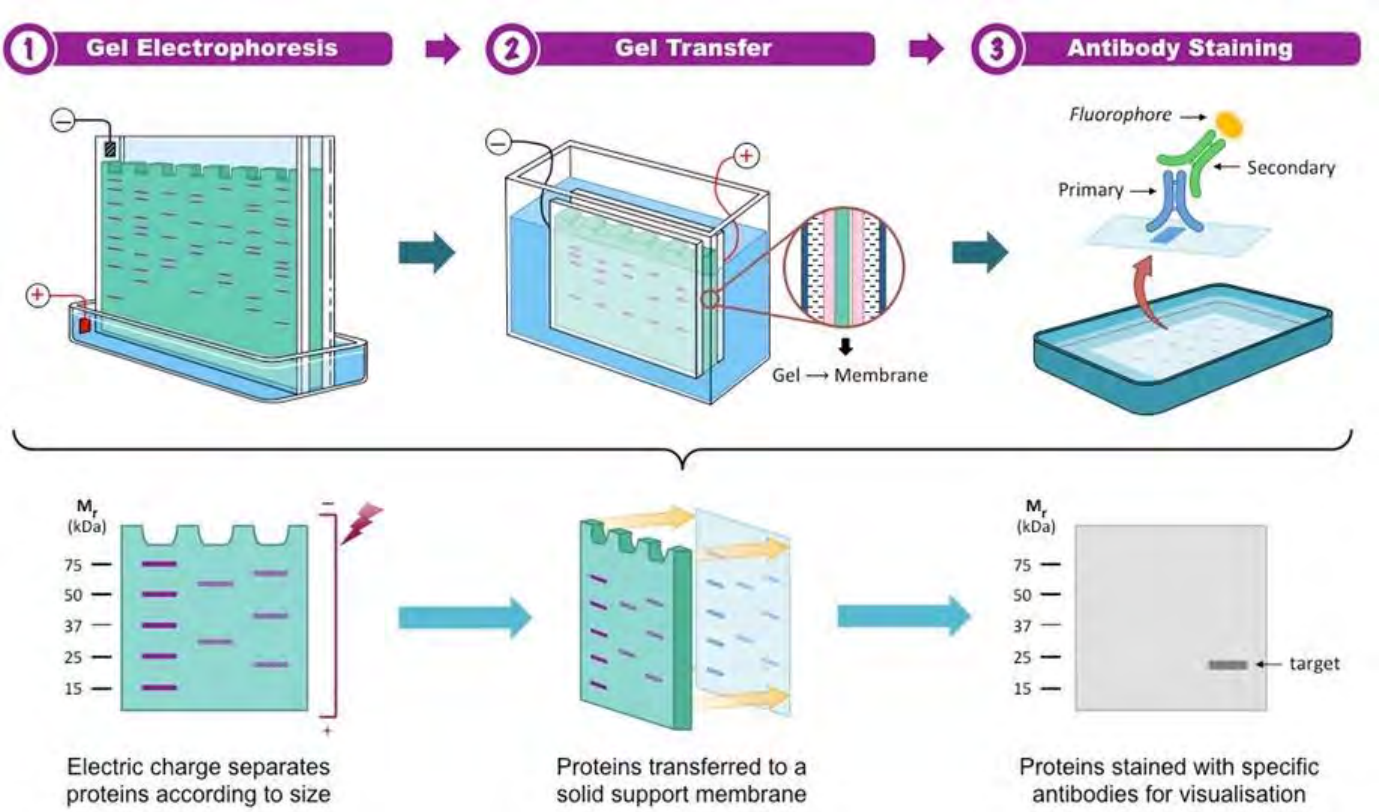
View under **UV light** to observe DNA bands.

Interpreting Gel Electrophoresis Results

Smaller DNA fragments move **faster** and travel **farther**.

Larger DNA fragments move **slower** and stay **closer to the wells**.

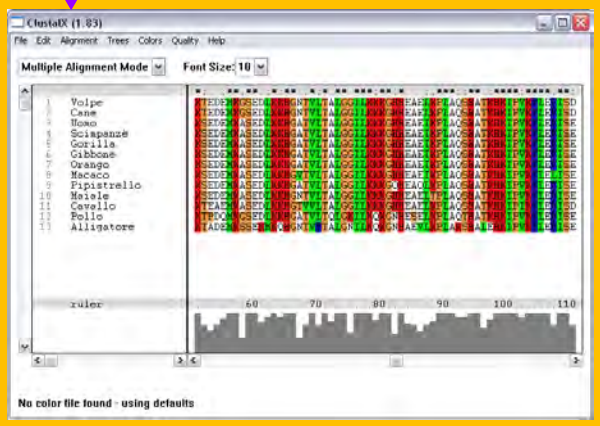
Compare unknown DNA bands with the **DNA ladder** to determine fragment sizes.



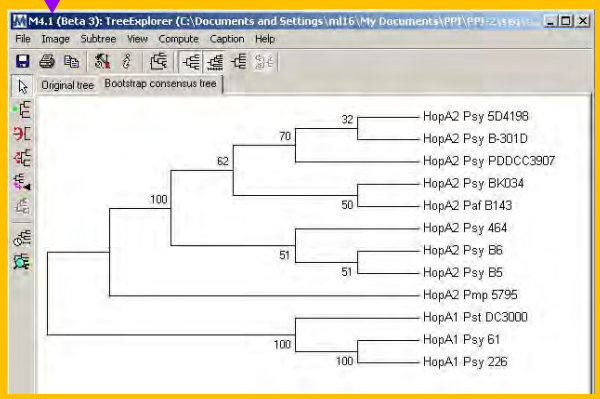
Major steps in Phylogenetic analyses

GTCGAAACCTGCATAGCAGAACGACCCGCGAACACGTTACACTACCAGGTGAGGGACGAGGGGTGCGCAA
GCTCCCCAAGTTTCAAACCCATGGTCGGGGACCACCCTTGGGTGGCCTCGTCCGAACAACGACCCCCCGG
CGCGGAATGCGCCAAGGAAATCAAACTGAACTGCACGCGTCCCCCCCGTTTGCGGGCGGCGGAAGCGTCT
TTCTAAAACACAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCA
TTAGGCCGAGGGCACGTCTGCCTGGGCGTCACACATCGCGTCGCCCCCCAACCCATCACTCCCTTGCGGG
AGTTGAGGCGGAGGGGCGGATAATGGCCTCCCGTGTCTCACCGCGCGGTTGGCCCAAATGCGAGTCCTTG
GCGATGGACGTCACGACAAGTGGTGGTTGTAAAAAGCCCTCTTCTCATGTCGTGCGGTGACCCGTCGCCA
GCAAAATCTCTCATGACCCTGTTGCGCCGAGCCTCGACGCGCGCTCCGACCGCGACCCC

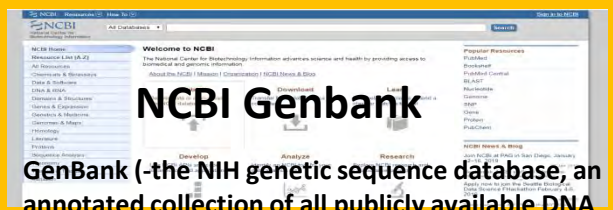
DNA sequence dataset preparation



DNA sequence alignment using ClustalX



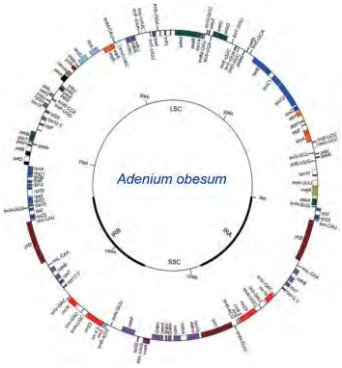
Molecular Phylogenetic analyses using MEGA

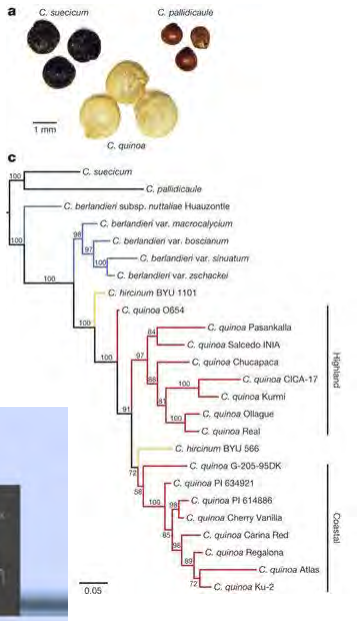


NCBI Genbank

GenBank (-the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences) has a very important role in molecular phylogeny and DNA barcoding.

Advances in next generation DNA sequencing and its application in systematics





Nanopore Portable DNA sequencer

❖Whole Chloroplast Genome Sequencing of *Adenium obesum*

- ❖**Chloroplast (cp) is a special subcellular organelle which contains the entire enzymatic machinery for photosynthesis.**
- ❖**Chloroplast contains its own small genome of 120–217 kb in size and 110-130 genes, consists of a circular double-stranded DNA.**
- ❖**The cp genome can be used to investigate molecular evolution and phylogenies.**
- ❖**The cp genomes are maternally inherited, which is beneficial in genetic engineering.**

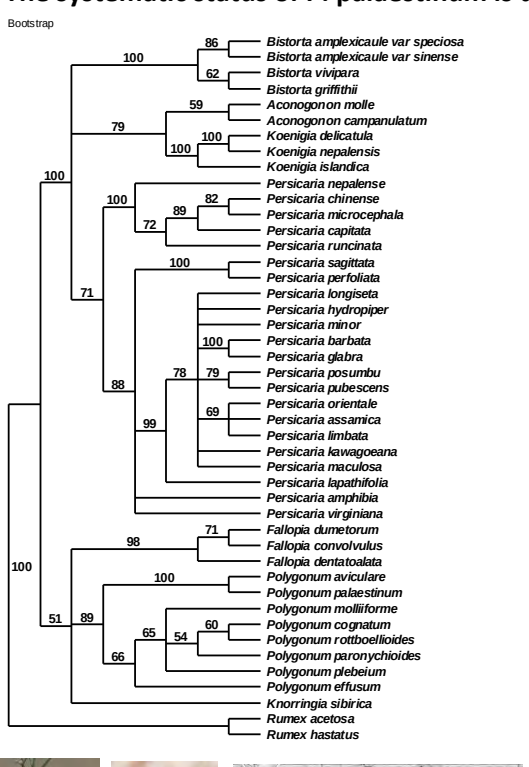
Whole genome sequencing

- ❖***Chenopodium quinoa* (quinoa) is a highly nutritious grain identified as an important crop** Jarvis 2017 Nature. ; 542(7641):307-312. **The genome of Chenopodium quinoa.**

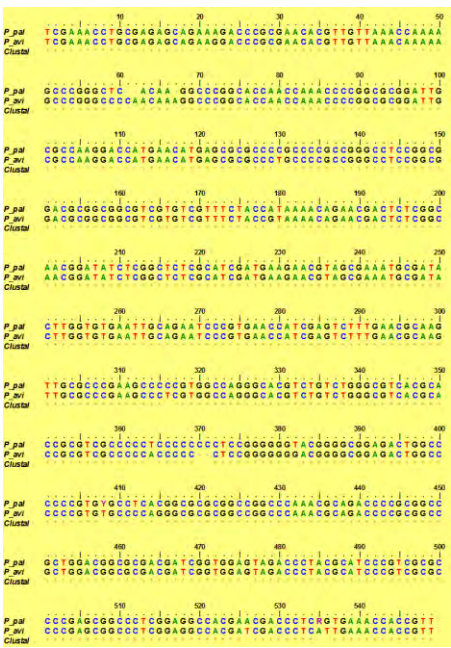
Molecular systematic studies on *Polygonum palaestinum* Zohary (polygonaceae) from Saudi Arabia using ITS sequences of nuclear ribosomal DNA

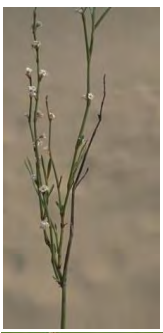
- **The taxonomy of the genus Polygonum is highly controversial because of diverse variation within species among the species has resulted into lack of consensus on taxonomic circumscription. Therefore, there is disagreement among the taxonomists that to which species should be retain within the genus Polygonum and to which species should be elevated to their own genus.**

- **The genus Polygonum in Saudi Arabia includes P. argyrocoleum Steud. ex Kunze, P. aviculare L. and P. palaestinum Zohary. Two out of these Polygonums of Saudi Arabia i.e. P. argyrocoleum and P. aviculare are common weed distributed throughout. The distribution of P. palaestinum is restricted to Harratal Harra area of Saudi Arabia.**
- **Decraene and Akeroyd (1988) have segregated Polygonum in the broad sense into two separate tribes, Polygoneae and Persicarieae.**
- **The systematic status of P. palaestinum is unresolved**



Polygonum palaestinum Zohary (polygonaceae)





Position in sequence alignment	*Polygonum aviculare*	*Polygonum palaestinum*
22	G	A
46	A	C
59	C	T
61	C	-
62	A	-
67	A	-
130	T	C
179	G	A
316	T	C
364	A	T
370	-	C
371	-	C
382	G	T
408	T	Y C T
413	C	T
415	G	C
526	T	A
535	A	R A G
536	T	G

Advance Experimental Taxonomy

Chapter 6- Molecular Markers and DNA sequences in Taxonomic study

Introduction

Traditional taxonomy relied on **morphology** and **phenotypic characteristics**.

Molecular markers and DNA sequences provide **precise and reliable identification** of species.

Essential in **phylogenetics, evolutionary studies, and species conservation**.

What are Molecular Markers?

Specific DNA sequences used to study **genetic variation**.

Help in identifying **species, subspecies, and populations**.

Useful in **phylogenetics, population genetics, and conservation biology**.

Types of Molecular Markers

- **RFLP (Restriction Fragment Length Polymorphism)**
- **RAPD (Random Amplified Polymorphic DNA)**
- **AFLP (Amplified Fragment Length Polymorphism)**
- **SSR (Simple Sequence Repeats) / Microsatellites**
- **SNP (Single Nucleotide Polymorphism)**

DNA Sequences in Taxonomy

Barcoding Genes used in species identification:

COI (Cytochrome c Oxidase I) – Animals

rbcL & matK – Plants

ITS (Internal Transcribed Spacer) – Fungi

Whole Genome Sequencing for deeper evolutionary insights.

DNA Barcoding in Taxonomy

A **standardized method** for species identification.

Uses short, specific DNA sequences.

Helps in identifying **cryptic species** and discovering new ones.

Advantages of Molecular Markers

- **High accuracy** in species identification
- Helps in **resolving complex taxonomic relationships**
- Works for **all life stages** (seeds, larvae, etc.)
- Useful in **conservation biology and forensic studies**

Challenges & Limitations

- Requires **advanced lab techniques & expertise**
- **Expensive & time-consuming**
- Risk of **contamination & errors in sequencing**
- **Database limitations** for rare or new species

Applications in Taxonomy

Phylogenetics – Understanding evolutionary relationships.

Conservation Biology – Identifying endangered species.

Ecology – Studying biodiversity and ecosystem changes.

Agriculture – Crop and pest identification.

Types of Molecular Markers

Restriction Fragment Length Polymorphism (RFLP):

Introduction

RFLP is a molecular technique used to detect **variations in DNA sequences**.

It helps in **species identification, genetic diversity analysis, and evolutionary studies**.

One of the first DNA-based markers used in taxonomy and phylogenetics.

Principle of RFLP

RFLP works by cutting **genomic DNA** with **restriction enzymes** at specific sequences.

Differences in DNA sequences cause **variations in fragment sizes**.

These fragments are separated using **gel electrophoresis**.

A labeled **probe hybridizes** with the DNA fragments to detect polymorphisms.

Key Concept:

If two individuals have **different sequences**, restriction enzymes cut their DNA differently, leading to **different banding patterns** on a gel.

This helps in identifying **species, genetic relationships, and mutations**.

Steps Involved in RFLP

1. DNA Extraction

Isolate DNA from plant cells.

2. DNA Digestion (Restriction Enzyme Treatment)

Use **specific restriction enzymes** to cut DNA at recognition sites.

3. Gel Electrophoresis

Separate DNA fragments based on size using **agarose gel electrophoresis**.

4. Southern Blotting

Transfer DNA from gel to a **nylon or nitrocellulose membrane**.

5. Hybridization with Labeled Probe

A **radioactive or fluorescent probe** binds to complementary DNA sequences.

6. Detection of Fragment Patterns

Expose the membrane to X-ray film or a detector to visualize the bands.

7. Data Analysis & Species Identification

Compare fragment patterns to **identify species, genetic variations, and phylogenetic relationships**.

Applications of RFLP

Plant Taxonomy & Phylogenetics

Differentiates plant species and determines evolutionary relationships.

Genetic Diversity & Conservation Biology

Assesses genetic variation in endangered plant species.

Crop Improvement & Plant Breeding

Identifies genes for **disease resistance and desirable traits** in crops.

Authentication of Medicinal Plants

Ensures purity and identification of medicinal plant species.

Forensic Botany

Used in plant-based forensic investigations.

Advantages of RFLP

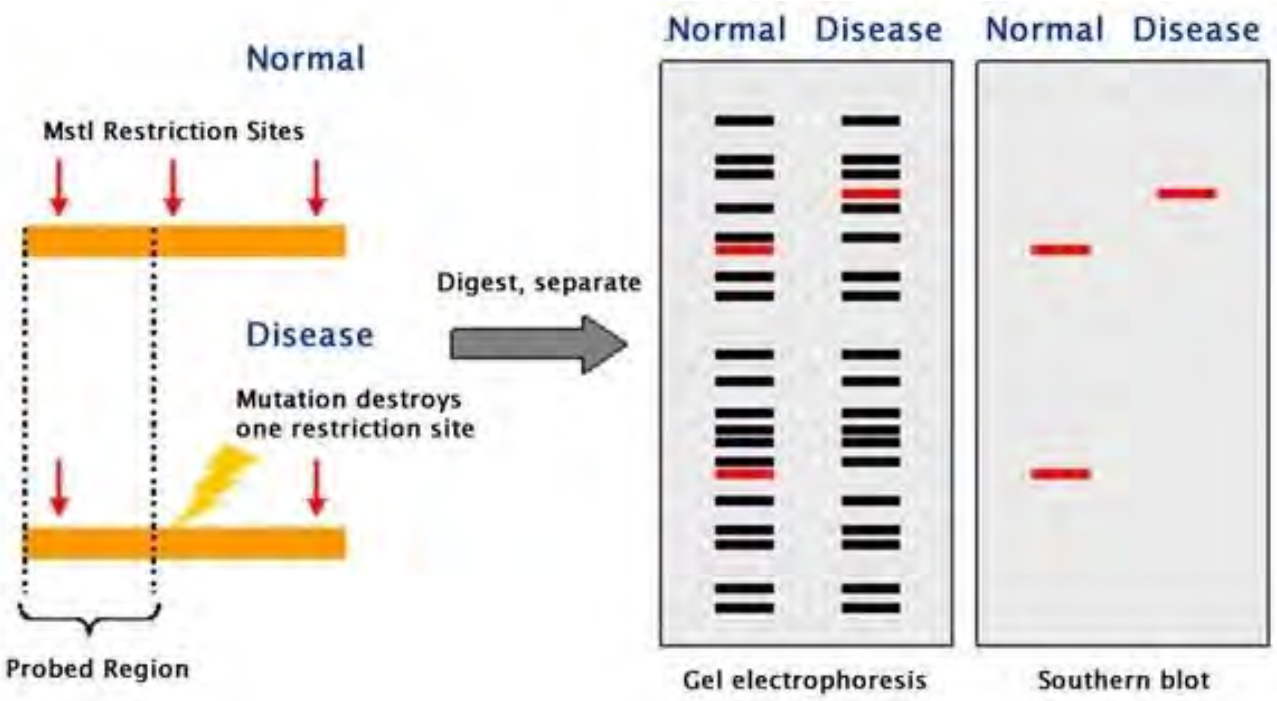
Highly Specific & Reliable – Detects actual **sequence variations** in DNA.
No Prior Sequence Knowledge Required – Works with **unknown sequences**.
Useful for Genetic Mapping – Helps in **linkage studies**.
Stable & Reproducible – Unlike RAPD, RFLP results are **consistent**.

Limitations of RFLP

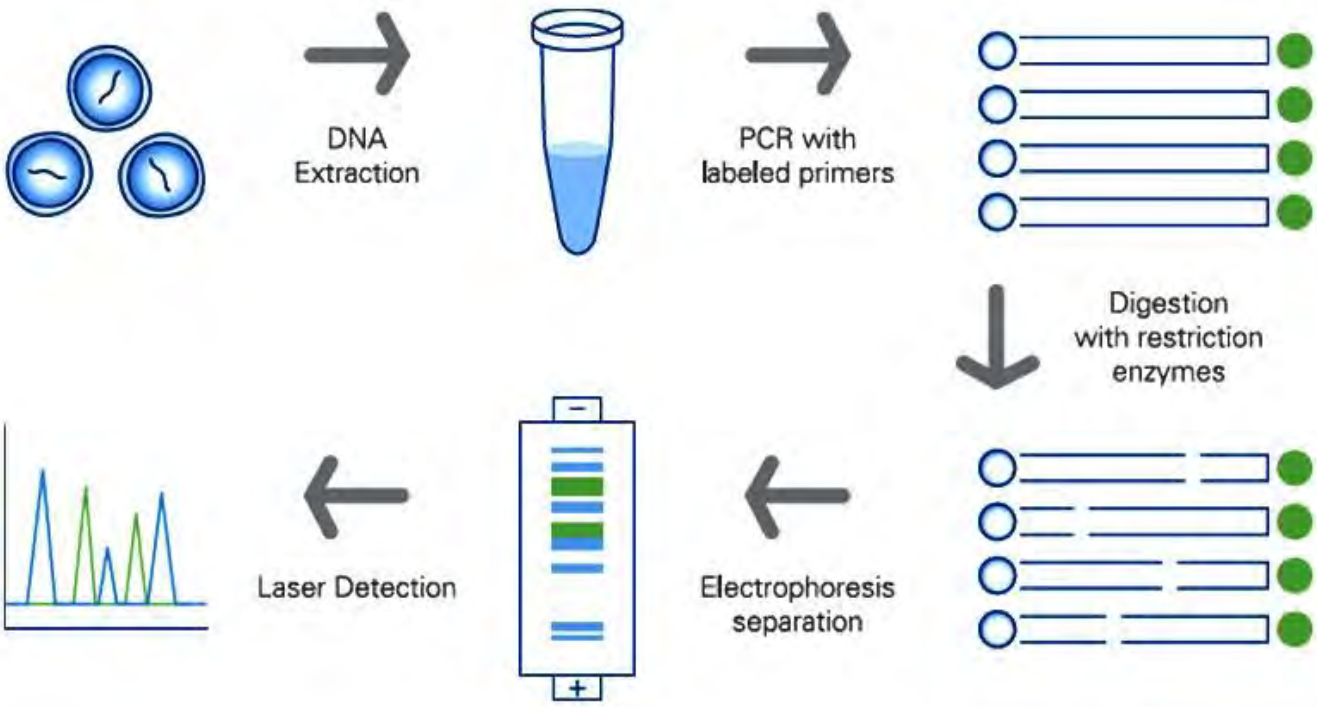
Time-Consuming & Labor-Intensive – Requires multiple steps (digestion, electrophoresis, blotting, hybridization).
Large DNA Sample Requirement – Needs a **high-quality DNA extract**.
Expensive & Requires Specialized Equipment – Southern blotting and labeled probes increase costs.
Limited Sample Throughput – Not suitable for large-scale screening compared to PCR-based markers.

Conclusion

RFLP is a **powerful technique** for plant taxonomy and genetic studies.
Despite its **limitations**, it remains valuable in **phylogenetics and plant breeding**.
Newer technologies like **SNP markers and Next-Generation Sequencing (NGS)** are replacing RFLP for faster and high-throughput analysis.



Restriction Fragment Length Polymorphism (RFLP)



RAPD (Random Amplified Polymorphic DNA)

Introduction

RAPD is a **PCR-based molecular marker technique** used to detect genetic variations.

It is a **dominant marker**, meaning it does not distinguish between heterozygous and homozygous loci.

Advantages: Fast, requires **small DNA samples**, and does not need prior sequence information.

Principle of RAPD

Uses **short, random primers** (10 nucleotides) to amplify **unknown DNA regions**.

The presence or absence of PCR-amplified DNA fragments creates a **unique banding pattern**.

Differences in banding patterns indicate **genetic polymorphism** among species or individuals.

Steps Involved in RAPD

1. DNA Extraction

Isolate DNA from plant tissues.

2. PCR Amplification

Use **random primers** to amplify multiple DNA regions.

3. Gel Electrophoresis

Separate amplified DNA fragments based on size using **agarose gel electrophoresis**.

4. Visualization of Bands

Stain the gel with **ethidium bromide or SYBR Green** and observe under UV light.

5. Data Analysis

Compare banding patterns to study **genetic diversity and relationships**.

Applications of RAPD

Plant Taxonomy & Phylogenetics – Identifies species and genetic relationships.

Genetic Diversity Studies – Assesses variation within and between plant populations.

Crop Improvement & Breeding – Selects genetically diverse parents for hybridization.

DNA Fingerprinting – Identifies plant varieties and cultivars.

Medicinal Plant Authentication – Ensures purity of herbal products.

Advantages of RAPD

Fast & Simple – No need for restriction digestion or Southern blotting.

Requires Small DNA Samples – Works with low-quality DNA.

No Prior Sequence Knowledge Needed – Can be used for any species.

Cost-Effective – Cheaper than RFLP and SSR markers.

Limitations of RAPD

Low Reproducibility – Results may vary due to **PCR conditions**.

Dominant Marker – Cannot distinguish between homozygous and heterozygous loci.

Sensitive to Contaminations – Requires careful handling to avoid errors.

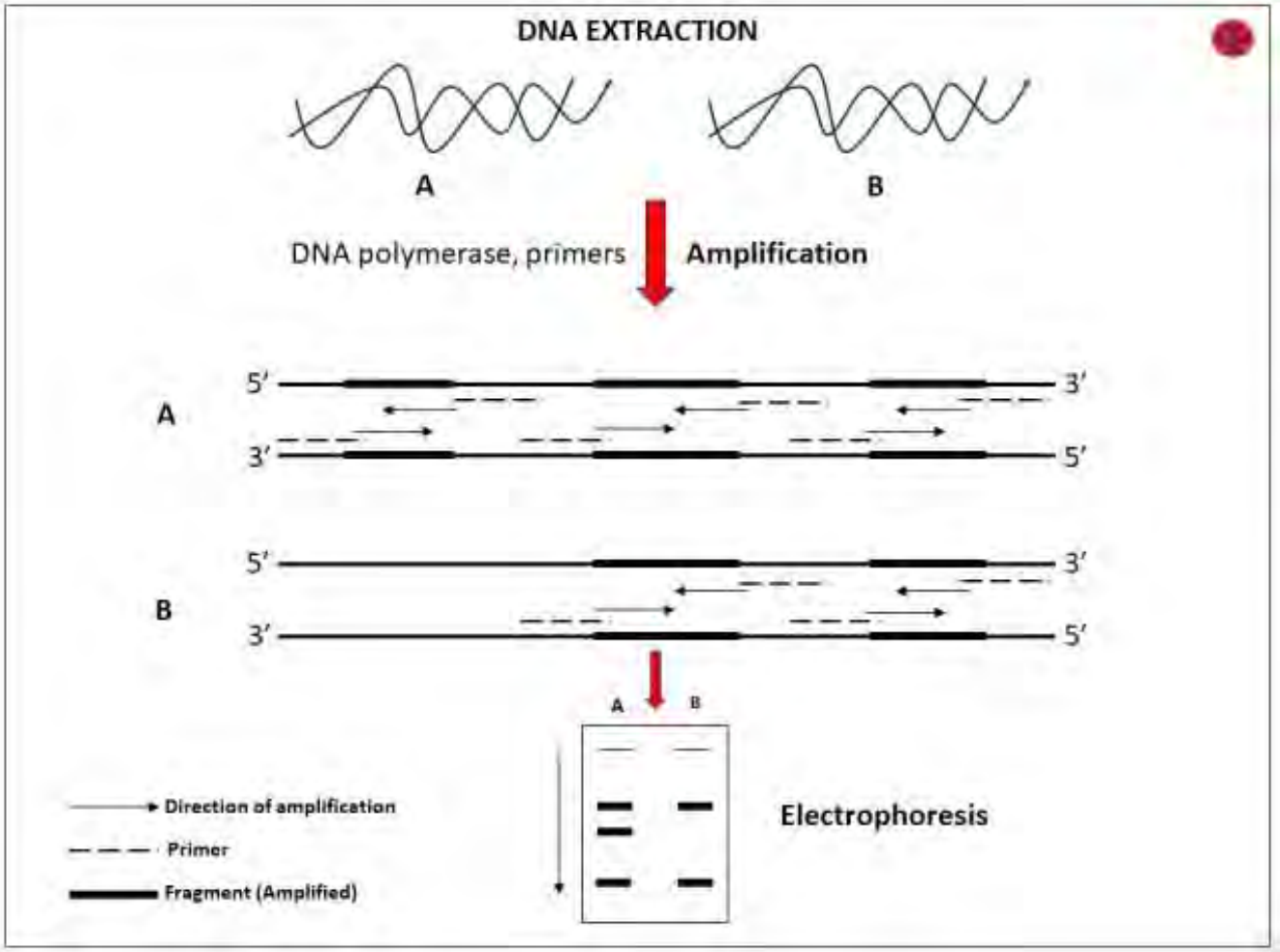
Limited Applicability for Fine-Scale Studies – Less informative than **SSR or SNP markers**.

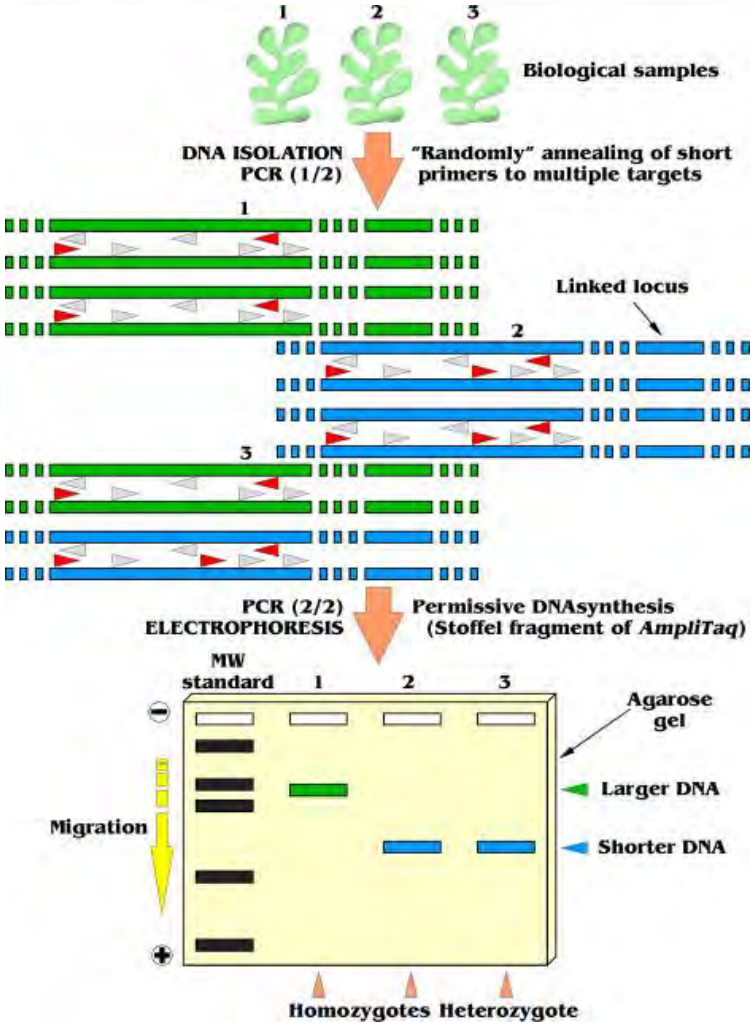
Conclusion

RAPD is a **quick and cost-effective** method for detecting genetic variation.

Despite **reproducibility issues**, it is widely used in **plant taxonomy, conservation biology, and crop improvement**.

Modern techniques like **AFLP, SSR, and SNP markers** provide more **accurate and reliable** results.


DNA EXTRACTION
A
B
DNA polymerase, primers
Amplification
5′
3′
A
B
Direction of amplification
Primer
Fragment (Amplified)
Electrophoresis


1
2
3
Biological samples
DNA ISOLATION
PCR (1/2)
"Randomly" annealing of short primers to multiple targets
Linked locus
PCR (2/2)
ELECTROPHORESIS
Permissive DNAsynthesis
(Stoffel fragment of AmpliTaq)
MW standard
Agarose gel
Larger DNA
Shorter DNA
Migration
Homozygotes
Heterozygote

AFLP (Amplified Fragment Length Polymorphism)

Introduction

AFLP is a **PCR-based molecular marker technique** that detects DNA polymorphisms.

Combines **RFLP (restriction enzyme digestion)** and **PCR amplification** for high-resolution genetic analysis.

Widely used in **plant taxonomy, phylogenetics, genetic mapping, and crop improvement**.

Principle of AFLP

AFLP involves **digesting genomic DNA** with **restriction enzymes**.

The resulting DNA fragments are **ligated to adapters** and selectively **amplified using PCR**.

The amplified products are separated via **gel electrophoresis** to detect genetic variations.

Steps Involved in AFLP

1. DNA Extraction

Isolate genomic DNA from plant tissue.

2. Restriction Digestion

Use **restriction enzymes** (e.g., EcoRI, MseI) to cut DNA at specific sites.

3. Adapter Ligation

Attach **synthetic adapters** to restriction fragments to serve as primer-binding sites.

4. Pre-Selective PCR Amplification

Amplify fragments using primers complementary to the adapters.

5. Selective PCR Amplification
Use primers with additional nucleotides to selectively amplify a subset of fragments.
6. Gel Electrophoresis & Visualization
Separate amplified DNA fragments using **polyacrylamide gel electrophoresis (PAGE)** or **capillary electrophoresis**.
7. Data Analysis
Compare banding patterns to determine **genetic diversity and species relationships**.

Applications of AFLP

Plant Taxonomy & Phylogenetics – Differentiates closely related species.
Genetic Diversity & Population Studies – Assesses variation within and between plant populations.
Crop Improvement & Breeding – Identifies markers linked to desirable traits.
DNA Fingerprinting – Used for plant variety protection.
Evolutionary Studies – Helps in reconstructing plant evolutionary histories.

Advantages of AFLP

High Polymorphism Detection – More sensitive than RAPD and RFLP.
No Prior Sequence Information Needed – Can be applied to any species.
Highly Reproducible – Overcomes RAPD's reproducibility issues.

Can Analyze Multiple Loci Simultaneously – Covers the entire genome efficiently.

Limitations of AFLP

Technically Complex & Time-Consuming – Requires multiple enzymatic and PCR steps.

Expensive – Needs specialized equipment like **fluorescent gel scanners** or **capillary electrophoresis**.

Dominant Marker – Cannot differentiate between homozygous and heterozygous loci.

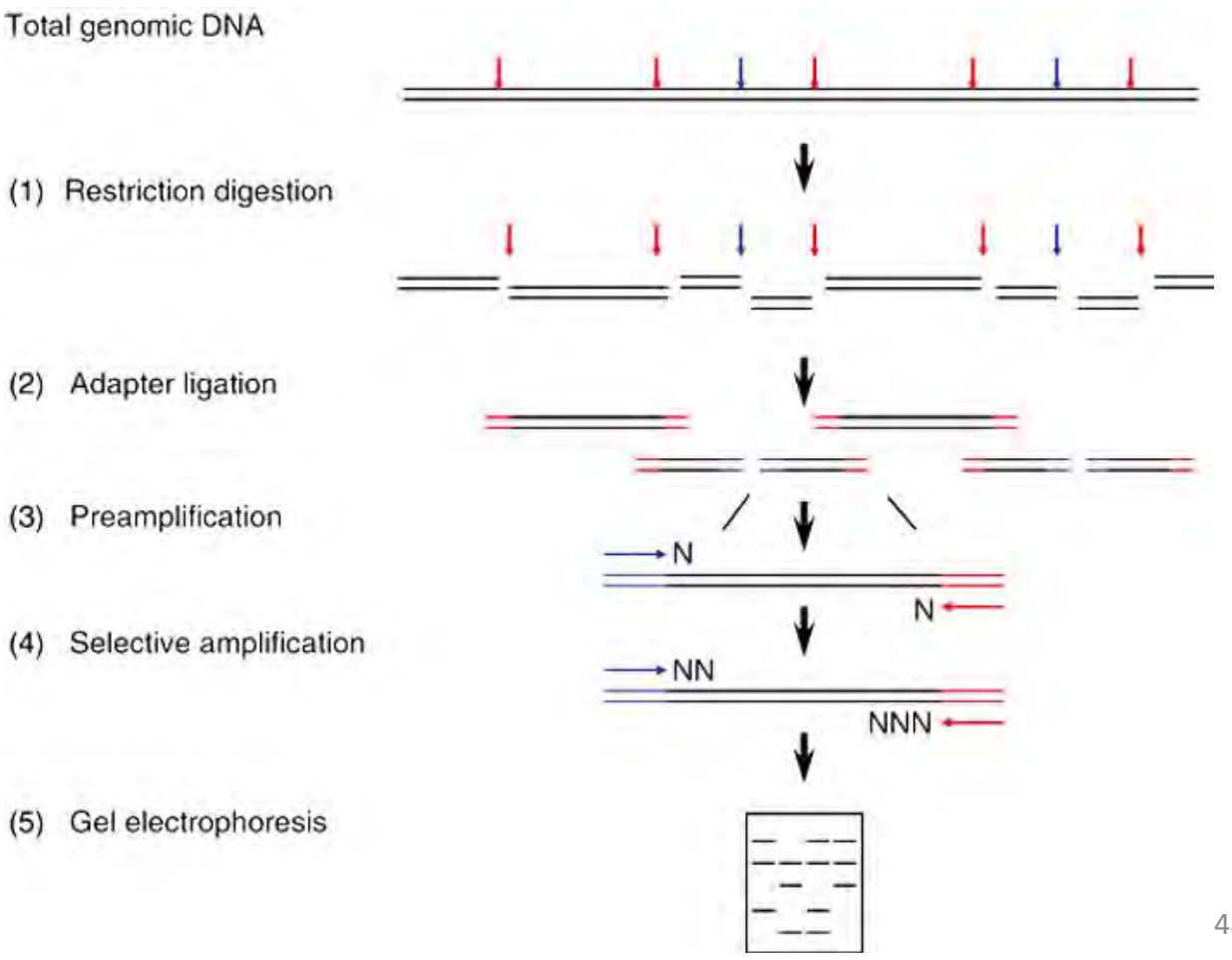
Requires High-Quality DNA – Degraded DNA may affect results.

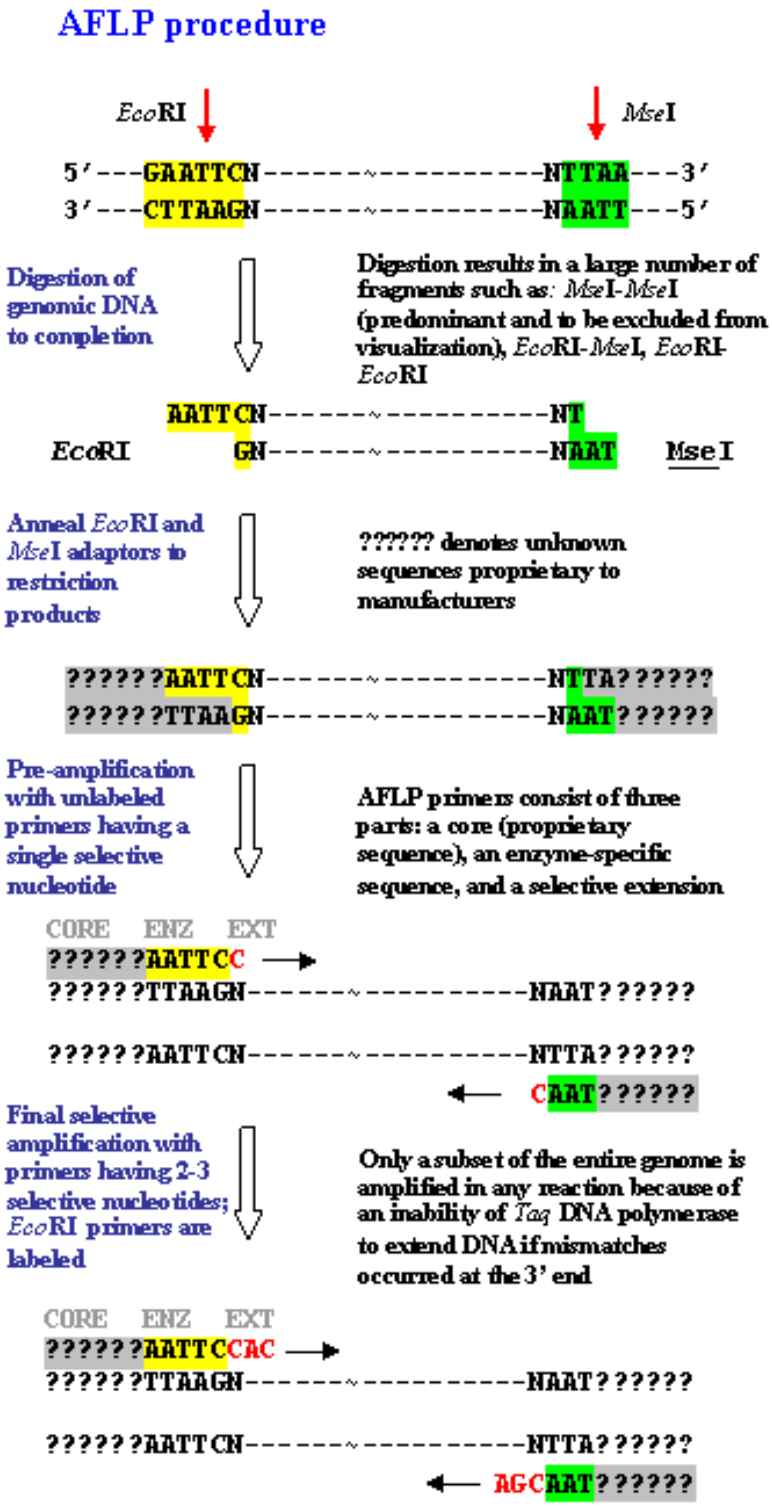
Conclusion

AFLP is a **powerful and highly polymorphic marker** for genetic analysis.

Despite its complexity, it is widely used in **taxonomy, breeding, and evolutionary studies**.

More advanced techniques like SNP markers and Next-Generation Sequencing (NGS) are replacing AFLP for genome-wide studies.





After final amplification, selectively amplified fragments are separated by gel electrophoresis and visualized autoradiographically. MseI-MseI fragments are excluded from the autorad because only EcoRI-directed primers are normally labeled. Typically, the autorad has 100-300 fingerprints with sizes ranging from 80 to 500 nucleotides. Only a subset (10-40) of these total bands is polymorphic between two related individuals, such as Arabidopsis thaliana Columbia and Landsberg erecta ecotypes.

Using 3-bp selective primer extensions gives 128 possible linker combinations. Therefore, 128 subsets of genomic DNA can be readily amplified. Thus, thousands of markers can be generated quite rapidly.

Simple Sequence Repeats (SSRs)

Introduction

Simple Sequence Repeats (SSRs), also known as **microsatellites**, are short tandemly repeated DNA sequences (e.g., **(CA)n, (GT)n, (AT)n**).

Highly polymorphic and widely used in **plant genetics, taxonomy, and breeding studies**.

Identified using **PCR amplification with specific primers flanking the repeat region**.

Principle of SSR Markers

SSRs consist of **1-6 nucleotide repeat motifs** that vary in length due to mutations.

PCR primers designed around these repeats amplify **different alleles** based on repeat number.

The amplified fragments are separated via **gel electrophoresis or capillary electrophoresis** to determine genetic variation.

Steps Involved in SSR Analysis

1. DNA Extraction

Isolate DNA from plant samples.

2. PCR Amplification

Use **specific primers** flanking the SSR region.

3. Gel Electrophoresis / Capillary Electrophoresis

Separate amplified DNA fragments based on size.

4. Band Visualization & Genotyping

Analyze banding patterns using **fluorescent dyes or silver staining**.

5. Data Analysis
Compare SSR alleles among different plant samples to assess genetic variation.

Applications of SSR Markers

Plant Taxonomy & Phylogenetics – Differentiates closely related species.
Genetic Diversity Studies – Measures variation within and between plant populations.
Crop Improvement & Marker-Assisted Breeding – Identifies genes for **disease resistance, yield, and stress tolerance**.
DNA Fingerprinting & Variety Identification – Protects plant breeders' rights.
Evolutionary & Conservation Studies – Assesses biodiversity and conservation priorities.

Advantages of SSR Markers

High Polymorphism – More variation than RAPD and AFLP.
Co-Dominant Inheritance – Distinguishes homozygous and heterozygous alleles.
Highly Reproducible & Reliable – Consistent results across different laboratories.
Requires Small DNA Quantities – Works with limited or degraded DNA.
Widely Distributed in Genomes – Found in both coding and non-coding regions.

Limitations of SSR Markers

Primer Design Complexity – Requires sequence information for flanking regions.

Time-Consuming & Costly – Developing new SSR markers can be labor-intensive.

Limited Genome Coverage – Compared to SNP markers and whole-genome sequencing.

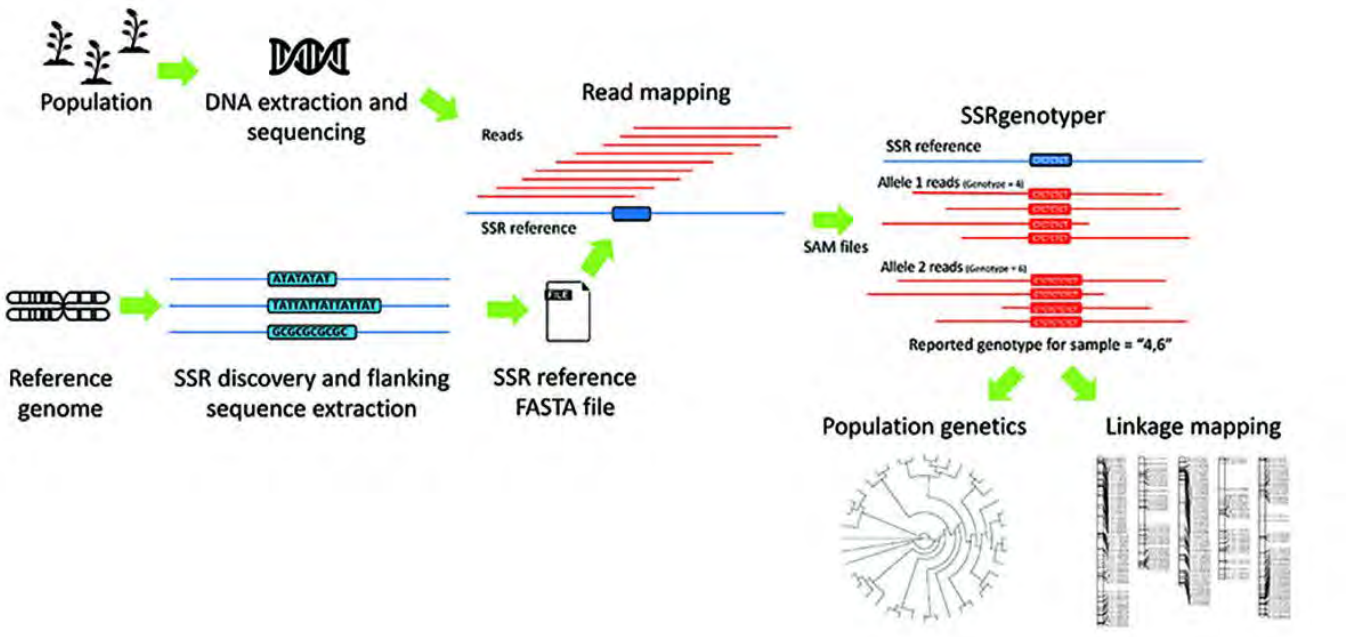
Potential PCR Artifacts – Stutter bands may complicate interpretation.

Conclusion

SSRs are **highly polymorphic, co-dominant, and reliable molecular markers**.

They are widely used in **genetic diversity studies, plant breeding, and DNA fingerprinting**.

Next-Generation Sequencing (NGS) and SNP markers are gradually replacing SSRs for high-throughput genotyping.



Single Nucleotide Polymorphisms (SNPs)

Introduction

Single Nucleotide Polymorphisms (SNPs) are single **base-pair variations** in the DNA sequence.

The most **abundant** and **stable** form of genetic variation in genomes.

Used in **plant breeding, evolutionary studies, and genetic mapping**.

Principle of SNP Markers

SNPs occur when a single nucleotide (A, T, C, or G) is altered at a specific location in the genome.

They can be **transition mutations (A↔G, C↔T)** or **transversion mutations (A↔C, G↔T, etc.)**.

Identified using high-throughput sequencing or genotyping techniques.

Types of SNPs

Coding Region SNPs

Can alter protein function (**nonsynonymous SNPs**) or have no effect (**synonymous SNPs**).

Non-Coding Region SNPs

Found in regulatory regions, affecting gene expression.

Intronic SNPs

Located in introns, may impact splicing.

Steps in SNP Genotyping

1. DNA Extraction

Isolate high-quality genomic DNA.

2. PCR Amplification

Amplify the DNA region containing the SNP.

3. SNP Detection Methods

DNA Sequencing (Sanger or NGS) – Directly reads SNPs.

TaqMan Assay – Uses fluorescent probes for allele discrimination.

Microarrays (SNP Chips) – Large-scale SNP genotyping.

KASP (Kompetitive Allele Specific PCR) – SNP-specific amplification.

4. Data Analysis

Compare SNP profiles to study genetic variation.

Applications of SNP Markers

Plant Breeding & Marker-Assisted Selection (MAS) – Identifies beneficial traits in crops.

Genetic Diversity & Phylogenetics – Determines evolutionary relationships.

Disease Resistance & Stress Tolerance – Finds SNPs linked to resistant genes.

Genome-Wide Association Studies (GWAS) – Identifies SNPs associated with traits.

Crop Improvement & Hybrid Development – Used in precision breeding programs.

Advantages of SNP Markers

Highly Abundant in Genomes – Found throughout the genome.

High Resolution & Accuracy – Detects fine genetic variations.

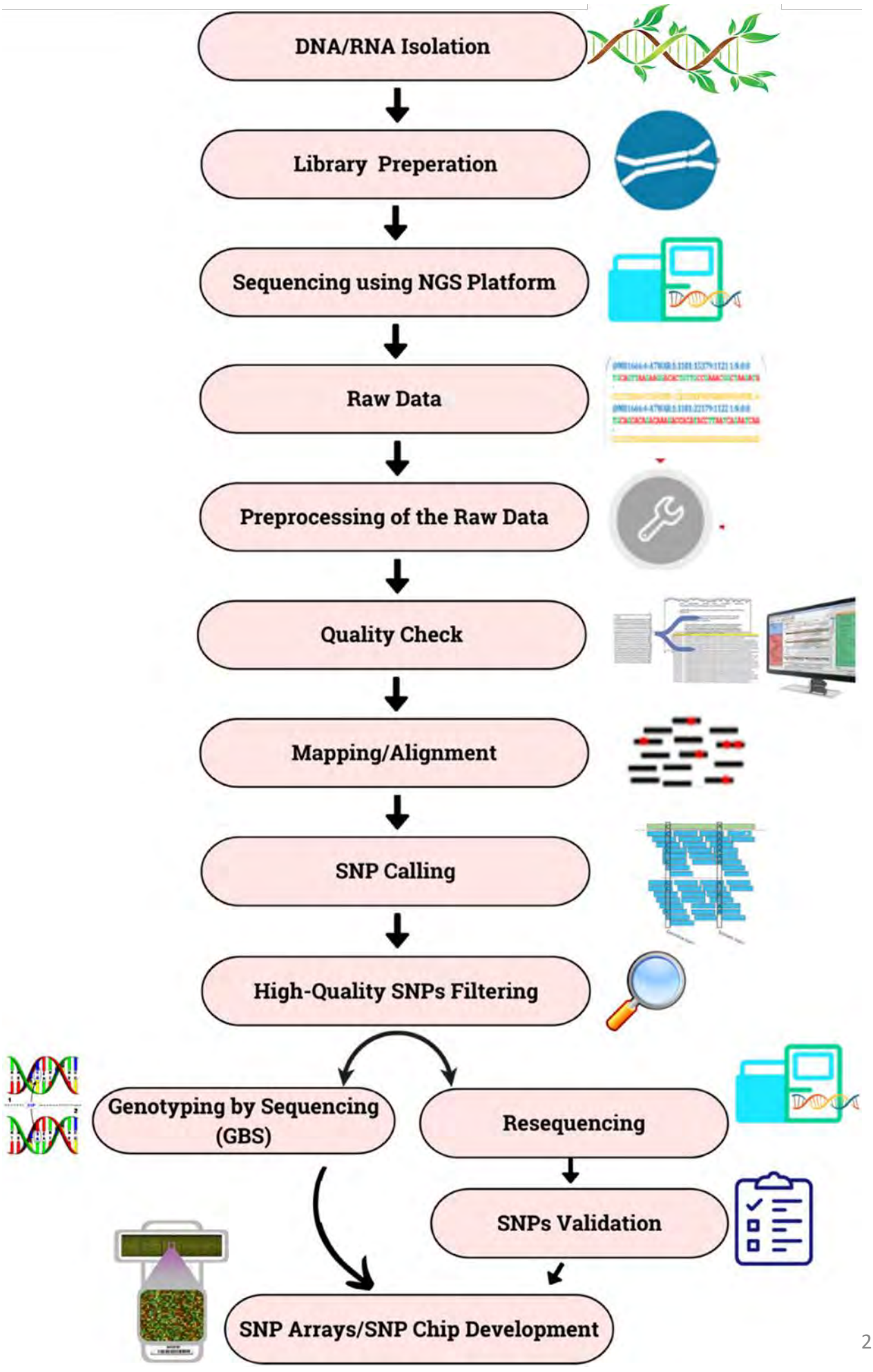
Codominant Markers – Differentiates between homozygous and heterozygous alleles.
Automation & High-Throughput Screening – Enables large-scale genotyping.
Stable & Reproducible – Lower mutation rate than SSRs.

Limitations of SNP Markers

High Initial Cost – Requires sequencing and SNP chip development.
Less Informative per Locus – SSRs have more alleles per marker.
Complex Data Analysis – Requires bioinformatics tools for large datasets.
Limited Genome Coverage (in Some Cases) – May require whole-genome sequencing.

Conclusion

SNPs are **powerful, high-resolution genetic markers** used in **breeding, taxonomy, and evolutionary studies**.
They provide **precise and stable genetic information**, making them essential for **modern genomics and crop improvement**.
Next-Generation Sequencing (NGS) and SNP chips enable large-scale SNP studies.


DNA/RNA Isolation
Library Preperation
Sequencing using NGS Platform
Raw Data
Preprocessing of the Raw Data
Quality Check
Mapping/Alignment
SNP Calling
High-Quality SNPs Filtering
Genotyping by Sequencing (GBS)
Resequencing
SNPs Validation
SNP Arrays/SNP Chip Development

Choosing molecular marker, and application of PCR in plant molecular taxonomy / DNA taxonomy

- In DNA sequencing method based practice of plant molecular taxonomy required DNA sequences.

- To obtain DNA sequence of a taxon required extraction of whole genomic DNA first. And then amplification of gene of interest. The amplification using gene interest is achieved by the polymerase chain reaction (PCR). The PCR results into billions of copies of gene of interest which can be observed in a gel under UV light. The amplified DNA later used for the purpose of DNA sequencing. So, for the cloning of the gene of interest using PCR requires primer. The primers are also called as molecular markers. To begin plant molecular taxonomy, selection of molecular marker is very critical and important.

- The most commonly used molecular marker in molecular taxonomy are ITS, rbcL, matK, psb, ndhF, trn gene.
- The molecular marker gene could be coding gene or non coding gene.

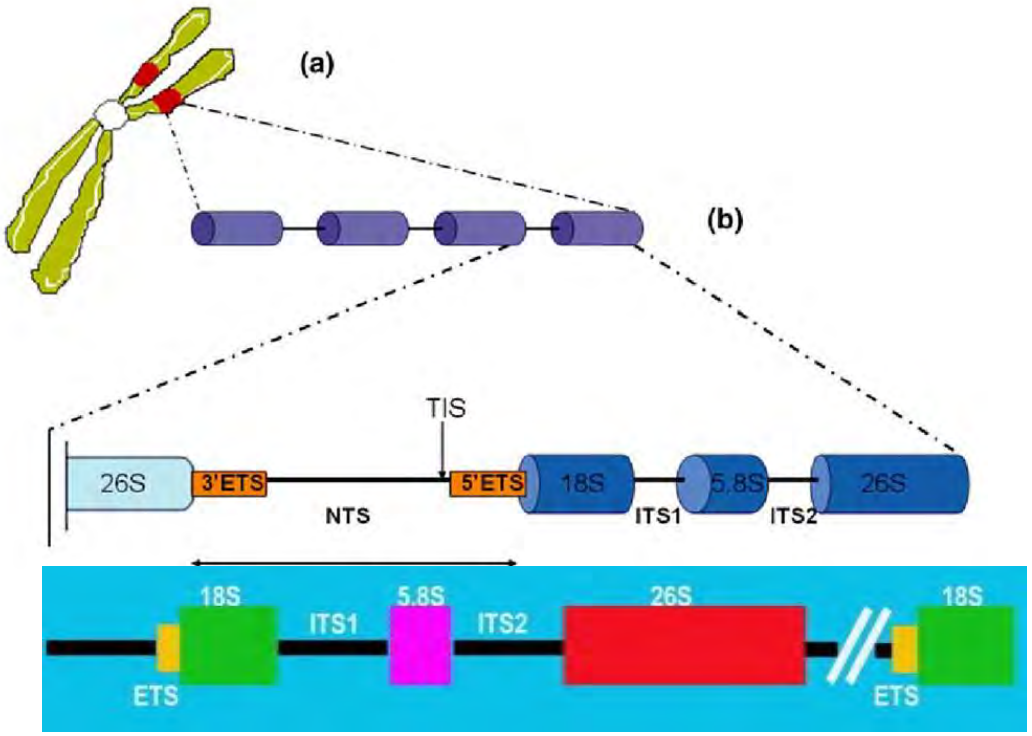
- Properties of ideal marker genes
  - A single-copy gene may be more useful than multiple-copy gene
  - The substitution rate should be optimum so as to provide enough informative sites and alignment should be easy.
  - Primers should be available to selectively amplify the marker gene

- The nuclear ribosomal locus coding for the large subunit is represented in tandem arrays in the plant genome.
- ITS is located between the 18 and 26S rRNA genes.
- The 5.8S region on the other hand is only about 160 bp long and highly conserved within major organism groups.
- The ITS region consists of three parts: the ITS1 and ITS2 and the highly conserved 5.8S rDNA exon located in between. The total length of this region varies between 500 and 750 bp in angiosperms while in other seed plants it can be much longer, up to 1,500–3,500 bp.
- Spacer DNA is a region of non-coding DNA between genes.
- In contrast to the coding regions, spacers evolve more quickly, like the internal transcribed spacer (ITS) region, which is extensively used as a marker for phylogenetic reconstruction at different levels.
- The ITS is present in virtually all organisms. The advantages of this region are: (1) easy PCR amplification, with several universal primers available for a various kind of organisms; (2) multicopy structure; (3) moderate size allowing easy sequencing; and (4) it has a high degree of variation even between closely related species.,
- variability is due to frequently occurring nucleotide polymorphisms or to common insertions/deletions in the sequence.

- As DNA of ITS regions is removed and it is not part of the mature RNA molecule, they are considered noncoding regions of the genome

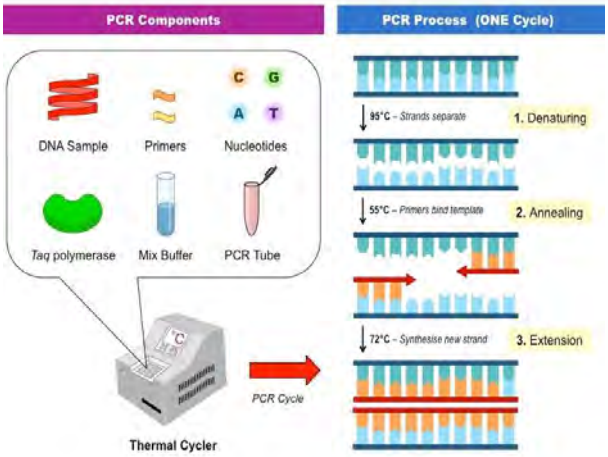
A fascinating feature of biological life is the common use of the DNA genetic code and its subsequent processing into functional units of protein through the intermediate RNA molecule.
The transcription of DNA into RNA and translation of RNA into protein are both highly
regulated and compartmentalized in all living organisms.

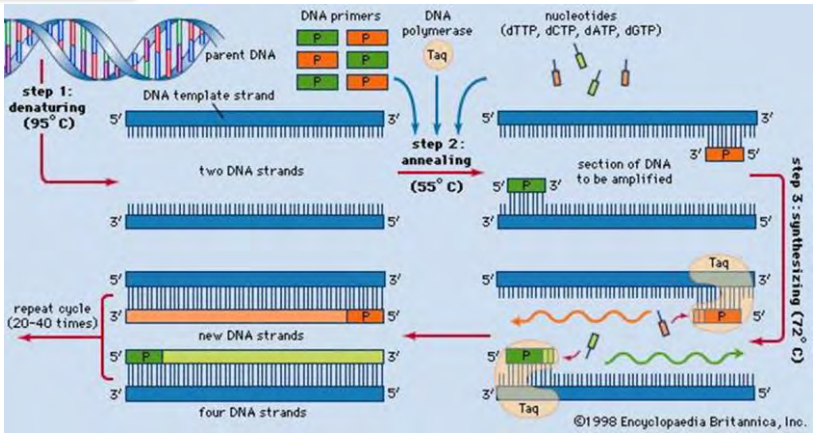
The cellular factory responsible for the production of protein is the ribosome. As the essential functions of ribosomes are critical for survival, their physical parameters have been conserved in all forms of life.
Some components within the ribosomal factories have, however, changed sometimes. These similarities, as well as the changes within genetic material can be used as tools for the identification of organisms

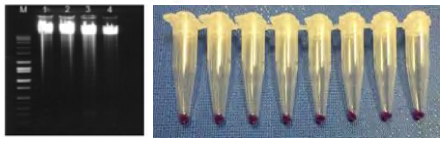


MARKER	SEQUENCE	REFERENCE
ITS1 F	TCCGTAGGTGAACCTGCGG	White et al. (1990)
ITS4 R	TCCTCCGCTTATTGATATGC	White et al. (1990)
rbcLa F	ATGTCACCACAAACAGAGACTAAAGC	Levin (2003)
rbcLa R	GTAAAATCAAGTCCACCRCG	Kress and Erickson (2007)
MatK 390 F	CGATCTATTCATTCAATATTTC	Cuenoud et al. (2002)
MatK 1326 R	TCTAGCACACGAAAGTCGAAGT	Cuenoud et al. (2002)
psbA-trnH F	GTTATGCATGAACGTAATGCTC	Sang et al. (1997)
psbA-trnH R	CGCGCATGGTGGATTCACAATCC	Tate and Simpson (2003)
trn L-F R	GGTTCAAGTCCCTCTATCCC	Taberlet et al. (1991)
trn L-F F	ATTTGAACTGGTGACACGAG	Taberlet et al. (1991)

PCR (Polymerase Chain Reaction)







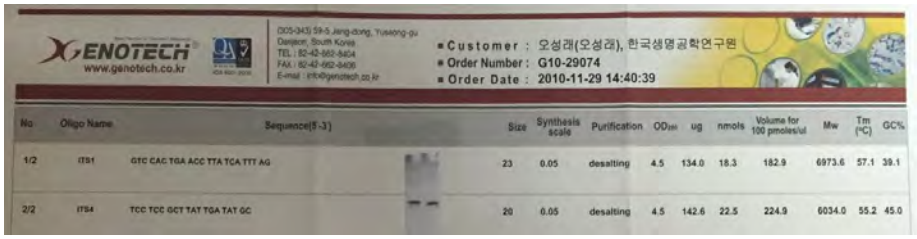
MARKER	SEQUENCE	REFERENCE
ITS1 F	TCCGTAGGTGAACCTGCGG	White et al. (1990)
ITS4 R	TCCTCCGCTTATTGATATGC	White et al. (1990)
rbcLa F	ATGTCACCACAAACAGAGACTAAAG C	Levin (2003)
rbcLa R	GTAAAATCAAGTCCACCRCG	Kress and Erickson (2007)
MatK 390 F	CGATCTATTCATTCAATATTTC	Cuenoud et al. (2002)
MatK 1326 R	TCTAGCACACGAAAGTCGAAGT	Cuenoud et al. (2002)
psbA-trnH F	GTTATGCATGAACGTAATGCTC	Sang et al. (1997)
psbA-trnH R	CGCGCATGGTGGATTCACAATCC	Tate and Simpson (2003)
trn L-F R	GGTTCAAGTCCCTCTATCCC	Taberlet et al. (1991)
trn L-F F	ATTTGAACTGGTGACACGAG	Taberlet et al. (1991)

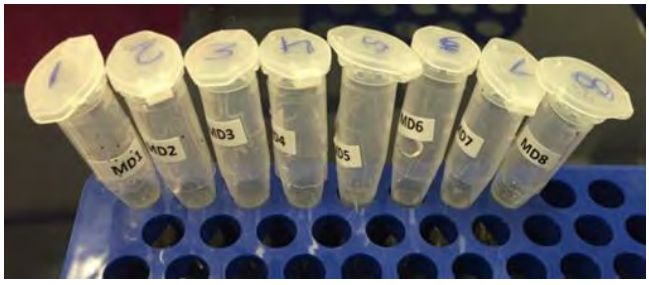
Contents of HF PCR premix Reaction size (20 µl reaction): 1. DNA polymerase 1µl, 2. Each dNTP (dATP, dCTP, dGTP, dTTP) 250 µM, 3. 10X reaction buffer Stabilizer and tracking dye2µl

Template DNA (1µl ~ 100 ng), Primer (1µl each of F and R, 5 ~20 pmole)

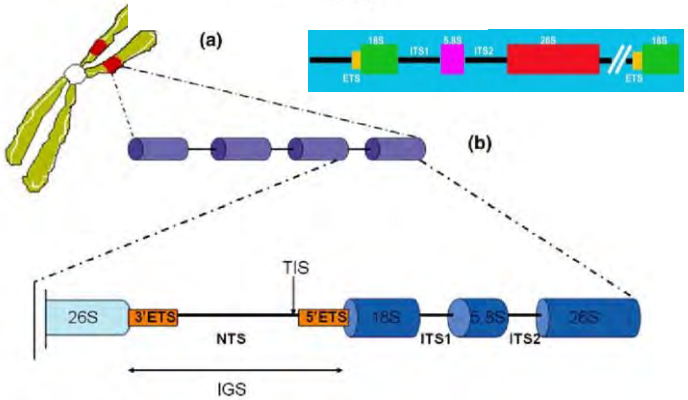
1/10th genomic DNA dilution: Add 10 µl total genomic DNA in 90 µl molecular grade distilled water.

Dilution of primer for stock solution (100 pmoles/ µl): nmols X 10 Distilled water (ddH_2O) = 100 pmoles/ µl (Stock)





PCR Parameters			
1	Initial Denaturation	94 °C for 5 minutes	
2	Denaturation	94 °C for 1minute	Number of cycles: 40
	Annealing	49 °C for 1minute	
	Extension	72 °C for 1minute	
3	Final extension	72 °C for 5 minutes	
4	Hold	4 °C	∞

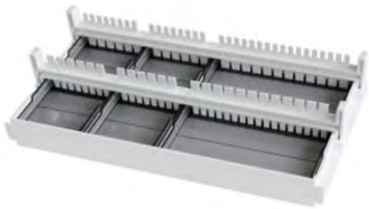


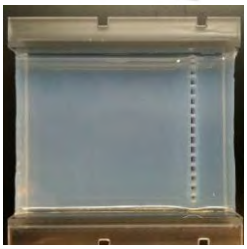
Agarose Gel Electrophoresis

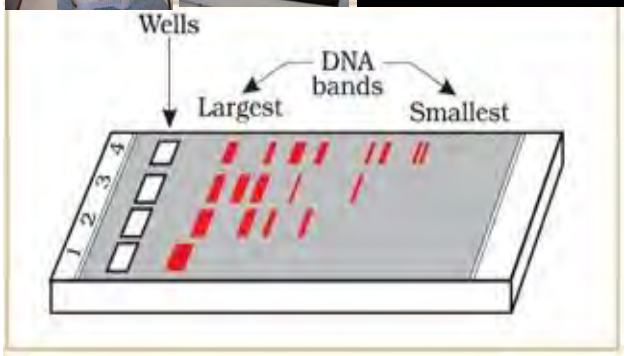
- The main purpose of agarose gel electrophoresis is to determine the presence or absence of genomic DNA or PCR products and quantify the size (length of the DNA molecule).
- Agarose gel electrophoresis is a widely used technique for the preparation and analysis of DNA. Electrophoresis is a method of separating DNA based on the rate of movement while under the influence of an electric field.
- Agarose is a polysaccharide purified from seaweed.
- An agarose gel is created by suspending dry agarose in a buffer solution, boiling until the solution becomes clear, and then pouring it into a casting tray and allowing it to cool. During electrophoresis, the gel is submersed in a chamber containing a buffer solution and a positive and negative electrode.
- The DNA to be analyzed is forced through the pores of the gel by the electrical current.
- Under an electrical field, DNA moves to the positive electrode (red) and away from the negative electrode (black).
- DNA itself is not visible within an agarose gel.
- The DNA visualized by the use of dye that binds to DNA.

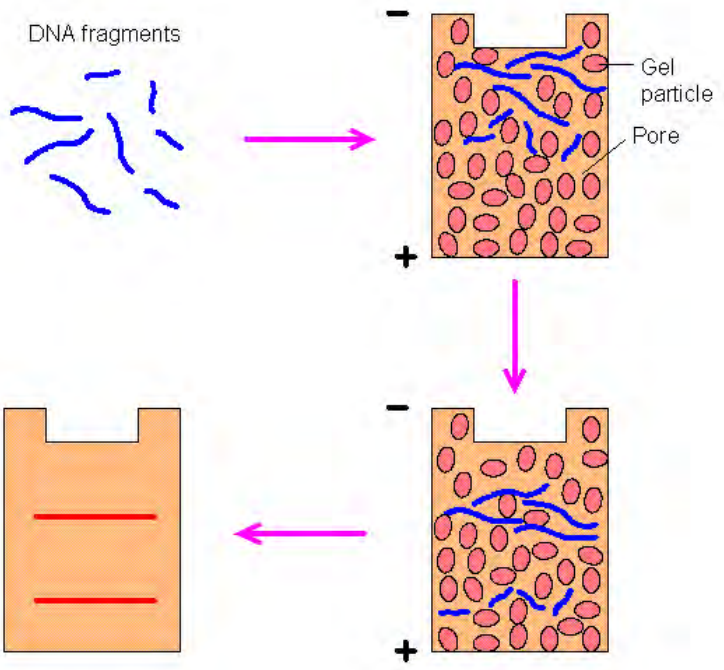
Agarose

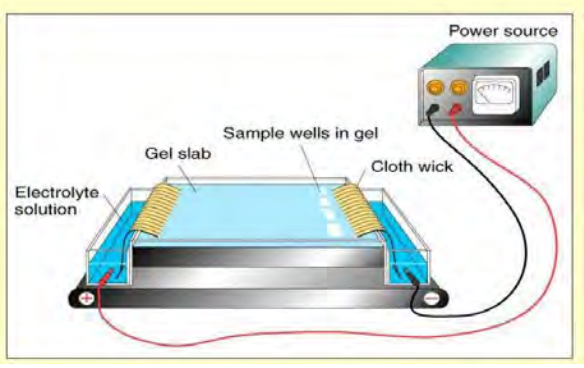
Buffer Solution







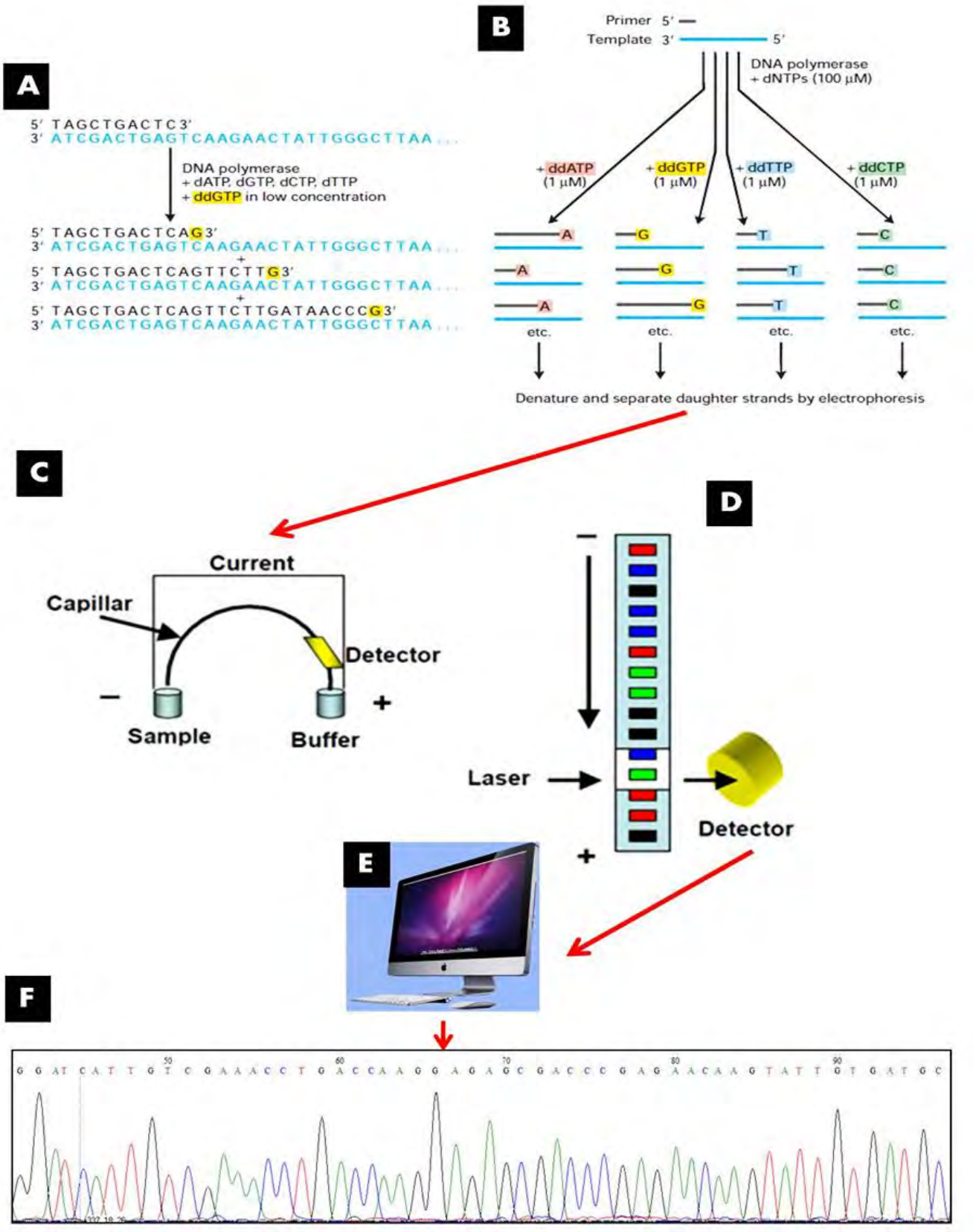




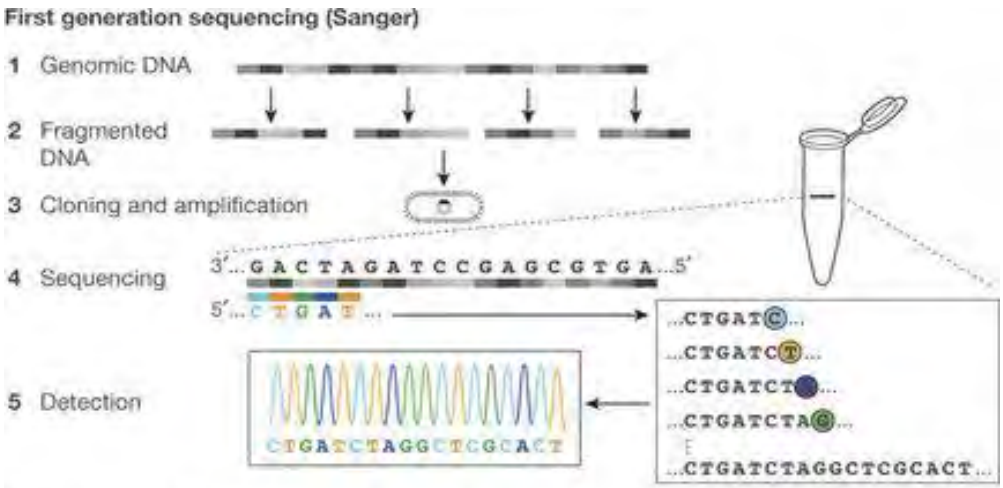
DNA sequencing

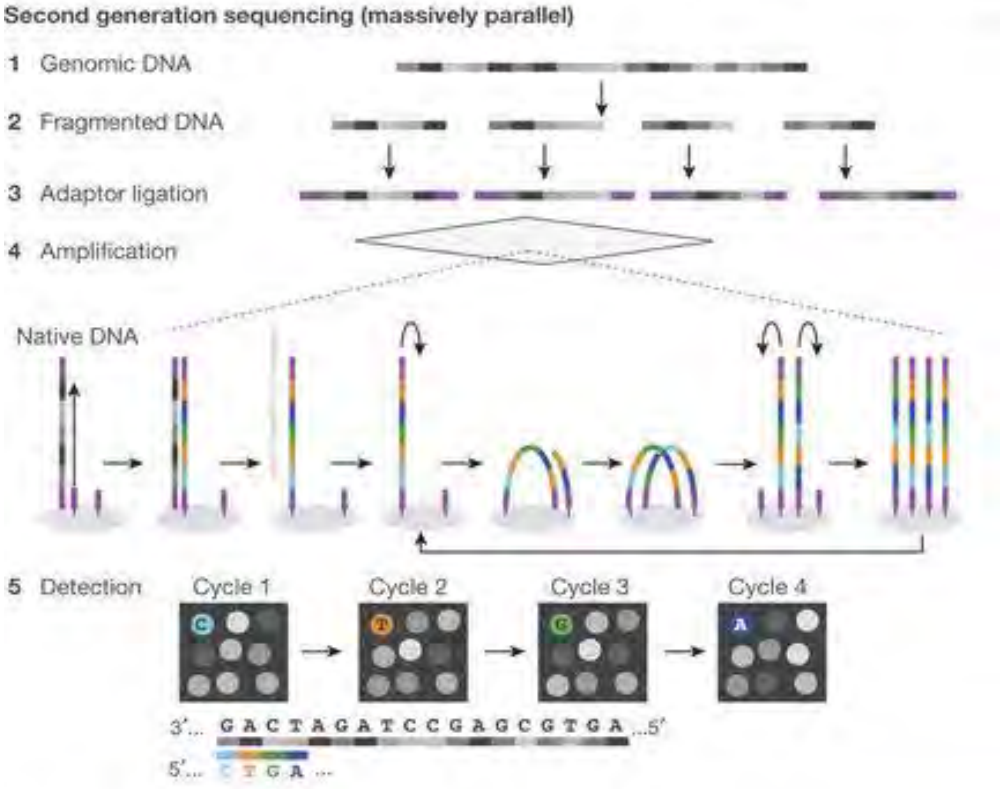
- DNA sequencing is the process of determining the sequence of nucleotides (A, T, C, and G) in a piece of DNA.
- In Sanger sequencing, the target DNA is copied many times, making fragments of different lengths. Fluorescent “chain terminator” nucleotides mark the ends of the fragments and allow the sequence to be determined.
- Next-generation sequencing techniques are new, large-scale approaches that increase the speed and reduce the cost of DNA sequencing.
- Sanger sequencing: The chain termination method
- Regions of DNA up to about 900 base pairs in length are routinely sequenced using a method called Sanger sequencing or the chain termination method.
- Ingredients for Sanger sequencing
- Sanger sequencing involves making many copies of a target DNA region. Its ingredients are similar to those needed for DNA replication in an organism, or for polymerase chain reaction (PCR), which copies DNA *in vitro*. They include:
- A DNA polymerase enzyme
- A primer, which is a short piece of single-stranded DNA that binds to the template DNA and acts as a "starter" for the polymerase
- The four DNA nucleotides (dATP, dTTP, dCTP, dGTP)
- The template DNA to be sequenced
- However, a Sanger sequencing reaction also contains a unique ingredient:
- Dideoxy, or chain-terminating, versions of all four nucleotides (ddATP, ddTTP, ddCTP, ddGTP), each labeled with a different color of dye
- Dideoxy nucleotides are similar to regular, or deoxy, nucleotides, but with one key difference: they lack a hydroxyl group on the 3’ carbon of the sugar ring. In a regular nucleotide, the 3’ hydroxyl group acts as a “hook," allowing a new nucleotide to be added to an existing chain.
- Once a dideoxy nucleotide has been added to the chain, there is no hydroxyl available and no further nucleotides can be added. The chain ends with the dideoxy nucleotide, which is marked with a particular color of dye depending on the base (A, T, C or G) that it carries.
- The DNA sample to be sequenced is combined in a tube with primer, DNA polymerase, and DNA nucleotides (dATP, dTTP, dGTP, and dCTP). The four dye-labeled, chain-terminating dideoxy nucleotides are added as well, but in much smaller amounts than the ordinary nucleotides.
- The mixture is first heated to denature the template DNA (separate the strands), then cooled so that the primer can bind to the single-stranded template. Once the primer has bound, the temperature is raised again, allowing DNA polymerase to synthesize new DNA starting from the primer. DNA polymerase will continue adding nucleotides to the chain until it happens to add a dideoxy nucleotide instead of a normal one. At that point, no further nucleotides can be added, so the strand will end with the dideoxy nucleotide.
- This process is repeated in a number of cycles. By the time the cycling is complete, it’s virtually guaranteed that a dideoxy nucleotide will have been incorporated at every single position of the target DNA in at least one reaction. That is, the tube will contain fragments of different lengths, ending at each of the nucleotide positions in the original DNA (see figure below). The ends of the fragments will be labeled with dyes that indicate their final nucleotide.
- After the reaction is done, the fragments are run through a long, thin tube containing a gel matrix in a process called capillary gel electrophoresis. Short fragments move quickly through the pores of the gel, while long fragments move more slowly. As each fragment crosses the “finish line” at the end of the tube, it’s illuminated by a laser, allowing the attached dye to be detected.

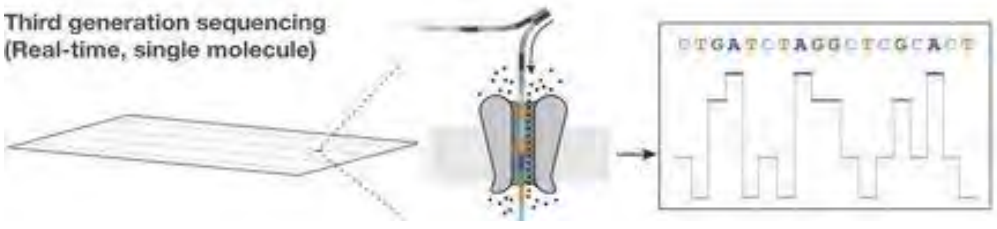
- **The smallest fragment (ending just one nucleotide after the primer) crosses the finish line first, followed by the next-smallest fragment (ending two nucleotides after the primer), and so forth. Thus, from the colors of dyes registered one after another on the detector, the sequence of the original piece of DNA can be built up one nucleotide at a time. The data recorded by the detector consist of a series of peaks in fluorescence intensity, as shown in the chromatogramabove. The DNA sequence is read from the peaks in the chromatogram.**



Advances in next generation DNA sequencing







BOX 1

The milestones listed below correspond to key developments in the evolution of sequencing technologies. This is a large topic, and we apologize for any omissions.

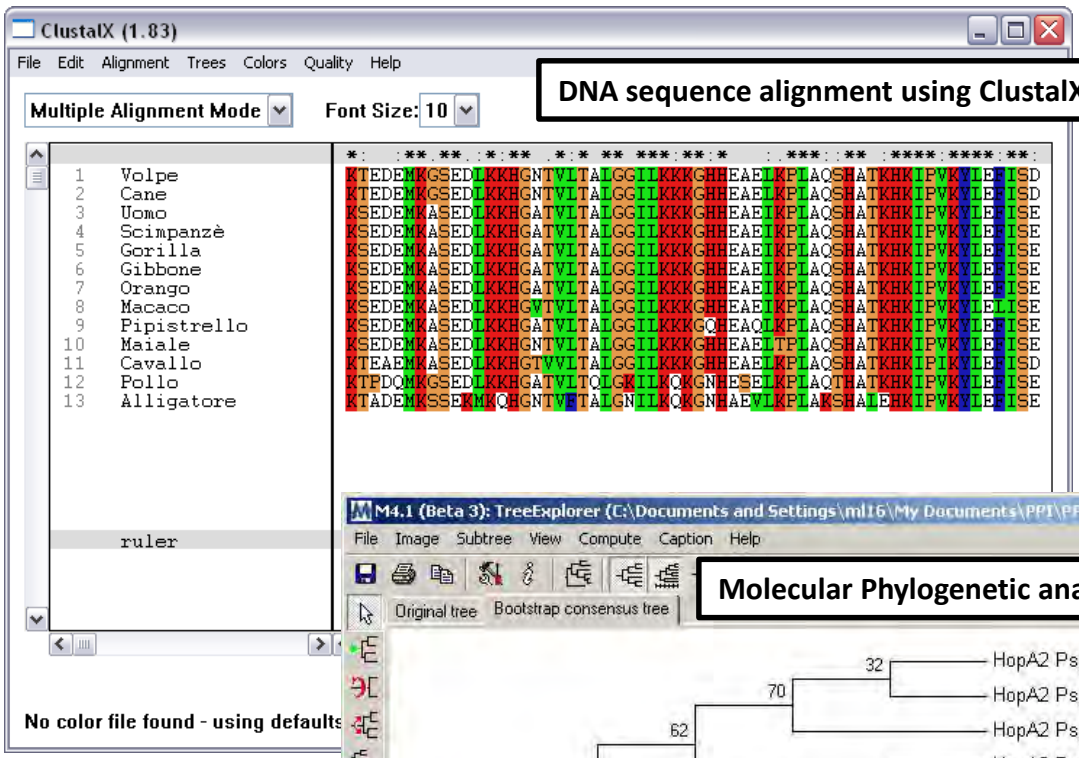
Technical milestones

1953: Sequencing of insulin protein[2]
1965: Sequencing of alanine tRNA[4]
1968: Sequencing of cohesive ends of phage lambda DNA[6]
1977: Maxam–Gilbert sequencing[9]
1977: Sanger sequencing[8]
1981: Messing's M13 phage vector[12]
1986–1987: Fluorescent detection in electrophoretic sequencing[14,15,17]
1987: Sequenase[18]
1988: Early example of sequencing by stepwise dNTP incorporation[139]
1990: Paired-end sequencing[23]
1992: Bodipy dyes[140]
1993: *In vitro* RNA colonies[37]
1996: Pyrosequencing[44]
1999: *In vitro* DNA colonies in gels[38]
2000: Massively parallel signature sequencing by ligation[47]
2003: Emulsion PCR to generate *in vitro* DNA colonies on beads[42]
2003: Single-molecule massively parallel sequencing-by-synthesis[33,34]
2003: Zero-mode waveguides for single-molecule analysis[57]
2003: Sequencing by synthesis of *in vitro* DNA colonies in gels[49]
2005: Four-colour reversible terminators[51–53]
2005: Sequencing by ligation of *in vitro* DNA colonies on beads[41]
2007: Large-scale targeted sequence capture[93–96]
2010: Direct detection of DNA methylation during single-molecule sequencing[65]
2010: Single-base resolution electron tunnelling through a solid-state detector[141]
2011: Semiconductor sequencing by proton detection[142]
2012: Reduction to practice of nanopore sequencing[143,144]
2012: Single-stranded library preparation method for ancient DNA[145]

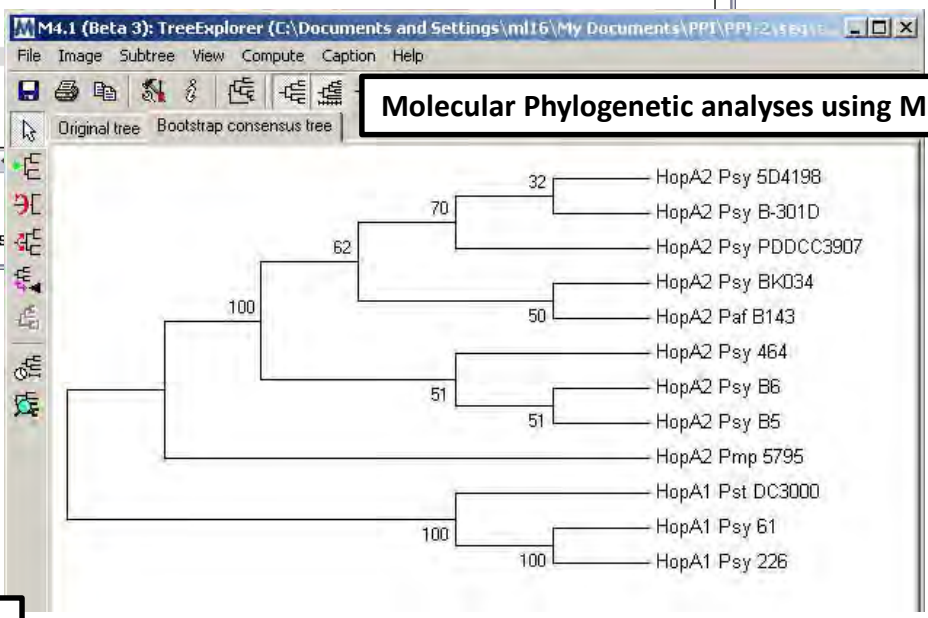
DNA sequence dataset preparation

GTCGAAACCTGCATAGCAGAACGACCCGCGAACACGTTACACTACCAGGTGAGGGACGAGGGGTGCGCAA
GCTCCCCAAGTTTCAAACCCATGGTCGGGGACCACCCTTGGGTGGCCTCGTCCGAACAACGACCCCCCGG
CGCGGAATGCGCCAAGGAAATCAAACTGAACTGCACGCGTCCCCCCCGTTTGCGGGCGGCGGAAGCGTCT
TTCTAAAACACAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCA
TTAGGCCGAGGGCACGTCTGCCTGGGCGTCACACATCGCGTCGCCCCCCAACCCATCACTCCCTTGCGGG
AGTTGAGGCGGAGGGGCGGATAATGGCCTCCCGTGTCTCACCGCGCGGTTGGCCCAAATGCGAGTCCTTG
GCGATGGACGTCACGACAAGTGGTGGTTGTAAAAAGCCCTCTTCTCATGTCGTGCGGTGACCCGTCGCCA
GCAAAATCTCTCATGACCCTGTTGCGCCGAGCCTCGACGCGCGCTCCGACCGCGACCCC

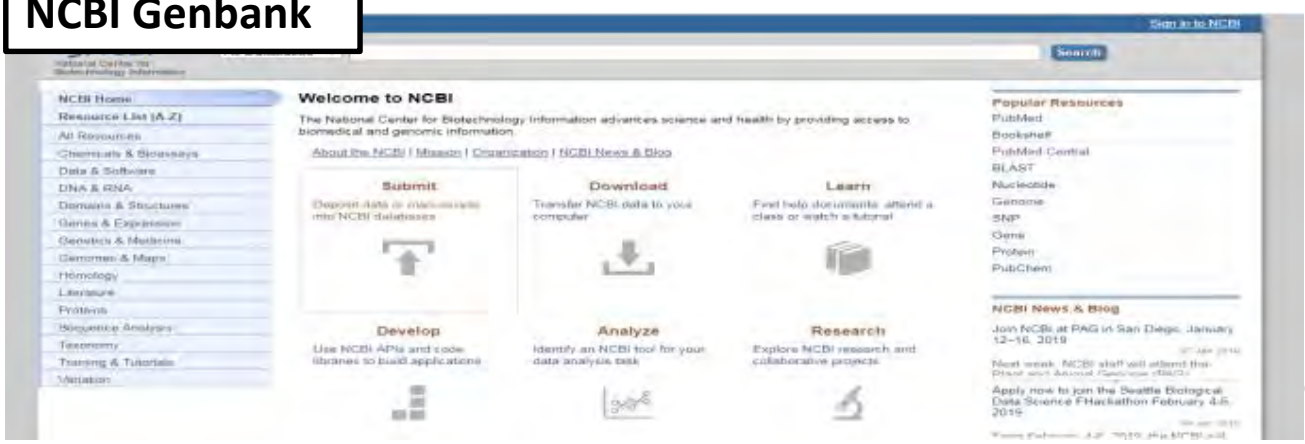
DNA sequence alignment using ClustalX



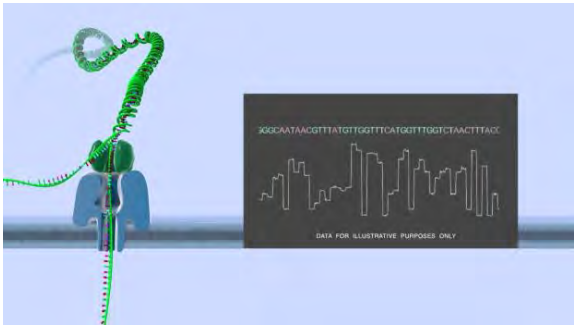
Molecular Phylogenetic analyses using MEGA

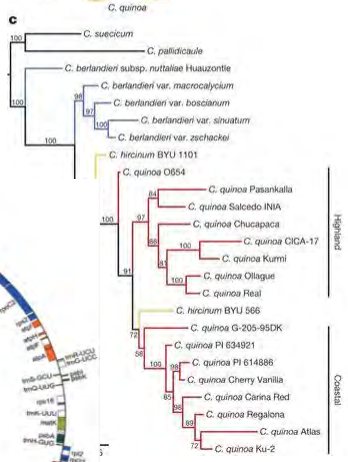


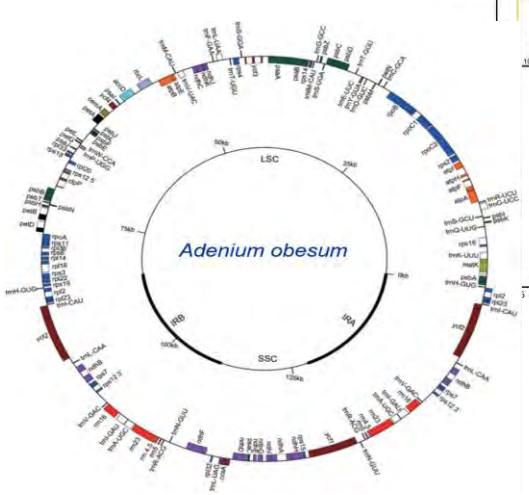
NCBI Genbank



Advances in next generation DNA sequencing and its application in systematics







❖**Whole Chloroplast Genome Sequencing of *Adenium obesum***

❖**Chloroplast (cp) is a special subcellular organelle which contains the entire enzymatic machinery for photosynthesis.**

❖**Chloroplast contains its own small genome of 120–217 kb in size and 110-130 genes, consists of a circular double-stranded DNA.**

❖**The cp genome can be used to investigate molecular evolution and phylogenies.**

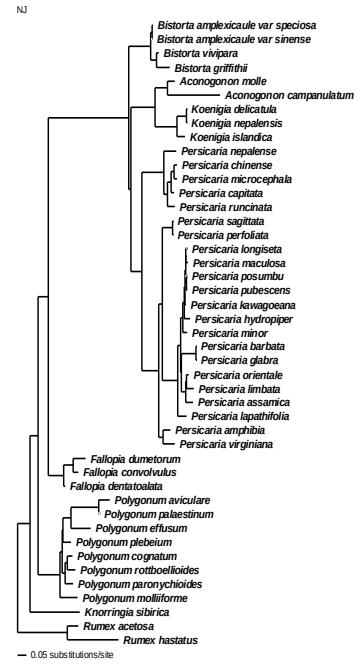
❖**The cp genomes are maternally inherited, which is beneficial in genetic engineering.**

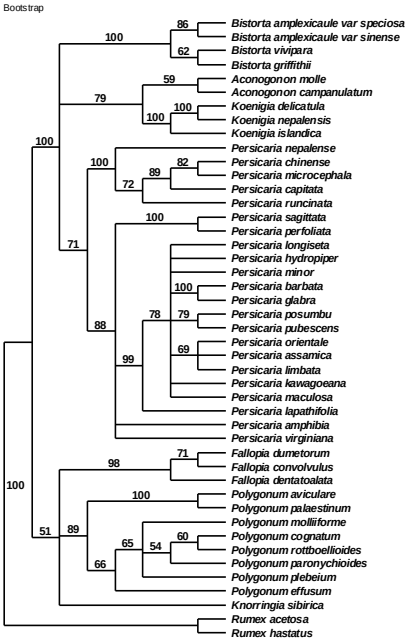
Whole genome sequencing

❖***Chenopodium quinoa* (quinoa) is a highly nutritious grain identified as an important crop** Jarvis 2017 Nature. ; 542(7641):307-312. **The genome of Chenopodium quinoa.**

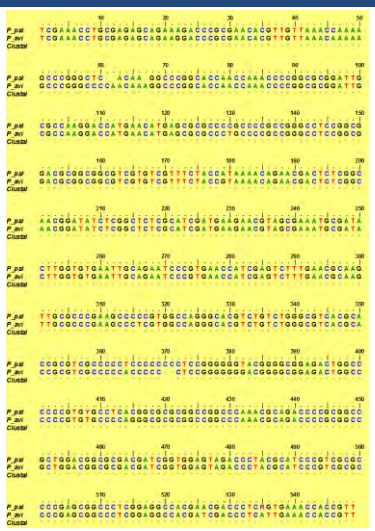
Molecular systematic studies on *Polygonum palaestinum* Zohary (polygonaceae) from Saudi Arabia using ITS sequences of nuclear ribosomal DNA

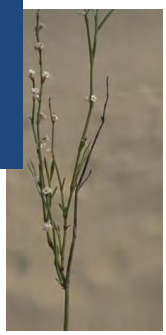
- The taxonomy of the genus Polygonum is highly controversial because of diverse variation within species among the species has resulted into lack of consensus on taxonomic circumscription. Therefore, there is disagreement among the taxonomists that to which species should be retain within the genus Polygonum and to which species should be elevated to their own genus.
- The genus Polygonum in Saudi Arabia includes P. argyrocoleum Steud. ex Kunze, P. aviculare L. and P. palaestinum Zohary. Two out of these Polygonums of Saudi Arabia i.e. P. argyrocoleum and P. aviculare are common weed distributed throughout. The distribution of P. palaestinum is restricted to Harratal Harra area of Saudi Arabia.
- Decraene and Akeroyd (1988) have segregated Polygonum in the broad sense into two separate tribes, Polygoneae and Persicarieae.
- The systematic status of P. palaestinum is unresolved





Polygonum palaestinum Zohary (polygonaceae)

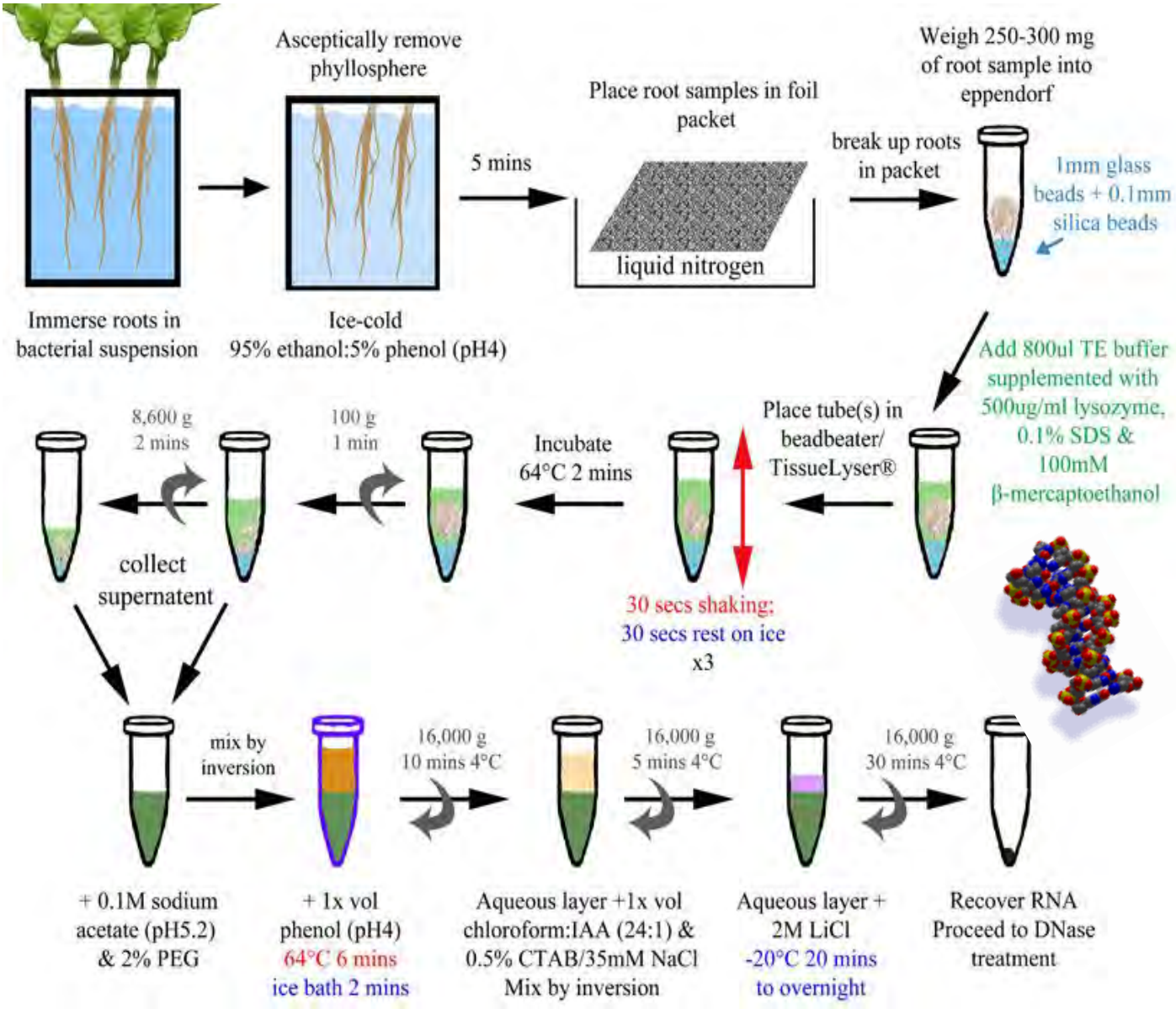




Position in sequence alignment	*Polygonum aviculare*	*Polygonum palaestinum*
22	G	A
46	A	C
59	C	T
61	C	-
62	A	-
67	A	-
130	T	C
179	G	A
316	T	C
364	A	T
370	-	C
371	-	C
382	G	T
408	T	Y C T
413	C	T
415	G	C
526	T	A
535	A	R A G
536	T	G

Sampling of leaf material for the molecular taxonomic study and DNA extraction

- **Doyle and Doyle (1990) is widely used protocol for DNA Extraction from plant tissue. But it involves preparation of several buffer manually. It takes long times. This method atleast take more than one day preparation and about whole day in DNA extraction. It also involves several times centrifugation. This method requires large amount of fresh leaves (10 gram or even more).**
- **Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15**
- **In contrast to manual method, there are several DNA extraction kit and automated DNA extraction machine is available like Qiagen automated DNA extraction machine, and Qiagen DNA extraction Kit.**
- **In Qiagen DNA extraction all the buffer are provided and ready to use. DNA can be extracted from small amount of 20 mg l dried leaf tissue or from very small piece of leaf collected from even old herbarium specimens. By using Qiagen DNA can be extracted in 3 hours. It do not required centrifugation manually.**



Chapter 7- Genetic diversity and DNA barcoding

Advance Experimental Taxonomy

Chapter 7- Genetic diversity and DNA barcoding

Introduction to Genetic Diversity

Definition:

Genetic diversity refers to the total number of genetic characteristics present within a species, population, or community. It represents the variety of genes and their alleles, which contribute to the uniqueness of individuals and populations.

Why is Genetic Diversity Important?

Genetic diversity plays a crucial role in the survival and adaptability of species. The greater the genetic variation within a population, the higher its chances of withstanding environmental changes, diseases, and other challenges.

Key Benefits of Genetic Diversity:

Increases Adaptability to Environmental Changes:

Species with high genetic diversity have a better chance of adapting to climate fluctuations, habitat destruction, and other ecological pressures.

Example: Some plant populations develop drought resistance due to genetic variation.

Reduces Susceptibility to Diseases:

A genetically diverse population is less likely to be wiped out by a single disease, as different individuals may have genetic resistance.

Example: Human populations with diverse genetic backgrounds show varied resistance to diseases like malaria and COVID-19.

Essential for Evolution and Survival:

Evolution occurs through natural selection, which acts on genetic variation. Without diversity, species cannot evolve to cope with new challenges.

Example: The development of antibiotic-resistant bacteria is a result of genetic mutations and selection.

Examples of Genetic Diversity in Nature:

Wild Species (High Genetic Diversity):

Cheetahs: Despite their low diversity, slight genetic differences have helped some individuals survive habitat changes.

Mangrove Trees: These trees have evolved various salt tolerance mechanisms due to high genetic variation.

Domesticated Species (Low Genetic Diversity):

Bananas (Cavendish variety): Clonal propagation has led to genetic uniformity, making them highly vulnerable to diseases like Panama disease.

Purebred Dogs: Selective breeding has reduced genetic variation, increasing the risk of inherited diseases.

Threats to Genetic Diversity:

Habitat destruction and deforestation

Climate change

Overexploitation of species

Genetic bottlenecks (e.g., cheetahs suffering from inbreeding)

Conservation Efforts to Maintain Genetic Diversity:

Establishing seed banks and gene banks

Protecting natural habitats

Promoting breeding programs that maintain genetic variability

Using genetic engineering and biotechnology to enhance diversity

Factors Affecting Genetic Diversity

Genetic diversity is influenced by several evolutionary processes that shape the genetic composition of populations over time. The main factors affecting genetic diversity include **natural selection, mutation, genetic drift, gene flow, and population bottlenecks/founder effects**.

1. Natural Selection

Natural selection is the process by which individuals with favorable traits survive and reproduce more successfully than others, passing on advantageous genes to the next generation.

How It Affects Genetic Diversity:

Increases diversity when beneficial mutations spread through a population.

Reduces diversity when harmful mutations are removed or when a single trait becomes dominant (directional selection).

Example:

Peppered Moth (Biston betularia):

Before the Industrial Revolution, lighter-colored moths were more common because they blended with tree bark.

During pollution, dark-colored moths survived better due to soot-covered trees, leading to a shift in allele frequency.

2. Mutation

A mutation is a random change in an organism's DNA sequence. It is the primary source of new genetic variation.

How It Affects Genetic Diversity:

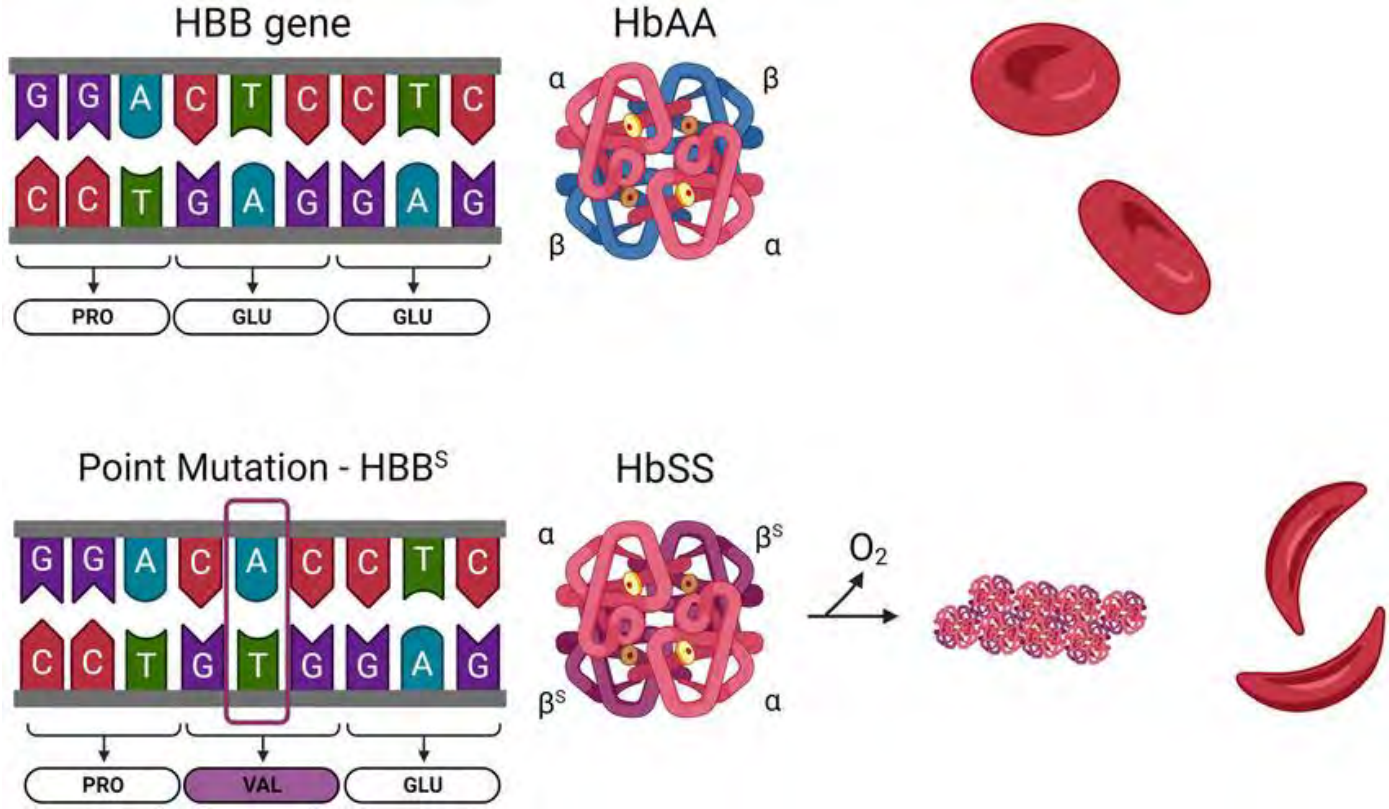
Increases genetic diversity by introducing new alleles into a population.

Can be beneficial, neutral, or harmful, depending on the effect on an organism’s survival and reproduction.

Example:

Sickle Cell Anemia (HbS mutation in humans):

The mutation causing sickle cell anemia also provides resistance to malaria in heterozygous individuals, maintaining its presence in some populations.



3. Genetic Drift

Genetic drift is the **random** fluctuation of allele frequencies in a population due to chance events. It has a stronger effect in small populations.

How It Affects Genetic Diversity:

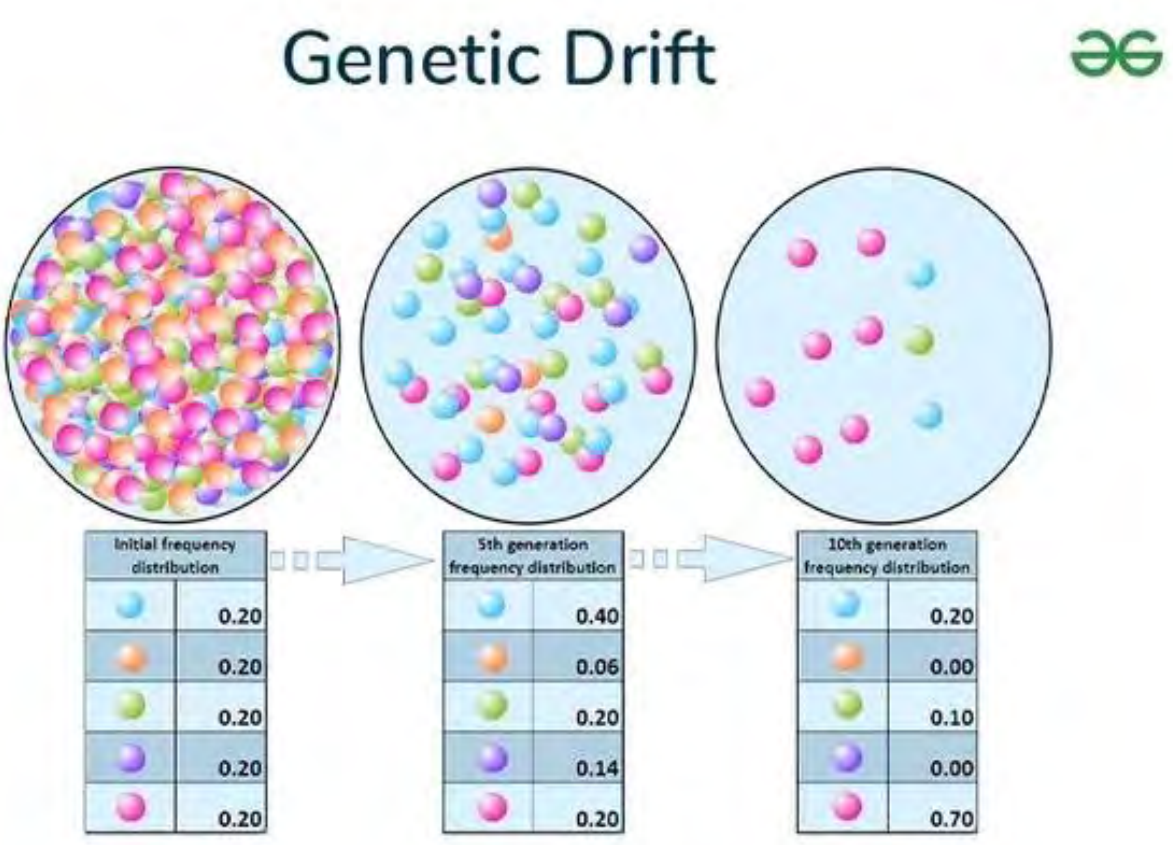
Reduces genetic diversity by causing the loss of alleles over generations.

Can lead to fixation of certain alleles (when an allele reaches 100% frequency).

Example:

Island Populations:

A small group of individuals colonizing an island may experience random shifts in allele frequency, reducing overall diversity.



4. Gene Flow (Migration)

Gene flow refers to the movement of genes between populations due to the migration of individuals or the transfer of gametes (e.g., pollen in plants).

How It Affects Genetic Diversity:

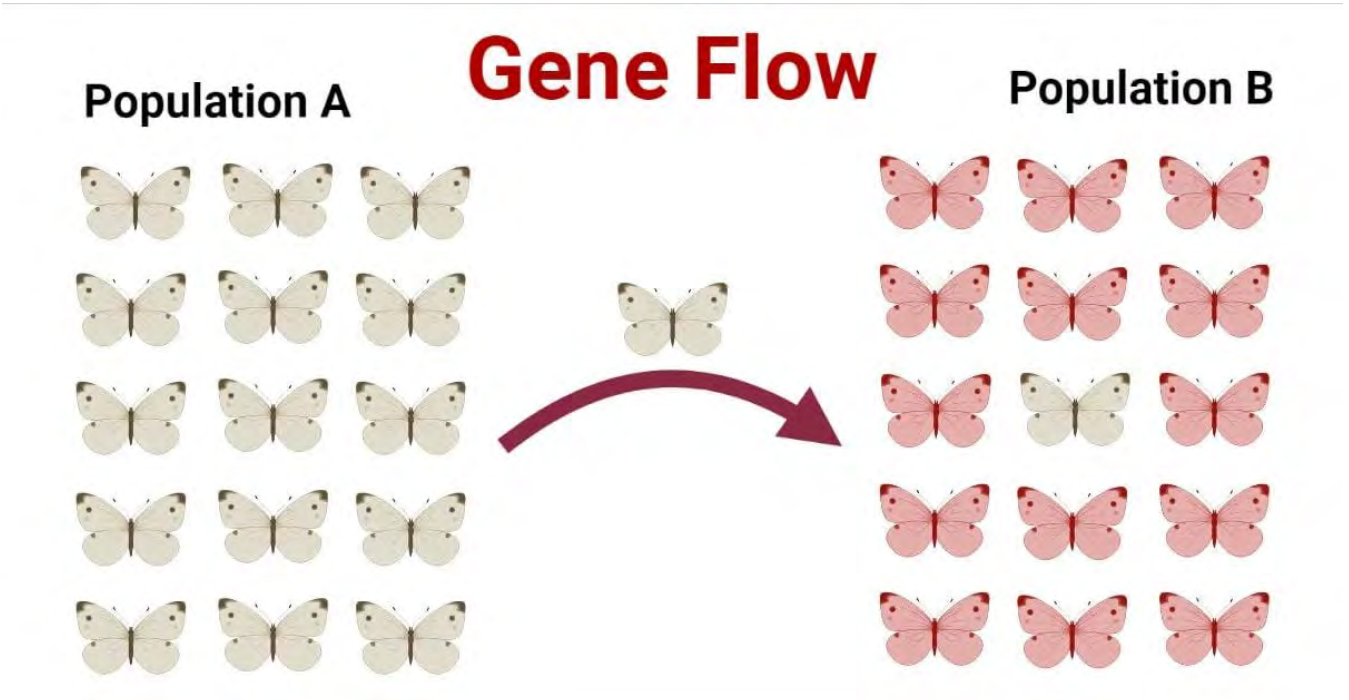
Increases genetic diversity by introducing new alleles into a population.

Can homogenize populations if high gene flow reduces genetic differences between them.

Example:

Human Migration:

Historically, interbreeding between different human populations (e.g., Neanderthals and modern humans) introduced beneficial genes, such as immunity-related genes.



5. Bottleneck & Founder Effects

These effects occur when a population undergoes a drastic reduction in size or is established by a small group of individuals, leading to lower genetic diversity.

A. Population Bottleneck

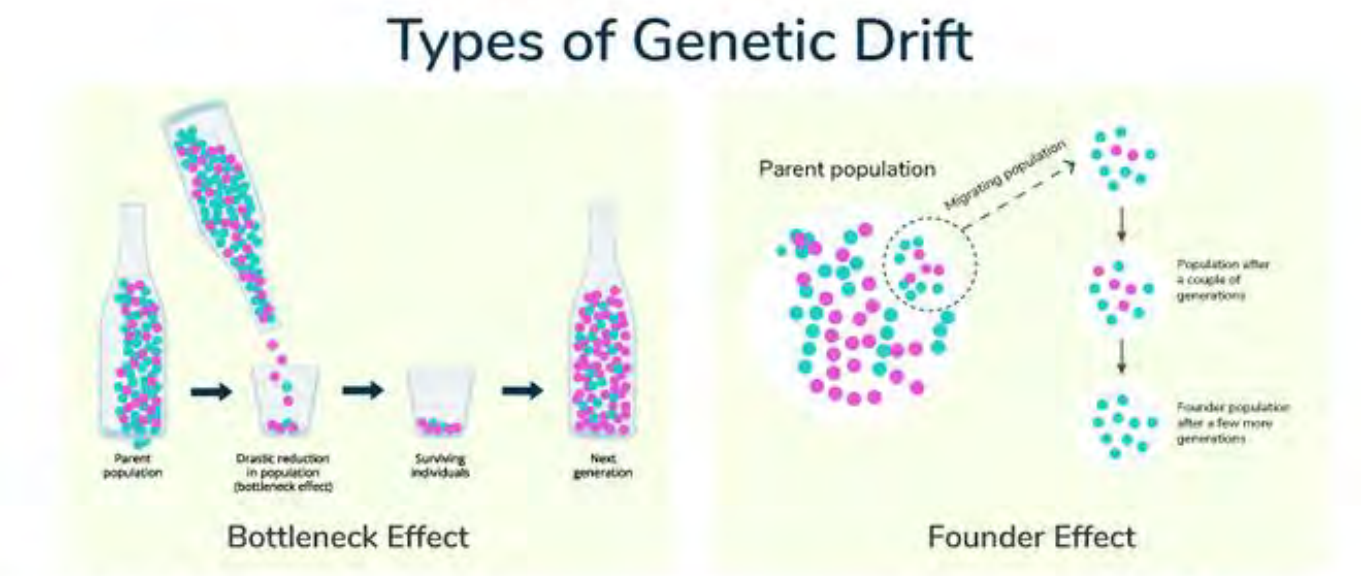
A bottleneck happens when a large portion of a population is suddenly wiped out due to a natural disaster, disease, or human activities.

How It Affects Genetic Diversity:

- **Greatly reduces genetic variation**, making populations more vulnerable to environmental changes and diseases.
- **Can lead to inbreeding**, increasing harmful mutations.

Example:

- **Cheetahs:**
    - Due to a past bottleneck, cheetahs have very low genetic variation, leading to poor reproductive success and high susceptibility to disease.



B. Founder Effect

The founder effect occurs when a small group of individuals establishes a new population in an isolated area. The gene pool is limited to the genetic variation present in the founding members.

How It Affects Genetic Diversity:

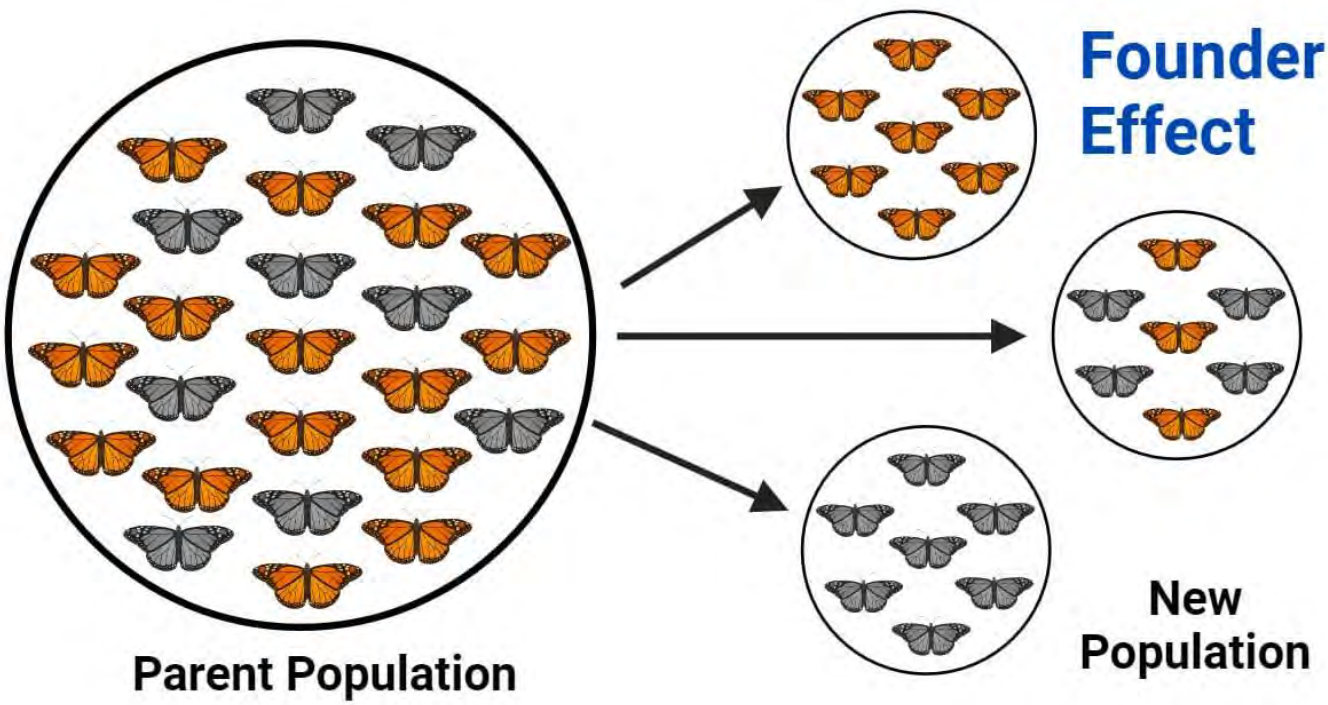
•**Reduces genetic variation**, as only a subset of alleles from the original population is present.

•**Can result in rare genetic disorders** if harmful alleles become more common.

Example:

•**Amish Population (Ellis–van Creveld Syndrome):**

- The Amish community, founded by a small group of European settlers, has a higher frequency of certain genetic disorders due to limited genetic diversity.



Measuring Genetic Diversity

Genetic diversity is a key indicator of a population's ability to adapt and evolve in response to environmental changes. It is measured using various molecular techniques that analyze variations at the DNA level. The four main methods used to measure genetic diversity are **Microsatellites (STRs), SNPs, Whole Genome Sequencing, and DNA Barcoding**.

1. Microsatellites (Short Tandem Repeats - STRs)

What Are Microsatellites?

Microsatellites, also known as Short Tandem Repeats (STRs), are repeating sequences of 2-6 base pairs found in non-coding regions of DNA. These sequences are highly variable among individuals, making them useful markers for genetic diversity studies.

How Are They Used?

DNA is extracted from a sample (blood, tissue, etc.).
PCR (Polymerase Chain Reaction) is used to amplify STR regions.
The number of repeats is analyzed using gel electrophoresis or capillary electrophoresis.
Genetic variation is assessed by comparing repeat numbers across individuals.

Applications:

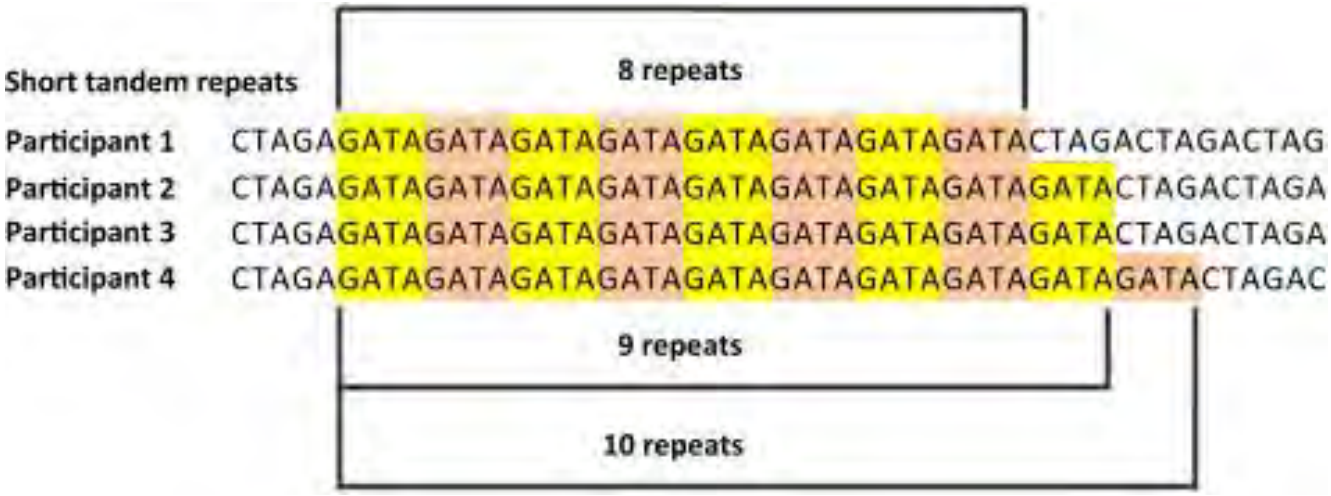
Used in population genetics to assess diversity and genetic relationships.

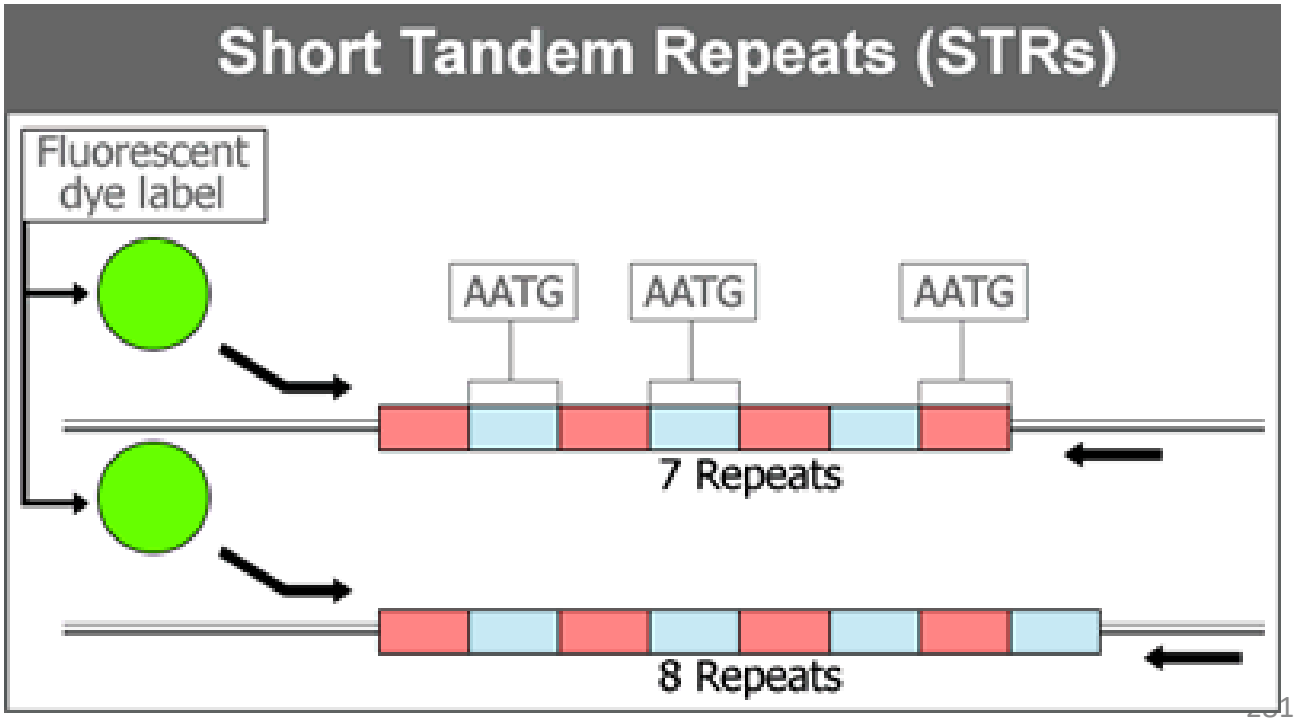
Commonly applied in forensic DNA profiling (e.g., crime scene investigations).

Useful in conservation biology for monitoring endangered species.

Example:

Forensic Science: The FBI uses 13 core STR markers for DNA fingerprinting in criminal investigations.





2. SNPs (Single Nucleotide Polymorphisms)

What Are SNPs?

Single Nucleotide Polymorphisms (SNPs) are variations in a single DNA base (A, T, C, or G) at a specific location in the genome. Unlike STRs, which involve repeated sequences, SNPs represent single base changes and are more stable across generations.

How Are They Used?

DNA is sequenced to identify SNPs at specific locations.

Computational analysis determines allele frequencies and genetic variation.

Genetic diversity is measured by comparing SNP differences among individuals or populations.

Applications:

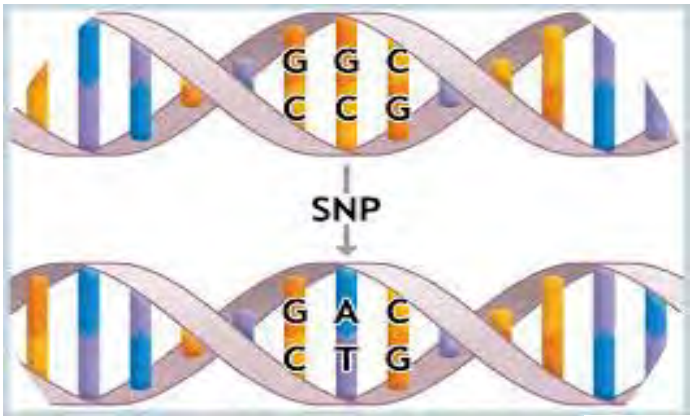
Used in medical genetics to identify disease-associated mutations.

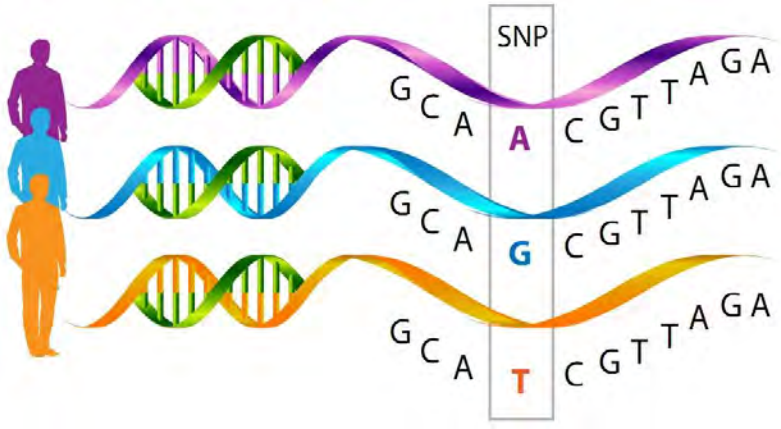
Helps in evolutionary biology to track human migration patterns.

Widely used in agriculture for crop and livestock breeding programs.

Example:

Personalized Medicine: SNP analysis helps predict an individual's risk for diseases like cancer and diabetes, allowing for personalized treatment.





3. Whole Genome Sequencing (WGS)

What Is Whole Genome Sequencing?

Whole Genome Sequencing (WGS) is a comprehensive method that determines the complete DNA sequence of an organism's genome. It provides the highest resolution for studying genetic diversity.

How Is It Done?

DNA is extracted and fragmented into smaller pieces.

Next-Generation Sequencing (NGS) technologies read the DNA sequences.

Bioinformatics tools analyze genetic variation across the entire genome.

Applications:

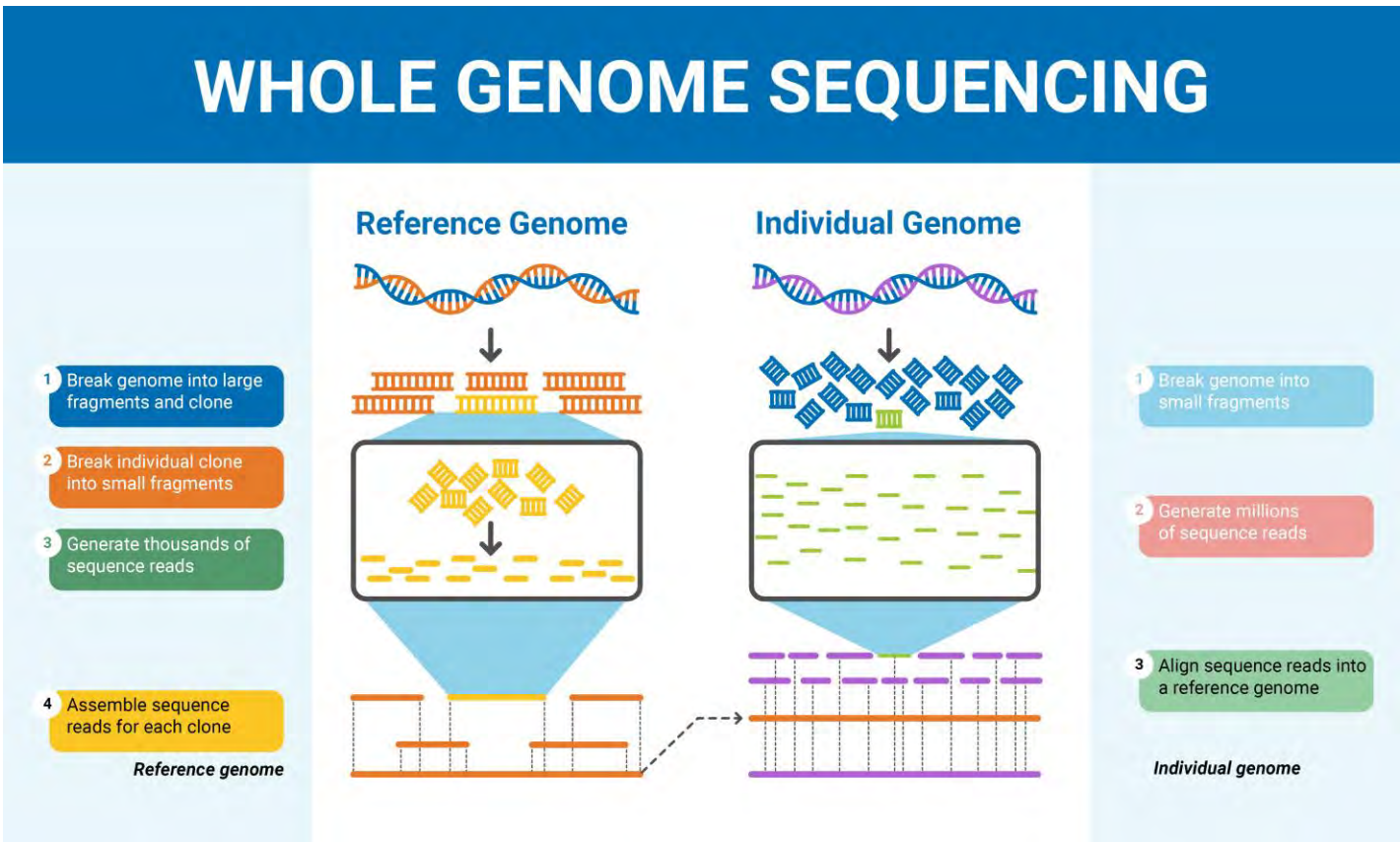
Provides the most detailed assessment of genetic diversity.

Helps in evolutionary studies and phylogenetics.

Used in precision medicine and rare disease diagnosis.

Example:

COVID-19 Variants: Scientists used WGS to track SARS-CoV-2 mutations and identify new variants of concern.



4. DNA Barcoding

What Is DNA Barcoding?

DNA barcoding is a technique used to identify species by analyzing short, standardized genetic regions. It is based on the idea that each species has a unique genetic "barcode."

Common Barcode Regions:

Animals: Mitochondrial *Cytochrome c Oxidase I (COI)* gene.

Plants: Chloroplast *rbcL* and *matK* genes.

Fungi & Bacteria: Internal Transcribed Spacer (*ITS*) and 16S rRNA genes.

How It Works:

DNA is extracted from a sample (e.g., leaf, feather, fish fillet).

PCR amplifies the barcode region.

Sequencing is performed to obtain the genetic code.

The sequence is compared to databases (e.g., BOLD, GenBank) for species identification.

Applications:

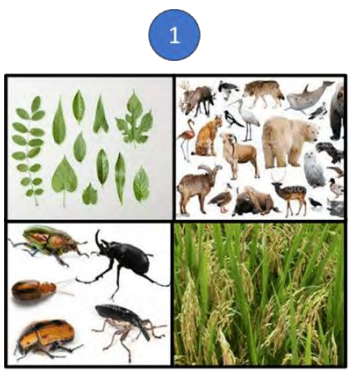
Used for species identification in biodiversity studies.

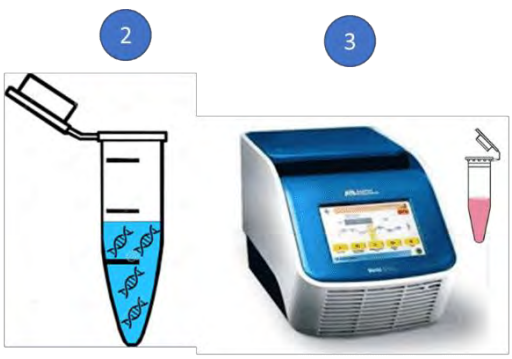
Helps detect food fraud and mislabeling in the seafood industry.

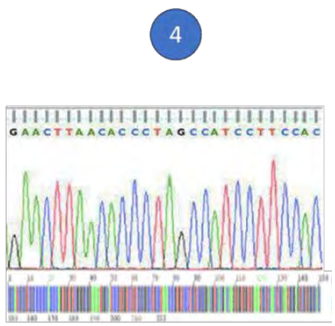
Aids in monitoring and preventing illegal wildlife trade.

Example:

Seafood Authentication: DNA barcoding revealed that nearly 20% of fish sold in markets are mislabeled, helping consumers avoid fraud.







Commonly Used Genes for DNA Barcoding

Different genetic markers are used for different groups of organisms based on mutation rates and conservation levels.

1. Animals: Mitochondrial COI (Cytochrome c Oxidase I) Gene

Why COI?

Mitochondrial DNA (mtDNA) evolves faster than nuclear DNA, allowing for better species differentiation.

The *COI* gene is **highly conserved within species but variable between species**.

Easy to amplify using universal primers.

Applications:

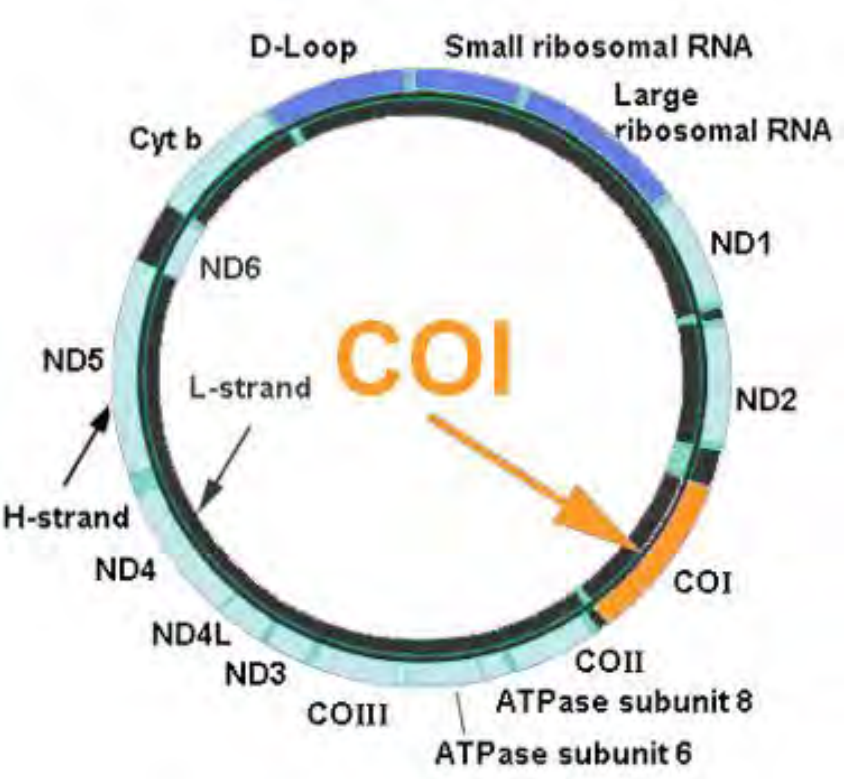
Identifying cryptic species (morphologically similar but genetically distinct).

Wildlife forensics (detecting illegal trade of endangered species).

Monitoring invasive species in ecosystems.

Example:

Fish Identification: DNA barcoding has been used to uncover seafood mislabeling, revealing that many products sold as "snapper" or "tuna" were actually different species.



2. Plants: Chloroplast rbcL and matK Genes

Why rbcL and matK?

Unlike animals, **mitochondrial DNA evolves too slowly in plants** for species identification.

Instead, genes from the chloroplast genome (**rbcL** and **matK**) are used because they exhibit **moderate variation** between species but are **highly conserved within a species**.

rbcL (Ribulose-1,5-bisphosphate carboxylase/oxygenase Large Subunit):

A universal plant barcode due to its **broad taxonomic coverage** and ease of sequencing.

matK (Maturase K):

Provides **higher species resolution** than *rbcL*, making it useful for distinguishing closely related species.

Applications:

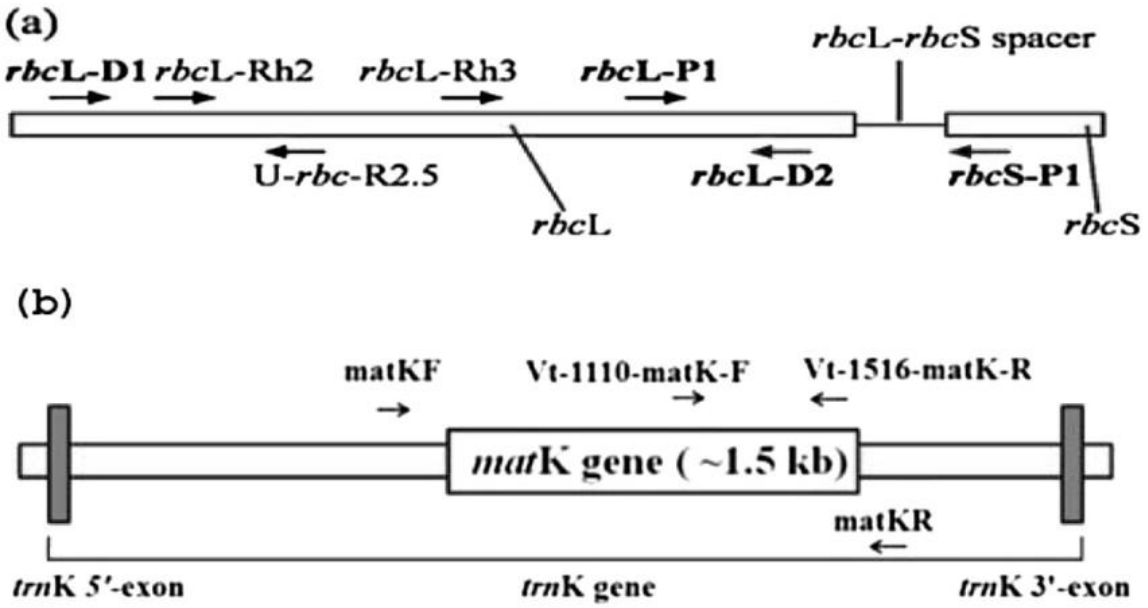
Authenticating medicinal plants (detecting adulteration in herbal products).

Monitoring illegal logging by identifying tree species from timber samples.

Studying plant biodiversity in rainforests and conservation areas.

Example:

Tea Authentication: DNA barcoding has been used to differentiate between high-value tea varieties (e.g., Darjeeling vs. ordinary black tea).



3. Fungi: Internal Transcribed Spacer (ITS) Region

Why ITS?

The *ITS* region, located between ribosomal RNA genes, shows **high variability among fungal species**.

Suitable for distinguishing fungi, including **pathogens, edible mushrooms, and decomposers**.

Applications:

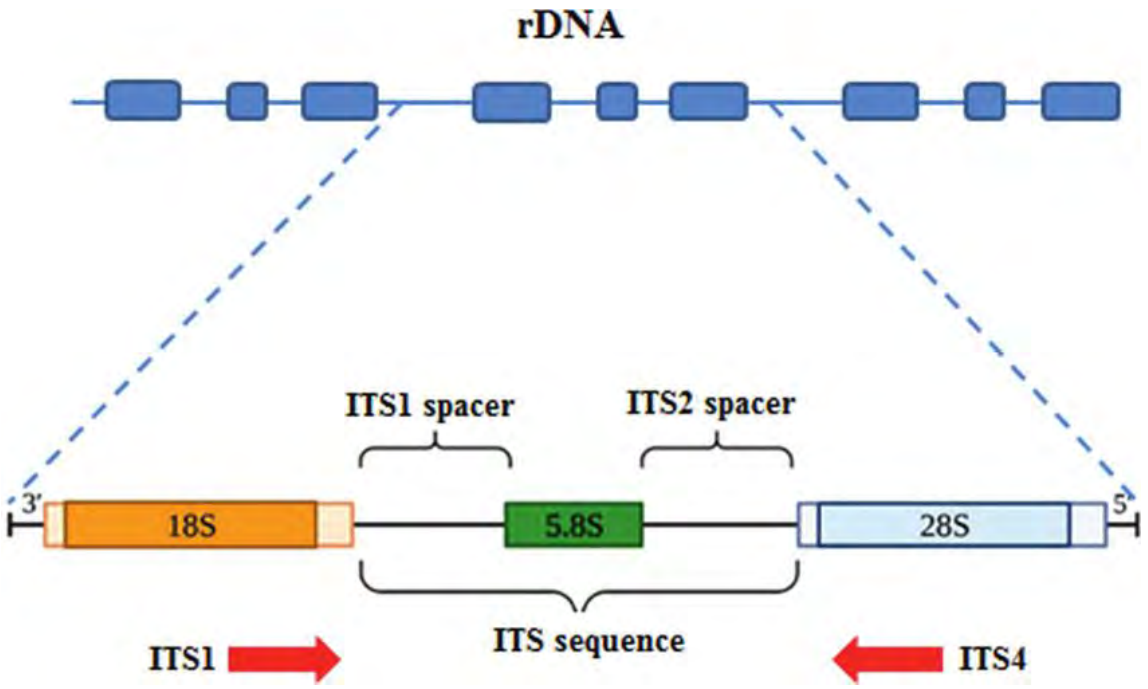
Identifying foodborne fungal contaminants (e.g., toxic mold in crops).

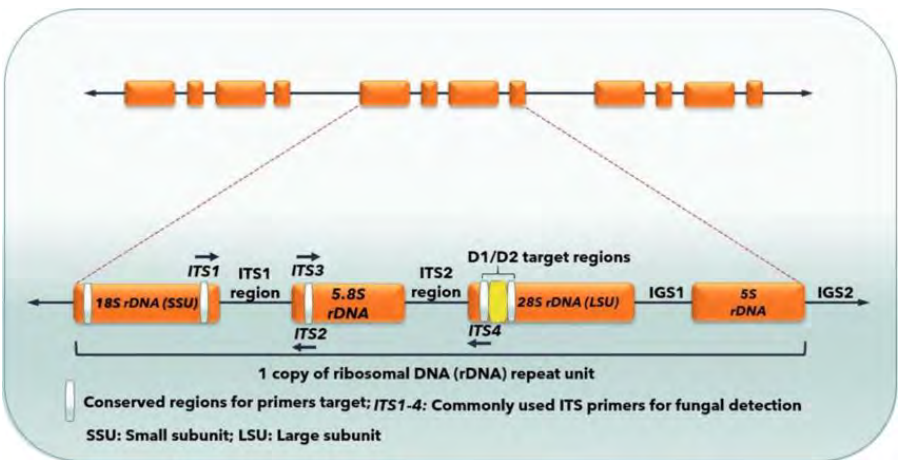
Discovering new fungal species in soil and forest ecosystems.

Detecting fungal infections in clinical diagnostics.

Example:

Medical Mycology: DNA barcoding has helped identify fungal pathogens like *Candida auris*, a drug-resistant yeast causing hospital outbreaks.





How DNA Barcoding Works

DNA barcoding follows a standardized process to identify species using genetic markers. The method involves six key steps: **sample collection, DNA extraction, PCR amplification, sequencing, database comparison, and species identification**.

1. Sample Collection

The first step involves **collecting a biological sample** from the organism of interest.

Samples can include:

Leaves, flowers, or seeds (for plants)

Feathers, hair, scales, or tissue (for animals)

Fungal spores, fruiting bodies, or soil samples (for fungi)

Water or environmental DNA (eDNA) (for microbial communities and aquatic species)

Example:

A scientist studying biodiversity in a rainforest might collect **leaves from different tree species** or **insect specimens** from a trap.

2. DNA Extraction

The collected sample is processed to **isolate DNA** using specialized chemicals and laboratory techniques.

Common DNA extraction methods:

CTAB method (for plants, using cetyltrimethylammonium bromide to remove polysaccharides)

Phenol-Chloroform extraction (for high-quality DNA from animals and fungi)

Commercial DNA extraction kits (e.g., Qiagen, Promega) for easy and rapid DNA isolation

Example:

A forensic scientist extracting DNA from a **feather** found at a crime scene to identify the bird species.

3. PCR Amplification (Using Barcode Primers)

Polymerase Chain Reaction (PCR) is used to **amplify** the barcode region of DNA.

Specific primers target the barcode genes:

COI (Cytochrome c Oxidase I) for animals

rbcL and matK for plants

ITS (Internal Transcribed Spacer) for fungi

The PCR reaction consists of:

- **Denaturation** (breaking DNA strands)
- **Annealing** (binding of primers to target regions)
- **Extension** (copying the DNA segment using DNA polymerase)

Example:

A marine biologist using PCR to amplify the COI gene from **fish muscle tissue** to identify the species of a seafood sample.

4. DNA Sequencing

The amplified DNA is **sequenced** using **Sanger sequencing** or **Next-Generation Sequencing (NGS)** techniques.

DNA sequencing reads the exact **order of nucleotides (A, T, C, G)** in the barcode region.

Output: A digital DNA sequence that represents the barcode for the species.

Example:

A conservationist sequencing the DNA of **tree samples** to determine if illegally logged timber is from a protected species.

5. Comparison with Barcode Databases (BOLD, GenBank)

The obtained DNA sequence is **compared against global reference databases**:

✔ **BOLD (Barcode of Life Data System)** – A dedicated DNA barcode database.

✔ **GenBank (NCBI Database)** – A comprehensive genetic sequence repository.

The sequence is matched with known species in the database using **bioinformatics tools** (e.g., BLAST).

Example:

A food safety officer analyzing the barcode sequence of **sushi samples** to check if they match the declared fish species in the database.

6. Species Identification

If the DNA sequence **matches a known species** in the database with high similarity (usually >98%), the species is identified.

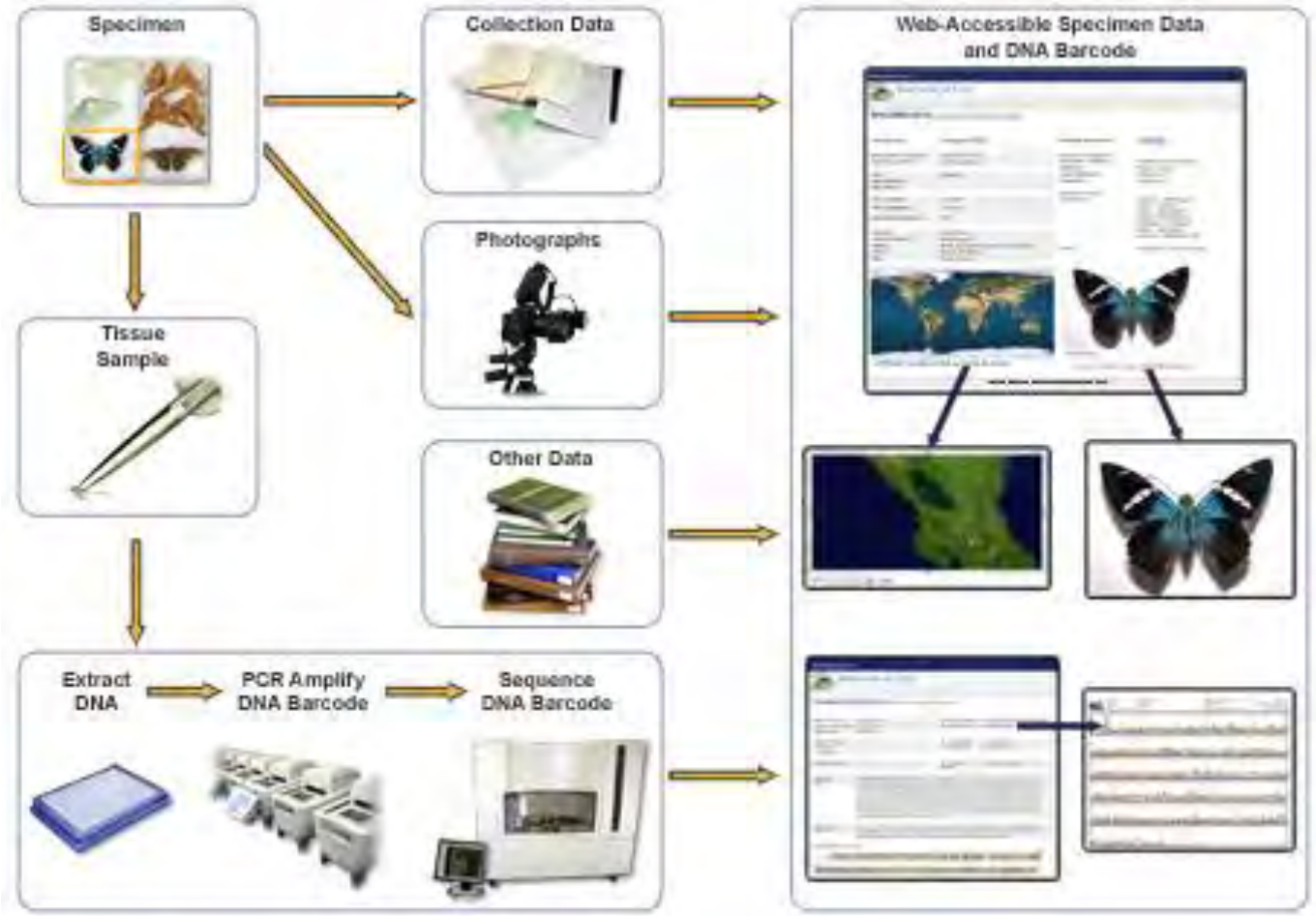
If there is **no exact match**, it may indicate:

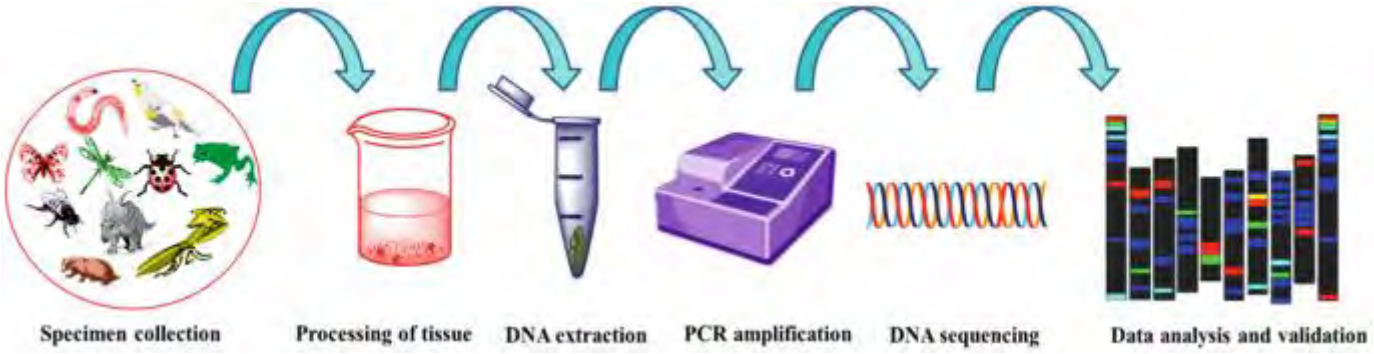
A **new species** (previously unrecorded in the database).

A **closely related species** that requires further analysis.

Example:

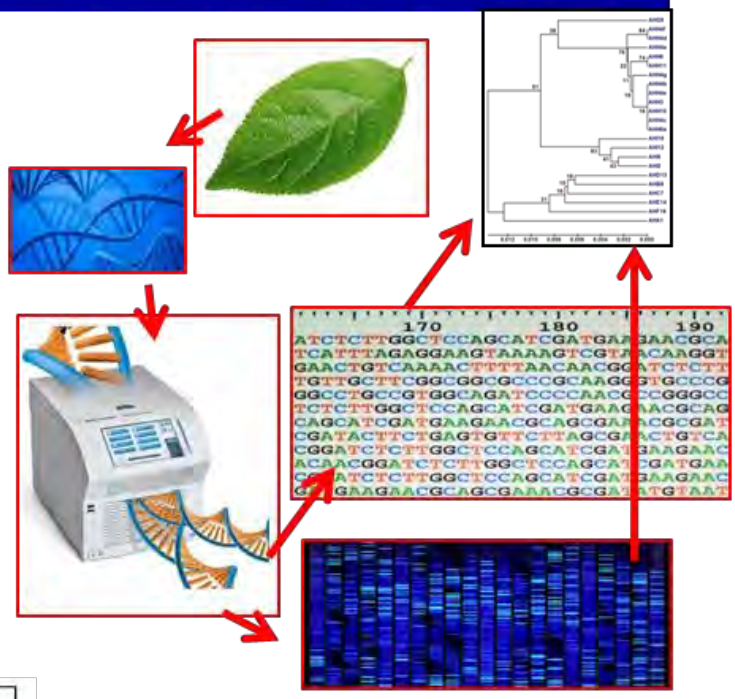
Researchers discovered **new frog species** in the Amazon rainforest using DNA barcoding when their DNA sequences didn’t match any known species in the database.


Specimen
Collection Data
Web-Accessible Specimen Data and DNA Barcode
Photographs
Tissue Sample
Other Data
Extract DNA
PCR Amplify DNA Barcode
Sequence DNA Barcode


Specimen collection
Processing of tissue
DNA extraction
PCR amplification
DNA sequencing
Data analysis and validation

GENETIC DIVERSITY

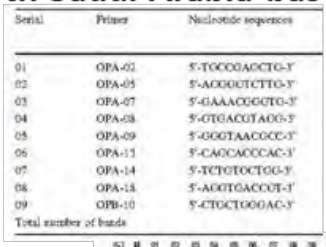
Genetic diversity is the total number of genetic characteristics in the genetic makeup of a species.



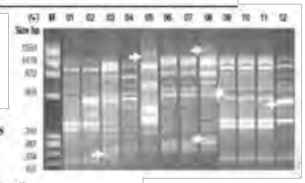
- **Molecular analyses comprise a large variety of DNA molecular markers, which can be employed for analysis of variation.**

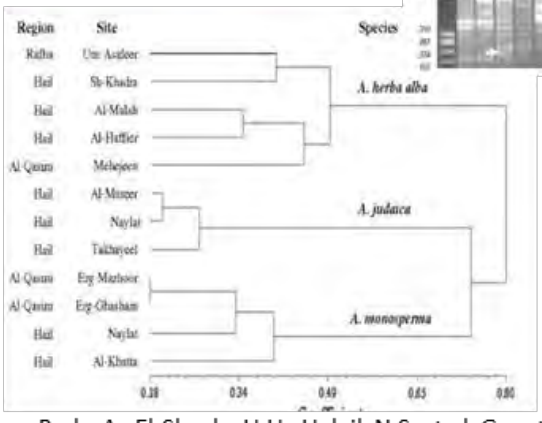
AFLP	Amplified Fragment Length Polymorphism
AP-PCR	Arbitrarily primed PCR
ARMS	Amplification Refractory Mutation System
ASAP	Arbitrary Signatures from Amplification
ASH	Allele-Specific Hybridization
ASLP	Amplified Sequence Length Polymorphism
ASO	Allele Specific Oligonucleotide
CAPS	Cleaved Amplification Polymorphic Sequence
CAS	Coupled Amplification and Sequencing
DAF	DNA Amplification Fingerprint
DGGE	Denaturing Gradient Gel Electrophoresis
GBA	Genetic Bit Analysis
IRAO	Inter-Retrotrasposon Amplified Polymorphism
ISSR	Inter-Simple Sequence Repeats
ISTR	Inverse Sequence-Tagged Repeats
MP-PCR	Microsatellite-Primed PCR
OLA	Oligonucleotide Ligation Assay
RAHM	Randomly Amplified Hybridizing Microsatellites
RAMPs	Randomly Amplified Microsatellite Polymorphisms
RAPD	Randomly Amplified Polymorphic DNA
RBIP	Retrotrasposon-Based Insertion Polymorphism
REF	Restriction Endonuclease Fingerprinting
REMAP	Retrotrasposon-Microsatellite Amplified Polymorphism
RFLP	Restriction Fragment Length Polymorphism
SAMPL	Selective Amplification of Polymorphic Loci
SCAR	Sequence Characterised Amplification Regions
SNP	Single Nucleotide Polymorphism
SPAR	Single Primer Amplification Reaction
SPLAT	Single Polymorphic Amplification Test
S-SAP	Sequence-Specific Amplification Polymorphisms
SSCP	Single Strand Conformation Polymorphism
SSLP	Single Sequence Length Polymorphism
SSR	Simple Sequence Repeats
STMS	Sequence-Tagged Microsatellite Site
STS	Sequence-Tagged-Site
TGGE	Thermal Gradient Gel Electrophoresis
VNTR	Variable Number Tandem Repeats
RAMS	Randomly Amplified Microsatellites

- **Genetic diversity of *Artemisia* in central and north Saudi Arabia based on RAPD**



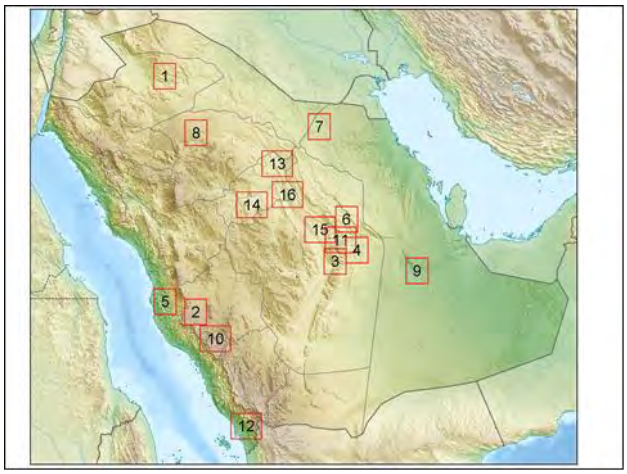
Serial	Primer	Nucleotide sequences
01	OPA-02	5'-TGCCGAGCTG-3'
02	OPA-05	5'-AGGGGTCTTG-3'
03	OPA-07	5'-GAAACGGGTG-3'
04	OPA-08	5'-GTGACGTAGG-3'
05	OPA-09	5'-GGGTAACGCC-3'
06	OPA-13	5'-CAGCACCCAC-3'
07	OPA-14	5'-TCTGTGCTGG-3'
08	OPA-18	5'-AGGTGACCGT-3'
09	OPB-10	5'-CTGCTGGGAC-3'
Total number of bands		

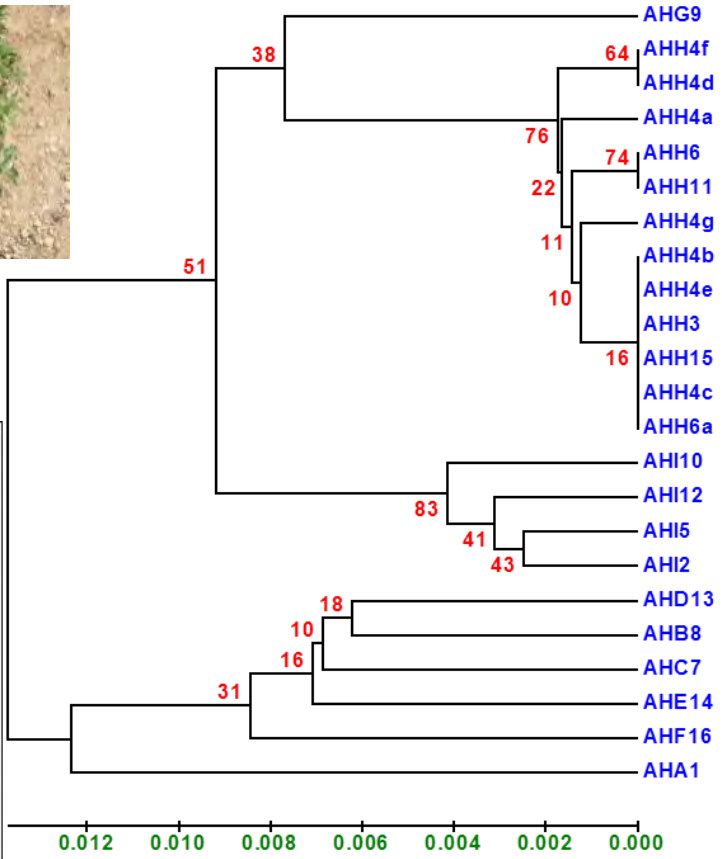




Badr, A., El-Shazly, H.H., Helail, N.S. et al. Genetic diversity of *Artemisia* populations in central and north Saudi Arabia based on morphological variation and RAPD polymorphism. Plant Syst Evol (2012) 298: 871)

Assessment of genetic diversity of *Anastatica hierochuntica* (kaff maryam) from Saudi Arabia based on Internal Transcribed Spacer sequences of nuclear ribosomal DNA gene





- *Anastatica hierochuntica* (Rose of Jericho) is among the common medicinal plants widely used in Hijaz, Najd, and Al Rub'Al Khali. The plant is prescribed in folk medicine for difficult labor, uterine hemorrhage and to facilitate the expulsion of dead fetuses. A total number of 23 population of Anastatica hierochuntica from Saudi Arabia were sequenced.
- The resulted UPGMA tree reveals that the populations of different geographic location sampled in the present study grouped into three major group.
- Group I consists of population from Hanifa valley, Summan, Rumah, Hair area, Riyadh, Khurma, and Khoris;
- Group II consists of population from Al-Baha, Jeedah, Ranyah and Zazan; and
- Group III consists of population from Hail, Darb Al Hafer, Qasim Buraydah, Afif, and Marat), and the groups were according to their geographic locations;
- however it was interesting to note that population collected from the geographic location of Haradh and Buseita (Tabarjal) and were nested within the group I and II respectively, which might be due to evolution under reproductive isolation and different environmental conditions, and this may be most probably due to long distance distribution, and possibility of genetic exchange among the populations of *Anastatica hierochuntica* distributed in Saudi Arabia.

293

DNA barcoding

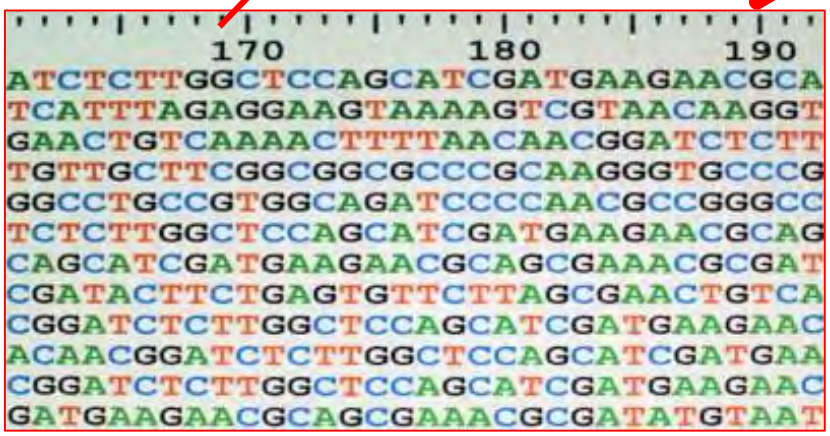
- DNA barcoding is a system for fast and accurate species identification that makes ecological system more accessible by using short DNA sequence instead of whole genome and is used for eukaryotes. The short DNA sequence is generated from standard region of genome known as marker. This marker is different for various species like CO1 cytochrome c oxidase 1 for animals, matK for plants and Internal Transcribed Spacer (ITS) for fungus. DNA barcoding has many applications in various fields like preserving natural resources, protecting endangered species, controlling agriculture pests, identifying disease vectors, monitoring water quality, authentication of natural health products and identification of medicinal plants.

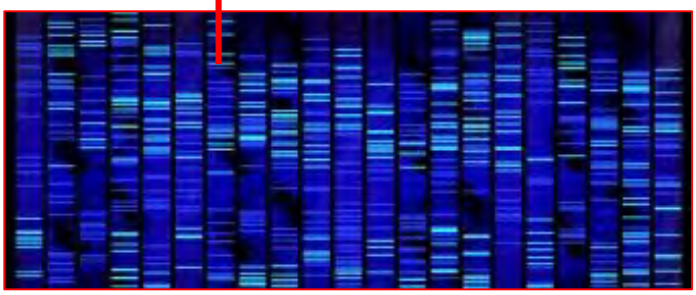
❖ **DNA barcoding can speed up identification of species.**

❖ **DNA barcoding can provide an avenue to encourage new participants into taxonomy.**

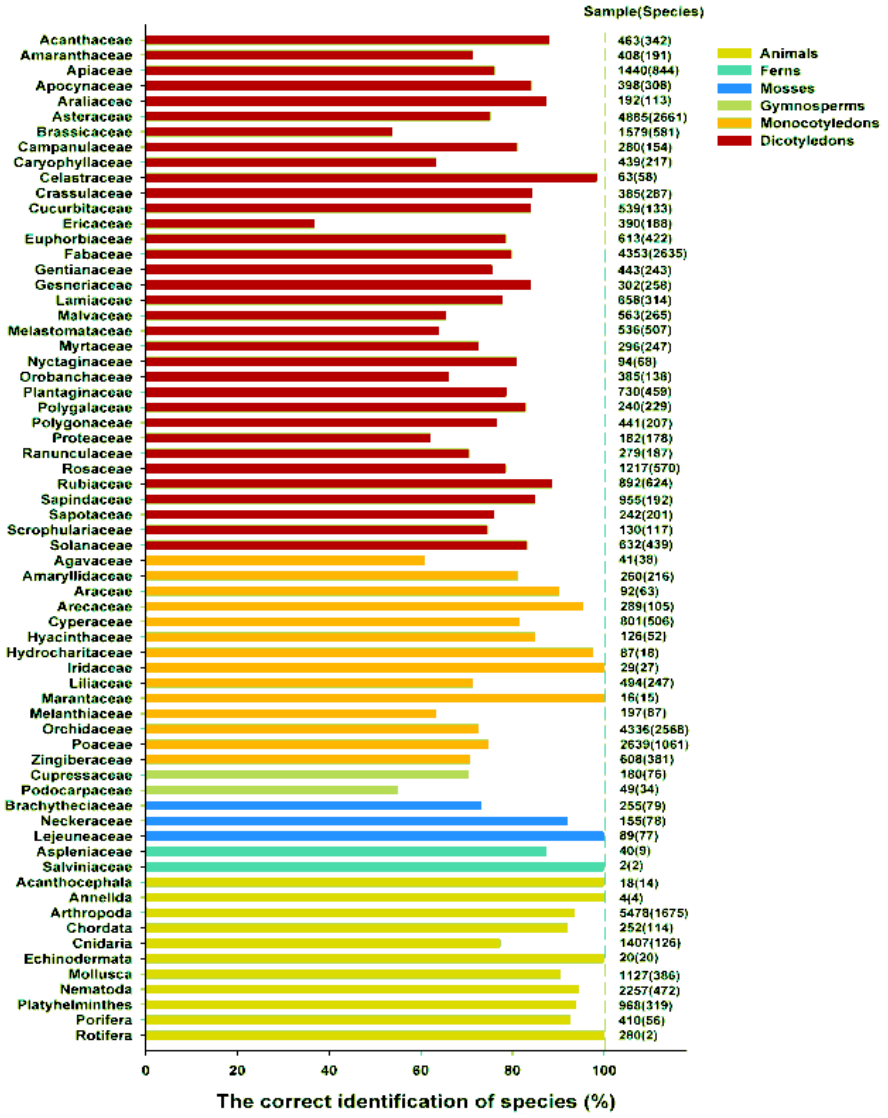
❖ **Raw drug authentication / Medicinal plant identification or authentication**

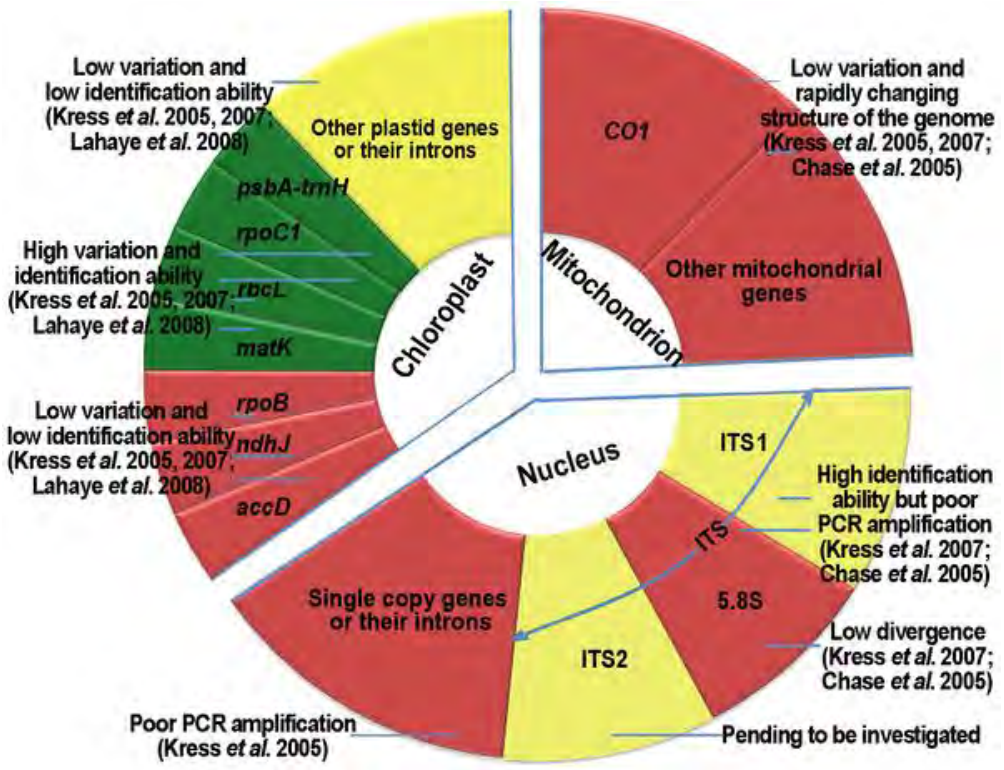
- In DNA barcoding, complete data set can be obtained from a single specimen irrespective to morphological or life stage characters.
- The core idea of DNA barcoding is based on the fact that the highly conserved stretches of DNA, either coding or
- non coding regions, vary at very minor degree during the evolution within the species.
- Sequences suggested to be useful in DNA barcoding include cytoplasmic mitochondrial DNA (e.g. cox1) and chloroplast DNA (e.g. rbcL, trnL-F, matK, ndhF, and atpB rbcL), and nuclear DNA (ITS)
- The term “DNA barcode” for global species identification was first coined by Hebert in 2003.
- The ideal DNA barcode region is reliably amplified and sequenced across large assemblages of taxa and provides a high level of species discrimination






Phylogeny Programs


Sample(Species)
Acanthaceae 463(342)
Amaranthaceae 408(191)
Apiaceae 1440(844)
Apocynaceae 398(308)
Araliaceae 192(113)
Asteraceae 4885(2661)
Brassicaceae 1579(581)
Campanulaceae 280(154)
Caryophyllaceae 439(217)
Celastraceae 63(58)
Crassulaceae 385(287)
Cucurbitaceae 539(133)
Ericaceae 390(188)
Euphorbiaceae 613(422)
Fabaceae 4353(2635)
Gentianaceae 443(243)
Gesneriaceae 302(258)
Lamiaceae 658(314)
Malvaceae 563(265)
Melastomataceae 536(507)
Myrtaceae 296(247)
Nyctaginaceae 94(68)
Orobanchaceae 385(138)
Plantaginaceae 730(459)
Polygalaceae 240(229)
Polygonaceae 441(207)
Proteaceae 182(178)
Ranunculaceae 279(187)
Rosaceae 1217(570)
Rubiaceae 892(624)
Sapindaceae 955(192)
Sapotaceae 242(201)
Scrophulariaceae 130(117)
Solanaceae 632(439)
Agavaceae 41(38)
Amaryllidaceae 260(216)
Araceae 92(63)
Arecaceae 289(105)
Cyperaceae 801(506)
Hyacinthaceae 126(52)
Hydrocharitaceae 87(18)
Iridaceae 29(27)
Liliaceae 494(247)
Marantaceae 16(15)
Melanthiaceae 197(87)
Orchidaceae 4336(2568)
Poaceae 2639(1061)
Zingiberaceae 608(381)
Cupressaceae 180(76)
Podocarpaceae 49(34)
Brachytheciaceae 255(79)
Neckeraceae 155(78)
Lejeuneaceae 89(77)
Aspleniaceae 40(9)
Salviniaceae 2(2)
Acanthocephala 18(14)
Annelida 4(4)
Arthropoda 5478(1675)
Chordata 252(114)
Cnidaria 1407(126)
Echinodermata 20(20)
Mollusca 1127(386)
Nematoda 2257(472)
Platyhelminthes 968(319)
Porifera 410(56)
Rotifera 280(2)
Animals
Ferns
Mosses
Gymnosperms
Monocotyledons
Dicotyledons
0
20
40
60
80
100
The correct identification of species (%)


Low variation and low identification ability (Kress et al. 2005, 2007; Lahaye et al. 2008)
Other plastid genes or their introns
psbA-trnH
rpoC1
High variation and identification ability (Kress et al. 2005, 2007; Lahaye et al. 2008)
rbcL
matK
Low variation and low identification ability (Kress et al. 2005, 2007; Lahaye et al. 2008)
rpoB
ndhJ
accD
Chloroplast
CO1
Low variation and rapidly changing structure of the genome (Kress et al. 2005, 2007; Chase et al. 2005)
Other mitochondrial genes
Mitochondrion
Nucleus
ITS1
High identification ability but poor PCR amplification (Kress et al. 2007; Chase et al. 2005)
ITS
5.8S
Low divergence (Kress et al. 2007; Chase et al. 2005)
ITS2
Pending to be investigated
Single copy genes or their introns
Poor PCR amplification (Kress et al. 2005)

Chapter 8- Interpretation of molecular phylogenetic trees

Advance Experimental Taxonomy

Chapter 8- Interpretation of molecular phylogenetic trees

1. What is a Phylogenetic Tree?

A **phylogenetic tree** is a branching diagram that represents the evolutionary relationships among species based on common ancestry. It helps visualize how species have evolved from common ancestors over time.

2. Key Features of a Phylogenetic Tree

Branches – Represent evolutionary lineages.

Nodes – Points where a single lineage splits into two or more; represents a common ancestor.

Root – The most ancestral node of the tree, showing the common origin of all organisms in the tree.

Clades – Groups of organisms that include a common ancestor and all its descendants (monophyletic group).

Outgroup – A species or group that is closely related to but not part of the main group being studied.

Branch Length – Can represent genetic changes or time elapsed since divergence.

3. Types of Phylogenetic Trees

Rooted Tree – Shows a common ancestor at the base and evolutionary divergence.

Unrooted Tree – Does not indicate a common ancestor, but shows relationships.

Cladogram – Only shows branching patterns; does not represent time or genetic differences.

Phylogram – Branch lengths are proportional to the number of genetic changes.

Chronogram – Branch lengths represent evolutionary time.

4. Methods to Construct Phylogenetic Trees

Morphological Data (based on traits such morphology, physiology, biochemistry etc.)

Molecular Data (DNA, RNA, protein sequences)

Computational Methods:

- **Maximum Parsimony** – Simplest explanation with the fewest changes.
- **Maximum Likelihood** – Uses probability models to determine the most likely tree.
- **Bayesian Inference** – Uses statistical probability to estimate tree structures.

5. Importance of Phylogenetic Trees

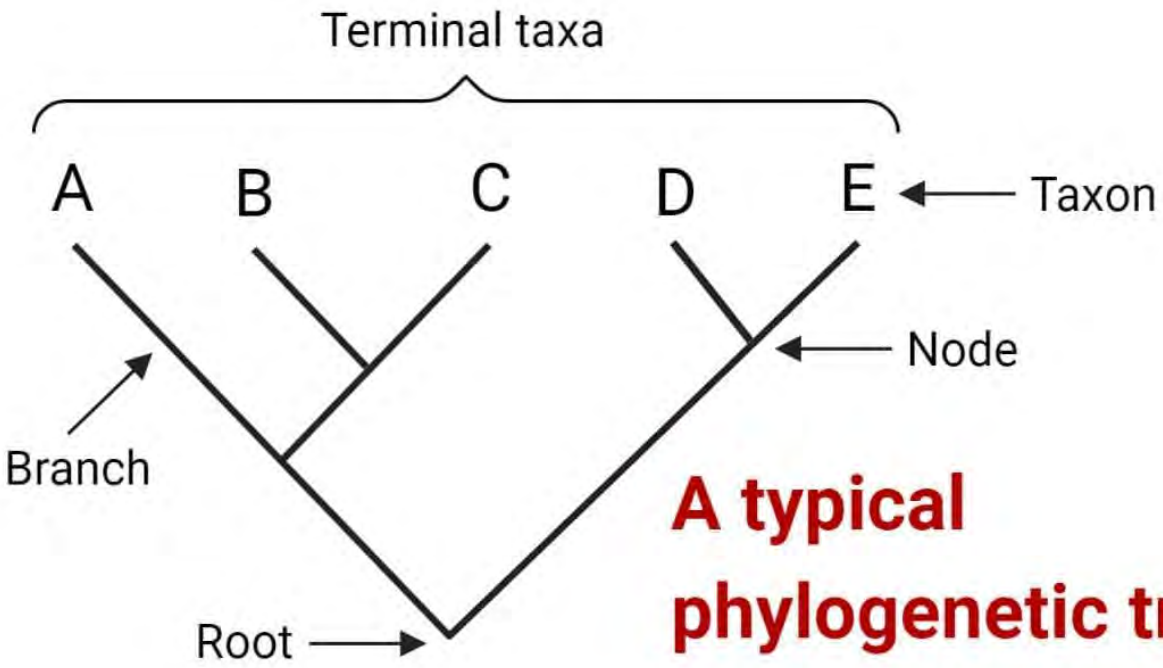
Understanding evolutionary relationships among species.

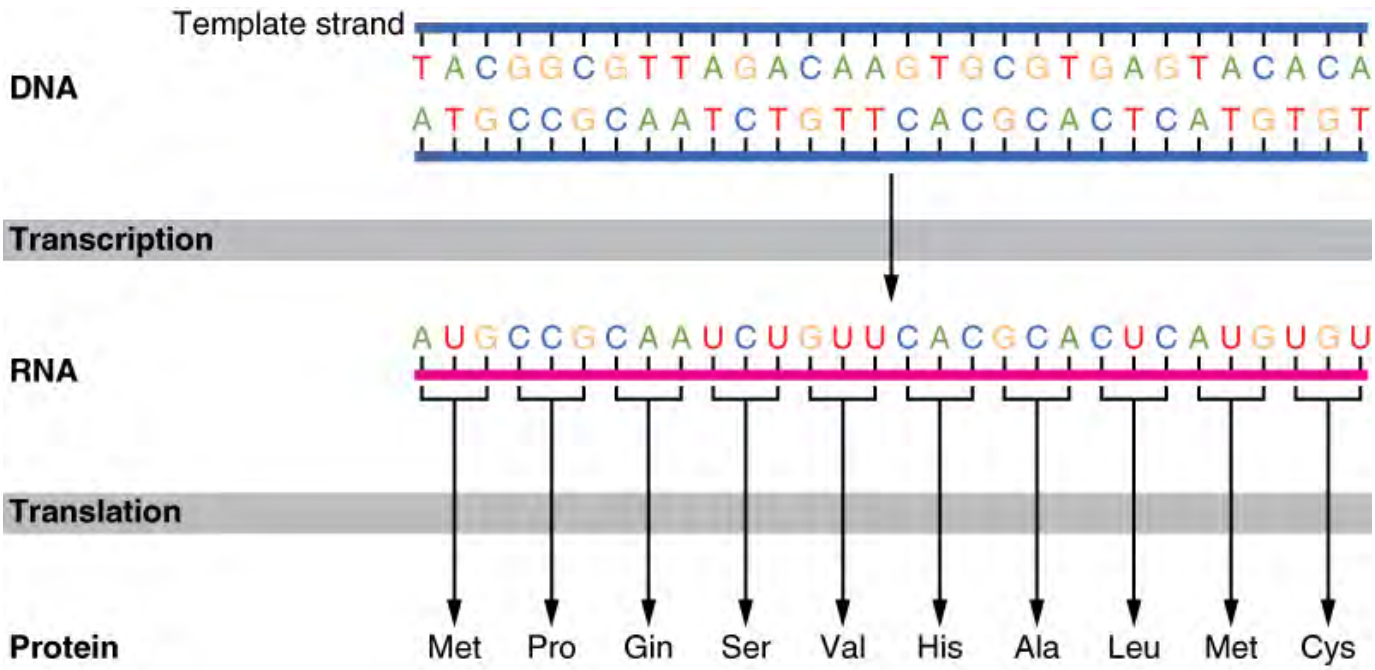
Tracing the origins and evolution of diseases (e.g., tracking viruses like COVID-19).

Identifying common ancestors and evolutionary traits.

Classifying organisms systematically (taxonomy).

Helping in conservation efforts by identifying evolutionary distinct species.


Terminal taxa
A
B
C
D
E
Taxon
Node
Branch
Root
A typical phylogenetic tree


Template strand
DNA
T A C G G C G T T A G A C A A G T G C G T G A G T A C A C A
A T G C C G C A A T C T G T T C A C G C A C T C A T G T G T
Transcription
RNA
A U G C C G C A A U C U G U U C A C G C A C U C A U G U G U
Translation
Protein
Met
Pro
Gin
Ser
Val
His
Ala
Leu
Met
Cys

Construction of Phylogenetic Trees based on; Morphological Data

1. Phenetics (Numerical Taxonomy)

Definition

Phenetics, also known as **Numerical Taxonomy**, is a classification method that groups organisms based on **overall similarity** rather than evolutionary relationships. It uses **quantitative** data (measurable traits) to create objective classifications.

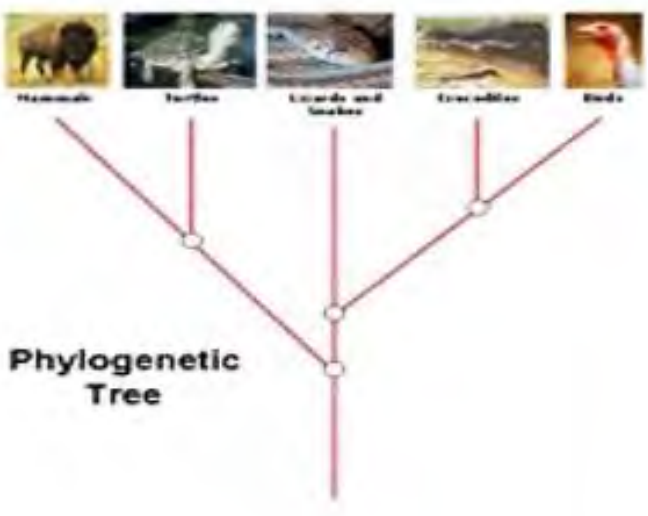
Key Features of Phenetics

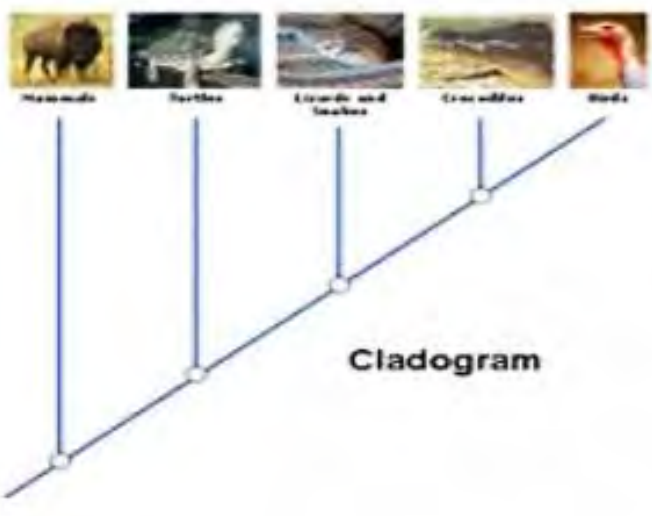
Based on **observable characteristics** (morphology, physiology, biochemistry, etc.).

Uses **mathematical models and statistical techniques** for classification.

Does **not consider evolutionary history** (unlike cladistics and phylogenetics).

Relies on **similarity indices** (e.g., Jaccard's coefficient, Euclidean distance).





Principles of Phenetics (Numerical Taxonomy)

Main Principles of Phenetics

1.1. Equal Weightage to All Characteristics

All traits (morphological, physiological, biochemical, genetic) are considered equally important.

No single characteristic is given priority over others.

1.2. Overall Similarity-Based Classification

Organisms are grouped based on their total similarity rather than evolutionary relationships.

Similarity is calculated using **statistical and mathematical models**.

1.3. Quantitative Approach

Uses **numerical values** instead of subjective descriptions.

Data is converted into a numerical matrix for analysis.

1.4. Use of Statistical Methods

Various **statistical techniques** are used to compare organisms:

Jaccard's Similarity Coefficient (for binary data)

Euclidean Distance (for continuous traits)

Cluster Analysis (to form groups)

1.5. Phenograms Instead of Phylogenies

The output is a **phenogram**, a tree-like diagram that shows similarity.

Unlike a phylogenetic tree, a phenogram **does not show evolutionary history**.

1.6. Large Dataset Handling

Phenetics allows classification based on a **large number of characteristics**.

This is useful in microbial taxonomy, where thousands of traits are analyzed simultaneously.

Selection of Characters in Phenetics (OTU – Operational Taxonomic Unit)

In **Phenetics (Numerical Taxonomy)**, the classification of organisms is based on overall similarity using **Operational Taxonomic Units (OTUs)**. OTUs are the basic units of classification, which can be **species, genera, populations, or even individual organisms**.

1. What is an OTU?

An **Operational Taxonomic Unit (OTU)** is a group of organisms classified together based on similarities in characteristics.

It can be **species, genus, strain, or any hierarchical taxonomic level**.

Used extensively in microbial taxonomy, genetics, and ecological studies.

2. Selection of Characters for OTU Classification

In phenetics, the selection of **characters (traits)** is crucial for an accurate and objective classification. The selected traits must be:

A. Types of Characters Considered

Morphological Characters – External structures (size, shape, color, number of segments, etc.).

Physiological & Biochemical Characters – Growth rate, enzyme production, metabolic activity.

Genetic Characters – DNA sequences, protein structures, chromosome numbers.

Ecological Characters – Habitat preference, distribution, behavior.

Molecular Characters – Protein electrophoresis, DNA fingerprinting.

B. Criteria for Selecting Characters

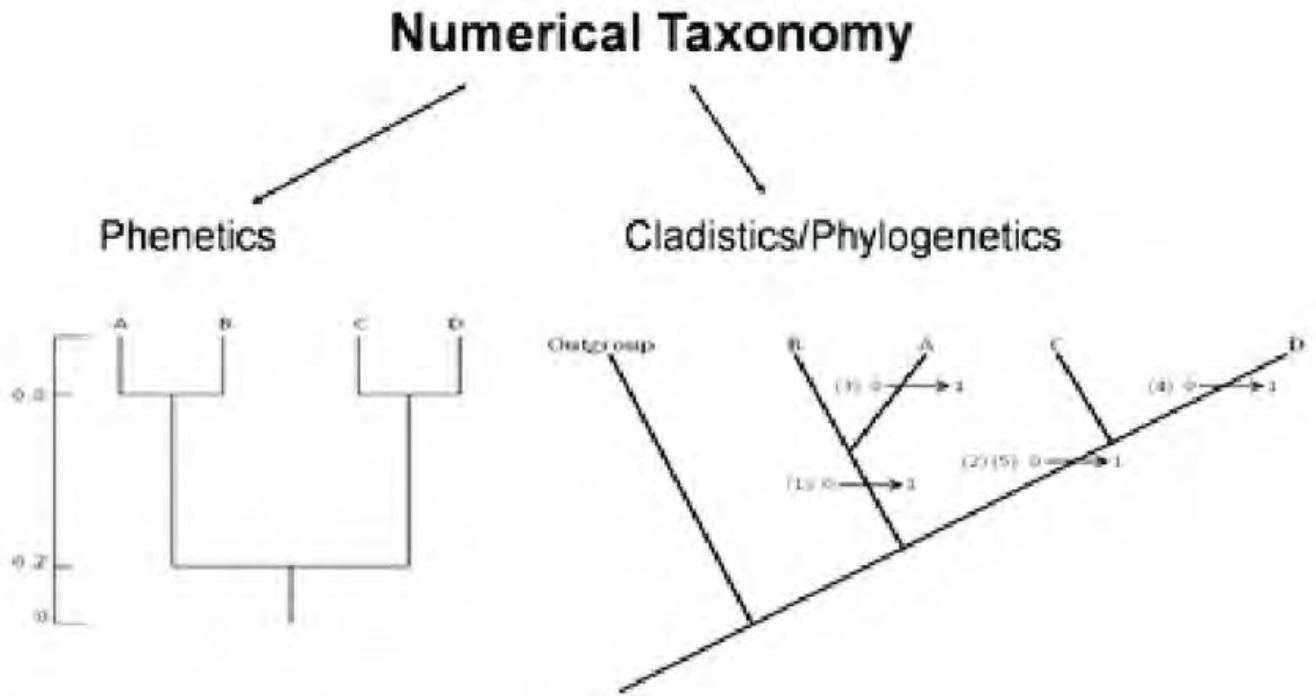
Measurable & Quantifiable – Traits should be numerically comparable (e.g., height, length, enzyme activity).

Stable & Consistent – The selected traits should not vary due to environmental conditions.

Genetically Controlled – Traits should be heritable and not influenced by temporary environmental factors.

Large Sample Size – Data from multiple individuals should be collected for accuracy.

Non-Redundant – Avoid selecting traits that provide redundant information (e.g., leaf size and leaf length may be correlated).



Continuous vs. Discontinuous Variation

Variation → Differences in traits among plants.

Types of Variation:

Continuous Variation → Traits show a **range** (e.g., leaf size, plant height).

Discontinuous Variation → Traits have **distinct categories** (e.g., flower color, seed shape).

Continuous Variation

Gradual differences, no fixed groups.

Controlled by multiple genes (**polygenic inheritance**).

Influenced by environment.

Forms a **bell-shaped curve**.

Examples:

Leaf size in the same species.

Plant height in pea plants.

Fruit weight in mango trees.

Discontinuous Variation

Clear-cut categories, no intermediates.

Controlled by one or few genes (**monogenic inheritance**).

Not influenced by environment.

Data shown as **bar graphs**.

Examples:

Flower color in pea plants (Purple or White).

Seed shape in peas (Round or Wrinkled).

Pod color in beans (Green or Yellow).

Character Weighting in Phenetics

•**Phenetics** → Classification based on overall similarity.

•**Character Weighting** → Assigning importance to traits.

•**Why is it needed?**

- Not all traits are equally important.
- Some traits provide more useful information.
- Avoids bias from less significant traits.

Types of Characters

•**Morphological** → Leaf shape, flower structure.

•**Physiological** → Growth rate, drought tolerance.

•**Biochemical** → Enzyme activity, protein markers.

•**Genetic** → DNA sequences, chromosome number.

Methods of Character Weighting

Equal Weighting → All traits treated equally.

Differential Weighting → More important traits get higher weight.

Statistical Weighting → Uses mathematical models (e.g., PCA).

Removing Redundant Traits → Avoids over-representation of similar traits.

Importance of Character Weighting

Improves accuracy in classification.

Reduces errors and bias.

Helps in plant taxonomy and evolutionary studies.

Coding in Phenetics

Phenetics (Numerical Taxonomy) → Classification based on overall similarity.

Coding → Converting biological traits into numerical data for analysis.

Why is Coding Important?

Standardizes data for comparison.

Allows statistical analysis.

Helps in cluster analysis and tree construction.

Types of Data Used

Morphological → Leaf shape, flower color.

Physiological → Growth rate, drought resistance.

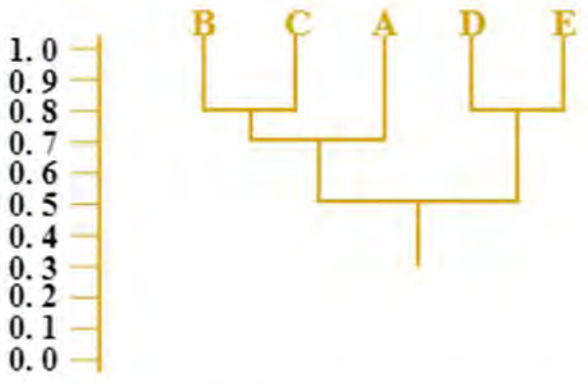
Biochemical → Protein markers, enzyme activity.

Molecular → DNA sequences, genetic markers.

Characters

Taxa	1	2	3	4	5	6	7	8	9	10
A	1	1	1	1	1	1	1	1	1	0
B	1	1	1	0	0	1	1	1	0	0
C	1	1	1	1	0	1	1	1	0	1
D	1	1	0	0	0	1	0	0	0	0
E	1	1	0	0	0	0	1	0	0	0

	A	B	C	D	E
A	-				
B	0.7	-			
C	0.7	0.8	-		
D	0.4	0.7	0.5	-	
E	0.4	0.7	0.5	0.8	



Methods of Coding in Phenetics

- **Binary Coding (0,1)** → Presence (1) or absence (0) of a trait.
- **Multistate Coding (0,1,2,3…)** → Different states of a trait.
- **Quantitative Coding** → Measurable traits converted into numerical ranges.
- **Weighting in Coding** → More significant traits may get higher weight.

Example of Binary Coding:

Importance of Coding

Enables objective classification.
Improves computational analysis.
Helps in phylogenetic tree construction.

Trait	Species A	Species B	Species C
Red Flower (1) / White (0)	1	0	1
Large Leaf (1) / Small (0)	1	1	0

Phenetic Methodology (Steps 1-4)

Phenetics → Classification based on overall similarity.
Objective → Group organisms using measurable traits.
Steps in Phenetic Methodology:
1 **Selection of OTUs** → Species or individual organisms.
2 **Selection of Characters** → Morphological, physiological, biochemical, genetic.
3 **Character Coding** → Converting traits into numbers (binary, multistate, quantitative).
4 **Character Weighting** → Assigning importance to traits (equal or differential).

Similarity, Clustering & Classification (Steps 5-7)

5 **Similarity Measurement**
Uses statistical methods to compare OTUs.
Common Methods: Simple Matching Coefficient (SMC), Jaccard's Coefficient, Euclidean Distance.
6 **Clustering & Tree Construction**
Groups similar OTUs into phenograms.
Techniques Used: UPGMA, Neighbor-Joining Method.
7 **Interpretation & Classification**
Identifying meaningful groups (taxa).
Comparing with traditional taxonomy for validation.

Importance

- Objective and data-driven classification.
- Used in plant taxonomy and microbial studies.
- Helps in computational taxonomy.

How to Perform Phenetic Analysis

1 **Select OTUs** → Choose species or organisms for classification.
2 **Select Characters** → Identify traits (morphological, biochemical, genetic).
3 **Character Coding** → Convert traits into numerical data (binary, multistate, or quantitative).
4 **Calculate Similarity** → Use statistical methods like Simple Matching Coefficient (SMC) or Jaccard's Coefficient.
5 **Cluster Analysis** → Group similar organisms using UPGMA or Neighbor-Joining Method.
6 **Interpret Results** → Analyze the phenogram to classify organisms.

Example of Phenetic Analysis

Trait	Species A	Species B	Species C	Species D
Leaf Shape (Round-1, Oval-0)	1	0	1	1
Flower Color (Red-1, Yellow-0)	1	1	0	1
Seed Type (Smooth-1, Wrinkled-0)	0	1	0	0
Plant Height (Tall-1, Short-0)	1	0	1	1

$$SMC = \frac{\text{Number of Matching Traits}}{\text{Total Traits}}$$

1. Similarity Calculation

SMC Formula:

$$SMC = \frac{\text{Number of Matching Traits}}{\text{Total Traits}}$$

Species Pair	Matching Traits	SMC Value
A & B	2/4 = 0.50	50% Similarity
A & C	3/4 = 0.75	75% Similarity
A & D	4/4 = 1.00	100% Similarity
B & C	1/4 = 0.25	25% Similarity
B & D	2/4 = 0.50	50% Similarity
C & D	3/4 = 0.75	75% Similarity

2. Clustering and Interpretation

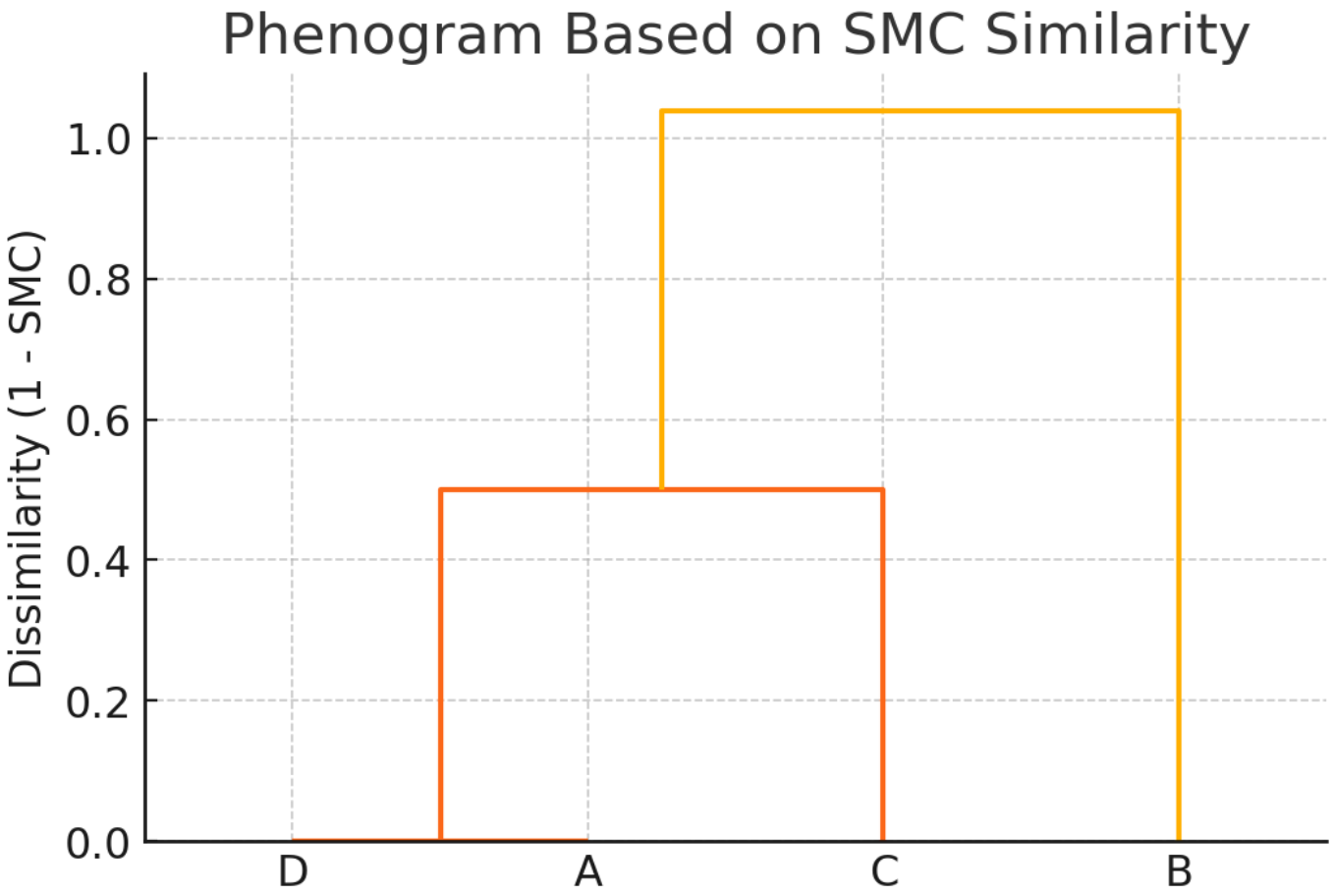
- **Species A and D are most similar (100%)** → Likely belong to the same group.
- **Species A and C have high similarity (75%)** → Closely related.
- **Species B and C have low similarity (25%)** → Distantly related.
- **A phenogram would show A and D clustering together first, then C, and finally B as the most distant.**

3. Conclusion

✓ **Species A and D** are closely related.
✓ **Species C is somewhat related to A and D.**
✓ **Species B is the most different.**

Phenogram based on the **SMC similarity values** from the table:

- **Species A & D** (100% Similarity) are clustered together first.
- **Species C** (75% Similarity with A & D) joins next.
- **Species B** (least similar) joins last.



Phenogram (dendrogram) based on the similarity values:

- **Species A and D** are clustered together first.
- **Species C** joins next, showing a moderate relationship with A and D.
- **Species B** is the most distinct and joins last.

Introduction to Cladistics

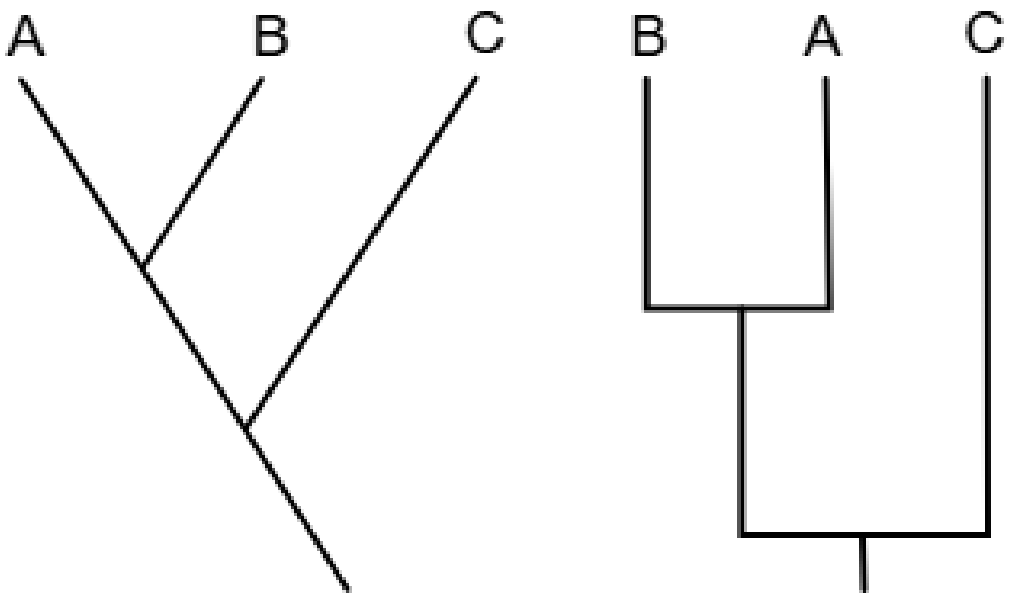
Definition: Cladistics is a method of classifying organisms based on common ancestry. It groups species into **clades** (ancestor + all descendants).

Key Concepts:

- **Clade:** Group of organisms with a common ancestor.
- **Cladogram:** A branching diagram showing evolutionary relationships.
- **Derived Character:** A trait unique to a clade.

Data Used in Cladistics:

- **Morphology:** Physical traits (e.g., bone structure).
- **Genetics:** DNA and molecular sequences.
- **Behavioral Data:** Evolutionary significance of behaviors.



Methodology of Cladistic Analysis

A. Evolutionary Assumptions

Plants evolved from a **common ancestor**.

Evolution follows **descent with modification**.

Cladistics uses **parsimony** (simplest explanation preferred).

B. Selection of Taxa

Choose related plant groups (e.g., **Algae, Bryophytes, Ferns, Gymnosperms, Angiosperms**).

C. Selection of Characters

Use **morphological traits** (e.g., presence of vascular tissue, seeds, flowers).

Example: **Leaf type, seed presence, reproductive structures**.

D. Coding of Character States

Convert traits into **binary (0,1) or multistate (0,1,2...)** codes.

Example: **Seeds (0 = absent, 1 = present)**.

E. Homologies of Characters and Character States

Identify **shared derived traits (synapomorphies)**.

Example: **Vascular tissue evolved before seeds**.

F. Construction of Morphocline

Arrange character states in an **evolutionary sequence**.

Example: **Algae → Bryophytes → Ferns → Gymnosperms → Angiosperms**.

G. Determination of Polarity

Identify if a trait is **ancestral (plesiomorphic)** or **derived (apomorphic)**.

Use an **outgroup (e.g., Algae)** for comparison.

H. Construction of Data Matrix

- Example Table:

I. Construction of Cladogram

- Use the **data matrix** to create a **cladogram**.
- Example:
  - Algae (Outgroup) → Bryophytes → Ferns → Gymnosperms → Angiosperms

Taxon	Vascular Tissue	Seeds	Flowers
Algae	0	0	0
Bryophytes	0	0	0
Ferns	1	0	0
Gymnosperms	1	1	0
Angiosperms	1	1	1

Construction of cladogram.

A **cladogram** is a diagram that represents evolutionary relationships among organisms based on shared derived characteristics.

1. Selected Plants

Saudi Arabia has diverse plant life, including xerophytic (desert-adapted) and halophytic (salt-tolerant) species. Here are six plant species with evolutionary significance:

1.Bryophytes (Mosses, Liverworts) – Example: *Riccia sp.*

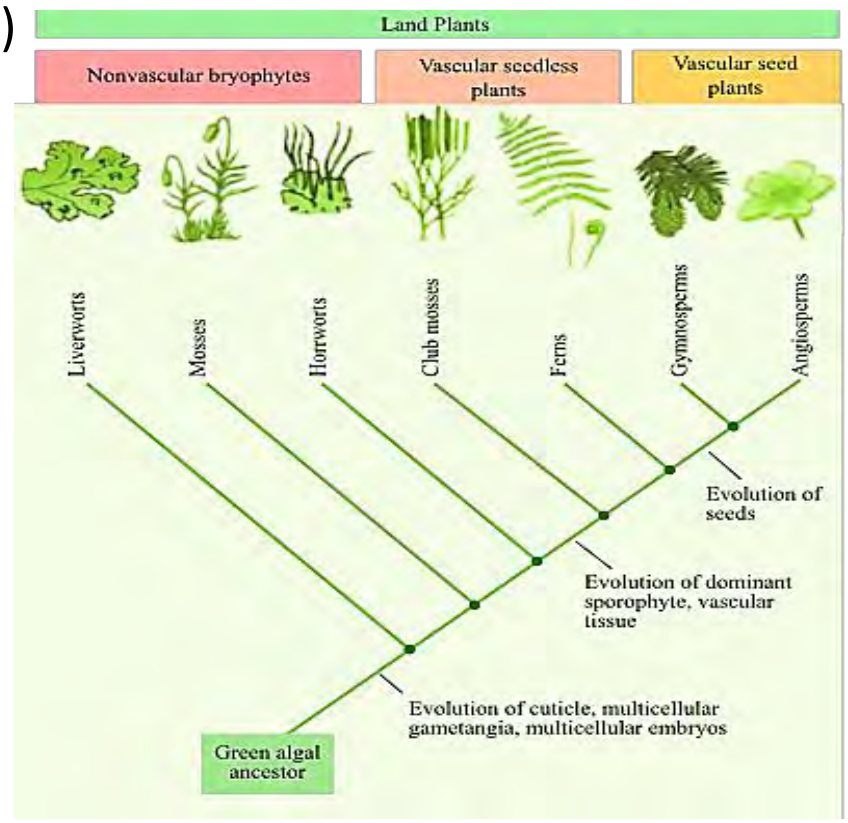
2.Ferns – Example: *Adiantum capillus-veneris* (Maidenhair fern)

3.Gymnosperms – Example: *Ephedra alata* (Joint pine)

4.Angiosperms (Monocots) – Example: *Phoenix dactylifera* (Date palm)

5.Angiosperms (Dicots – Shrubs) – Example: *Ziziphus spina-christi* (Christ's thorn jujube)

6.Angiosperms (Dicots – Herbs) – Example: *Calotropis procera* (Sodom apple)

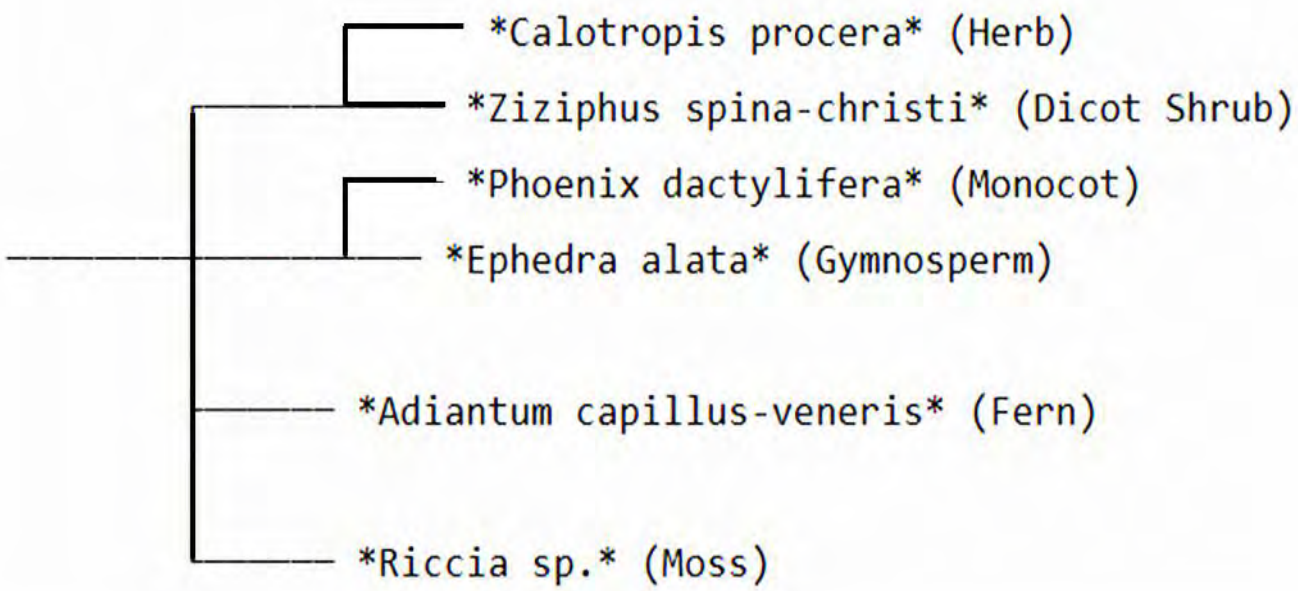


2. Evolutionary Traits for Cladogram

Plant Species	Vascular Tissue	Seeds	Flowers	Fruit	Woody Stem
Riccia sp. (Moss)	No	No	No	No	No
Adiantum capillus-veneris (Fern)	Yes	No	No	No	No
Ephedra alata (Gymnosperm)	Yes	Yes	No	No	Yes
Phoenix dactylifera (Date Palm – Monocot)	Yes	Yes	Yes	Yes	Yes
Ziziphus spina-christi (Jujube – Dicot)	Yes	Yes	Yes	Yes	Yes
Calotropis procera (Sodom Apple – Herb)	Yes	Yes	Yes	Yes	No

3. Cladogram Representation

1. **Bryophytes** (Mosses) are the most primitive.
2. **Ferns** evolved vascular tissue.
3. **Gymnosperms** evolved seeds.
4. **Angiosperms (Monocots & Dicots)** evolved flowers and fruits.
5. **Woody plants** are distinct from **herbs**.



Key Uses of Cladistics

Biological & Evolutionary Studies

Helps reconstruct evolutionary trees and understand species evolution.

Identifies common ancestors and evolutionary traits.

Taxonomy & Classification

More accurate than traditional classification (Linnaean system).

Helps in revising species grouping when new genetic or fossil evidence emerges.

Biodiversity & Conservation

Identifies evolutionarily unique species for conservation.

Tracks biodiversity loss due to environmental changes.

Medicine & Drug Discovery

Traces disease origins (e.g., viruses, bacteria evolution).

Helps find medicinal plants with useful chemical compounds.

Agriculture & Crop Improvement

Aids in breeding disease-resistant & drought-tolerant crops.

Helps develop eco-friendly pest control strategies.

Paleontology & Fossil Studies

Reconstructs extinct species and their relationships.

Studies mass extinctions and evolutionary survival patterns.

How to interpret molecular phylogenetic tree

1. Identify Tree Components

Branches: Represent evolutionary lineages.

Nodes: Indicate common ancestors.

- **Internal Nodes:** Represent hypothetical common ancestors.
- **External (Tip) Nodes:** Represent present-day species or sequences.

Root: The base of the tree, representing the most ancient common ancestor.

Branch Lengths: Sometimes proportional to genetic changes or evolutionary time.

2. Determine the Tree Type

Rooted Tree: Shows a clear evolutionary direction from an ancestor.

Unrooted Tree: Displays relationships without specifying ancestry.

Cladogram: Only shows branching order (relationship).

Phylogram: Shows branch lengths proportional to genetic change.

3. Read Evolutionary Relationships

Close branches indicate recent common ancestry.

Longer branches imply more genetic divergence.

Sister taxa share a more recent common ancestor than with others.

Outgroup (if present) is a reference point, often distantly related.

4. Evaluate Branch Support

Bootstrapping values (e.g., 70% or higher) indicate reliability of branches.

Bayesian posterior probabilities (closer to 1.0 indicate confidence).

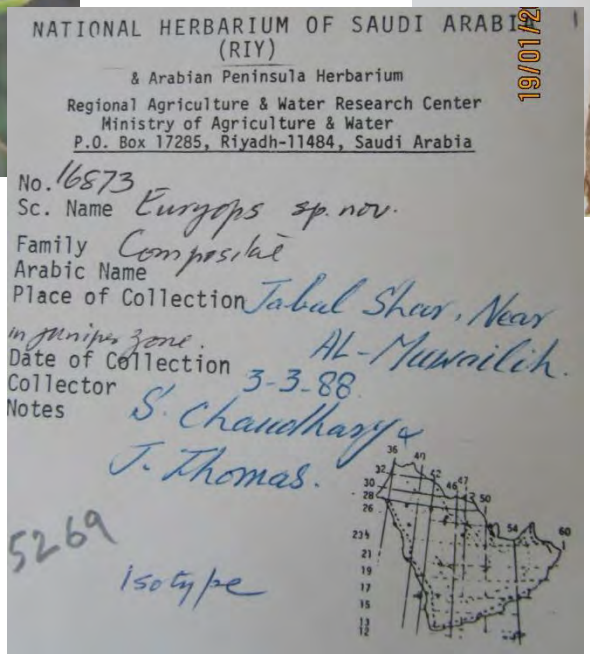
5. Compare with Evolutionary Hypotheses

Look for expected and unexpected relationships.

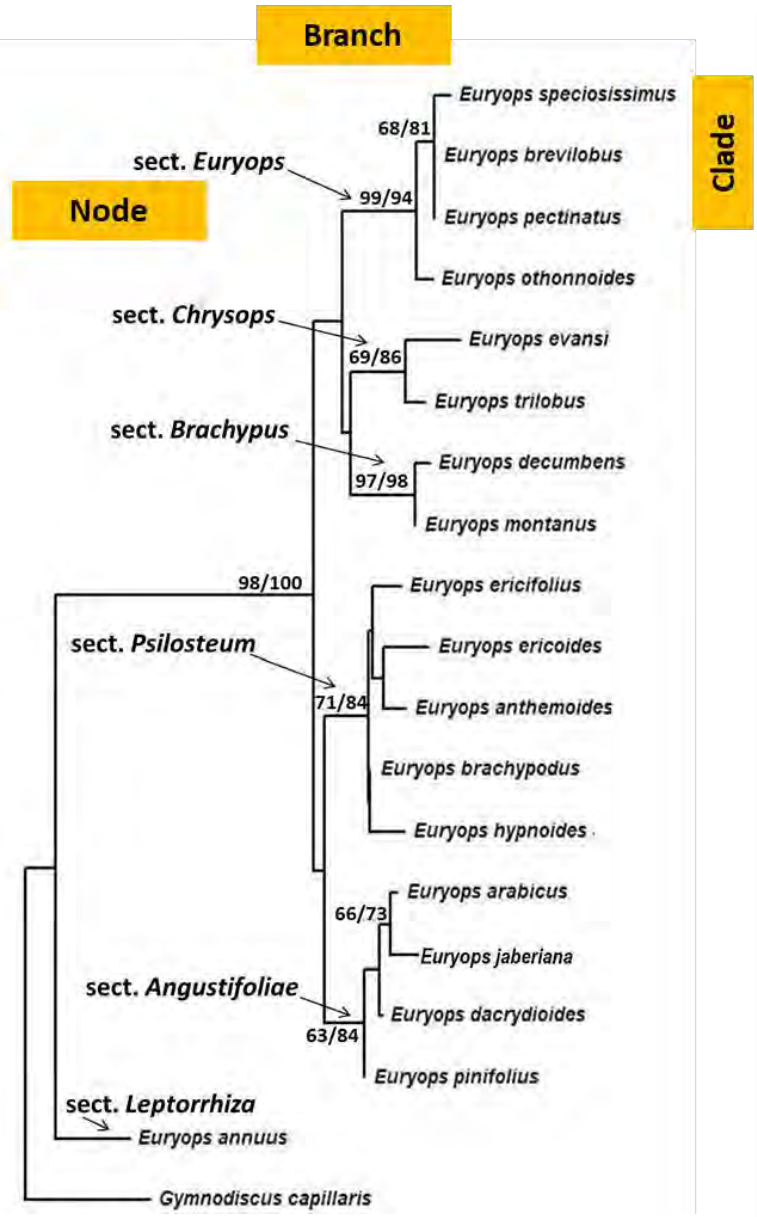
Consider molecular clock assumptions (if applicable).

Phylogenetic Implication of Molecular Genotyping of *Euryops jaberiana* Abedin & Chaudhary (Asteraceae)

- ❖The taxonomic status of *Euryops jaberiana* Abedin & Chaudhary (tribe Senecioneae, was evaluated (Ali et al., 2016) based on molecular phylogenetic analyses of internal transcribed spacer sequence (ITS) of nuclear ribosomal DNA (nrDNA) in order to ascertain its position within the genus.
- ❖In Saudi Arabia, the genus *Euryops* (family Asteraceae) is represented by two species, *viz*. *E. arabicus* Steud. *ex* Jaub. & Spach, and *E. jaberiana* Abedin & Chaudhary.
- ❖*E. arabicus* is is endemic to Arabian Peninsula, while *E. jaberiana* is endemic to northern Saudi Arabia.
- ❖Morphologically *E. jaberiana* very closely resembles with *E. arabicus* / very narrow differences in m morphological characters (Abedin and Chaudhary, 2000).

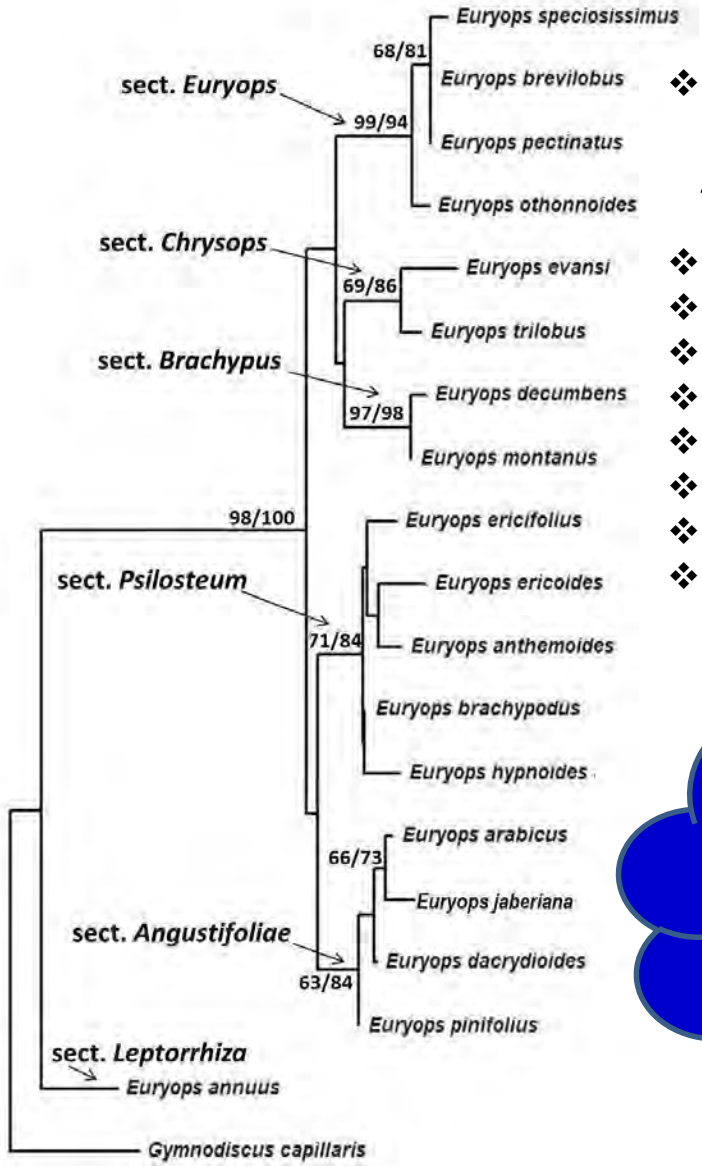


E. arabicus



- ❑ **In molecular taxonomic studies, the most convenient way of presenting taxonomic relationships among a group of organisms is the phylogenetic tree.**
- ❑ **Node: a branch point in a tree**
- ❑ **Branch: defines the relationship between the taxon**
- ❑ **Topology: the branching patterns of the tree**
- ❑ **Branch length: represents the number of changes that have occurred in the branch**
- ❑ **Clade: a group of two or more taxa closed together based on DNA sequences data analysis**
- ❑ **Maximum parsimony is an optimality criterion under which the phylogenetic tree that minimizes the total number of character-state changes is to be preferred.**
- ❑ **Bootstrap: Bootstrapping is a procedure where DNA sequence data run for the phylogenetic analysis, and the reported value is the percentage of bootstrap replicates, for examples 100 means that the node is well-supported, it showed in all tress.**

- ❖ **The key morphological features which differentiate *E. jaberiana* from *E. arabicus* are: leaves 3-lobed at the tips, pappus hairs transparent or rarely dull white, and achenes glabrescent, while in *E. arabicus*, the leaves are unlobed, pappus hairs are dull white and achene densely lanate hairy (Abedin and Chaudhary, 2000).**
- ❖ **The Maximum Parsimony analyses reveals that *E. jaberiana* nested within the clade of the section *Angustifoliae.***
- ❖ ***E. jaberiana* shows proximity with *E. arabicus* (66% bootstrap support.**



- **A total of eight specific nucleotide differences were detected between *E. jaberiana* and *E. Arabicus i.e.* at the alignment position:**
- **93 (A →T)**
- **116 (G → C)**
- **201 (T → C)**
- **443 (C → G)**
- **461 (T → G)**
- **531 (T → C)**
- **573 (C → T)**
- **611 (T → C)**

Thus on the basis of phylogenetic relationships of *E. jaberiana* within the genus and nucleotide differences, Ali et al. (2016) recognized *E. jaberiana* as a distinct species and different from *E. arabicus.*

```
                    10        20        30        40        50        60        70        80        90       100
                    |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Euryops_jabriana    TCGAAACCTGCATAGCAGAACGACCCGTGAACATGTAATAACAATCGGGTGTCCATGGTTTCCGACTATTGTTTGATTCTTTGGATACCCTGATAATGTG
Euryops_arabicus    ....................................................................................................T.......
Clustal Consensus   ********************************************************************************** *******

                    110       120       130       140       150       160       170       180       190       200
                    |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Euryops_jabriana    CGTCTTTGGTCAGCCGCTTGGGTCCTAATGATGTCACATTAACACAATAACAAACCCCCGGCACGGCATGTGCCAAGGAAAATAAAACTTAAGAAAAGCT
Euryops_arabicus    ..............C.....................................................................................
Clustal Consensus   *************** ********************************************************************

                    210       220       230       240       250       260       270       280       290       300
                    |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Euryops_jabriana    TGTATCATGTTACGTCGTTCGCGGGGTTTGCATGATACGTGGCTTCTTTATAATCATAAACGACTCTCGGCAACGGATATCTCGGCTCACGCATCGATGA
Euryops_arabicus    C.....................................................................................................
Clustal Consensus    ********************************************************************************************

                    310       320       330       340       350       360       370       380       390       400
                    |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Euryops_jabriana    AGAACGTAGCAAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCTTTTGGCCGAGGGCA
Euryops_arabicus    ....................................................................................................
Clustal Consensus   ****************************************************************************************************

                    410       420       430       440       450       460       470       480       490       500
                    |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Euryops_jabriana    CGTCTGCCTGGGCGTCACACATCGCGTCGCCCCCCACAACATCTCTTGATTGGGATGTTGTAATGGGGGCGGATATTGGTCTCCCGTTCCTAAGGTTCGG
Euryops_arabicus    ..........................................G.................G.....................................
Clustal Consensus   ****************************************** ***************** ***************************************

                    510       520       530       540       550       560       570       580       590       600
                    |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Euryops_jabriana    TTGGCTAAAATAGGAGTCCCCTTCGAAGGATGCACGATTAGTGGTGGTTGTCAAGACCTTCTTATCGACTCGCGCGTTACAAGTAGTAGGGAAGATCTCT
Euryops_arabicus    ............................C.....................................................T.....................
Clustal Consensus   **************************** ****************************************** ***********************

                    610       620       630       640
                    |....|....|....|....|....|....|....|
Euryops_jabriana    TCAAAGACCCTAATGTGTTGTCTTGTGACAATGCTTCGACCGCGA
Euryops_arabicus    ..........C.....................................
Clustal Consensus   ********** ************************************
```

PRACTICAL

Lab Activity No 01
Title of the Activity- DNA Extraction from plant tissue

Learning Objectives:

- Understand the process of extracting DNA from plant tissue.
- Learn about the role of different reagents in DNA extraction.
- Observe the physical appearance of DNA.

Aim:

To extract and observe DNA from plant tissue using a simple extraction method.

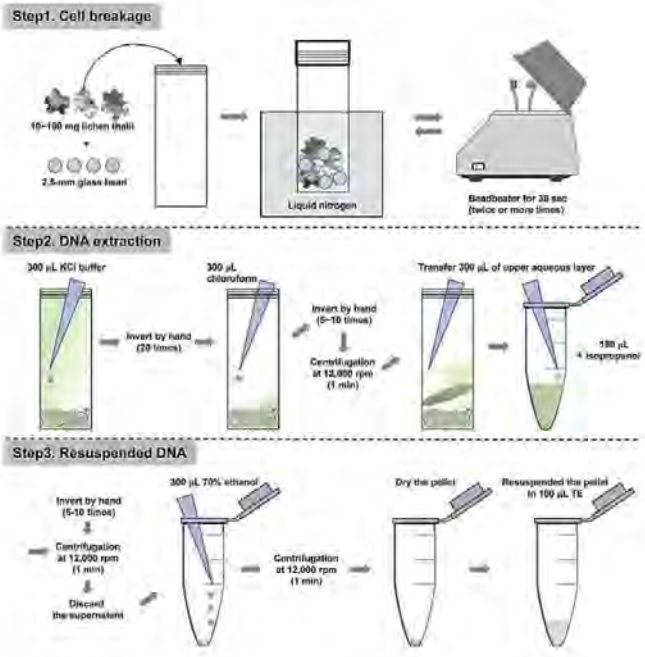
Apparatus & Materials:

- Fresh plant tissue (e.g., banana, spinach, or strawberry)
- Mortar and pestle
- 10% dishwashing liquid or shampoo (detergent solution)
- 5% salt solution (NaCl)
- Ice-cold ethanol or isopropanol
- Distilled water
- Beaker (100 mL)
- Glass rod or toothpick
- Filter paper and funnel
- Test tube

Theory:

DNA (deoxyribonucleic acid) is a genetic material present in all living cells. It can be extracted by breaking open the cells, dissolving the cell membranes, and precipitating the DNA using alcohol. The detergent helps in breaking down the lipid membranes, while the salt stabilizes the DNA molecules. Alcohol is used to precipitate the DNA as it is insoluble in ethanol.

Diagram:



Procedure:

1. Take fresh plant tissue (about 10 g) and grind it with a small amount of saltwater using a mortar and pestle to break open the cells.
2. Transfer the crushed tissue to a beaker and add 10 mL of detergent solution. Stir gently for 5–10 minutes to break the cell membranes.
3. Filter the mixture using filter paper and a funnel to remove solid debris. Collect the filtrate in a test tube.
4. Slowly add an equal volume of ice-cold ethanol along the sides of the test tube without mixing.
5. Allow the tube to stand for a few minutes. DNA will appear as white, stringy precipitate at the interface of the alcohol and filtrate.
6. Use a glass rod or toothpick to spool out and observe the DNA.

Observations Table:

Step	Observation
Grinding plant tissue	Mixture becomes thick and greenish
Adding detergent	Solution becomes frothy and slightly clearer
Filtering	Clear liquid is obtained
Adding ethanol	White, thread-like DNA appears

Result:
DNA was successfully extracted from plant tissue and appeared as a white, stringy precipitate.

Conclusion:
The experiment demonstrates that DNA can be isolated from plant cells using simple chemical treatments. The detergent breaks cell membranes, salt stabilizes DNA, and alcohol helps in precipitation.

Precautions:

- Use fresh plant tissue for better results.
- Do not vigorously shake after adding ethanol to avoid breaking DNA strands.
- Use ice-cold ethanol for better precipitation.
- Handle chemicals carefully to avoid spillage.

Lab Activity No 02
Title of the Activity- PCR Primer and Polymerase Chain Reaction (PCR)

Learning Objectives:

- Understand the principles of Polymerase Chain Reaction (PCR).
- Learn the function and design of primers in PCR.
- Perform PCR and analyze the amplification of DNA.

Aim:
To amplify a specific DNA sequence using Polymerase Chain Reaction (PCR).

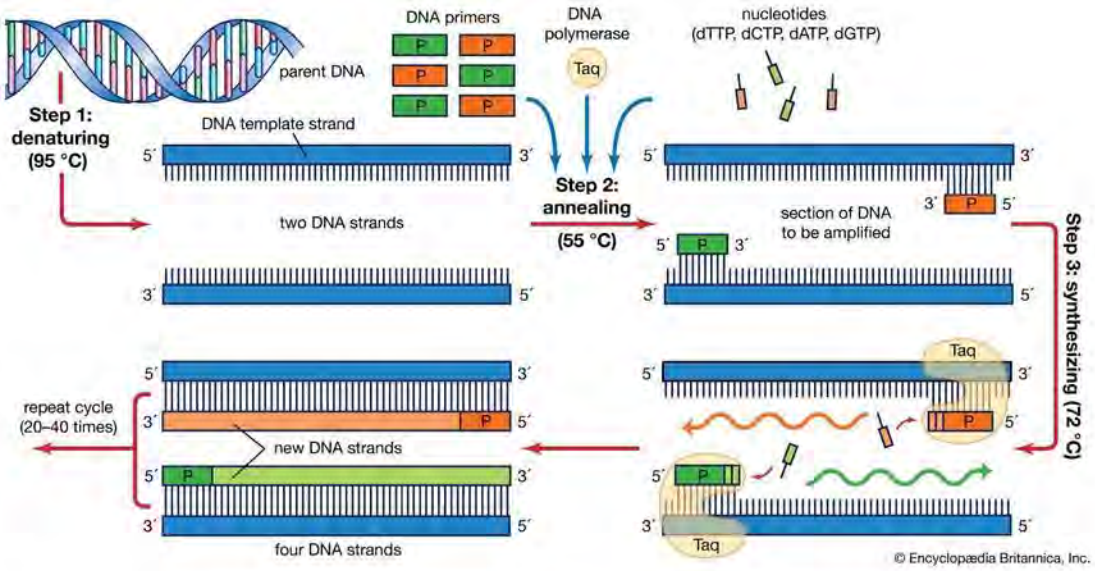
Apparatus & Materials:

- DNA template (target DNA sample)
- Forward and reverse primers

- Taq DNA polymerase
- dNTPs (deoxynucleotide triphosphates)
- PCR buffer with Mg^{2+} ions
- PCR tubes
- Thermal cycler (PCR machine)
- Micropipettes and tips
- Agarose gel and electrophoresis apparatus (for analysis)
- UV transilluminator

Theory:
Polymerase Chain Reaction (PCR) is a molecular biology technique used to amplify a specific DNA sequence. It involves repeated cycles of denaturation, annealing, and extension. Primers are short DNA sequences that define the region to be amplified. Taq polymerase synthesizes new DNA strands using the template. The exponential amplification of DNA makes PCR an essential tool in genetics, forensic science, medical diagnostics, and research.

Diagram:



Procedure:

1. **Preparation of PCR Master Mix:**
   - In a PCR tube, mix the following components:
     - 10–50 ng of template DNA
     - 0.5–1 µM of forward and reverse primers
     - 200 µM of dNTPs
     - 1X PCR buffer (with Mg^{2+})
     - 1–2 U of Taq DNA polymerase
     - Adjust the final volume with nuclease-free water.
2. **Placing the Tube in the Thermal Cycler:**
   - Set up the PCR conditions as follows:

Step	Temperature	Time	Function
Initial Denaturation	94-98°C	2-5 min	Breaks double-stranded DNA into single strands

Denaturation	94-98°C	30 sec	Separates DNA strands
Annealing	50-65°C	30 sec	Primers bind to complementary sequences
Extension	72°C	30-60 sec	DNA polymerase extends the new strand
Final Extension	72°C	5-10 min	Ensures complete DNA synthesis
Hold	4°C	Indefinite	Stores the PCR product

- Repeat **denaturation, annealing, and extension** steps for **25-40 cycles**.

3. **Agarose Gel Electrophoresis (for visualization):**
    - Prepare a **1-2% agarose gel** and load the PCR product.
    - Run electrophoresis at **100V for 30 minutes**.
    - Visualize DNA bands under a **UV transilluminator**.

Observations Table:

Step	Observation
Sample loaded in PCR machine	No visible change
Post-PCR (before gel electrophoresis)	Clear liquid in the tube
After gel electrophoresis	DNA bands visible under UV light

Result:
A distinct DNA band corresponding to the amplified target sequence is observed on the agarose gel, confirming successful PCR amplification.

Conclusion:
The experiment successfully demonstrates the amplification of a specific DNA sequence using PCR. The results confirm the efficiency of primers in targeting the desired DNA region, and the process can be applied in various molecular biology studies.

Precautions:
- Use **sterile** micropipette tips and PCR tubes to prevent contamination.
- Keep **Taq polymerase and dNTPs on ice** to maintain enzyme stability.
- Accurately set the **thermal cycler parameters** for optimal amplification.
- Use **ice-cold ethanol or isopropanol** for DNA precipitation if required.
- Properly handle the UV transilluminator to avoid exposure.

Lab Activity No 03
Title of the Activity- Gel Electrophoresis for DNA Quality

Learning Objectives:
- Understand the principle of gel electrophoresis and its role in DNA analysis.
- Learn how to assess DNA quality based on band patterns.
- Gain hands-on experience in preparing and running an agarose gel electrophoresis experiment.

Aim:
To analyze the quality of DNA using agarose gel electrophoresis by observing the integrity, purity, and size of DNA fragments.

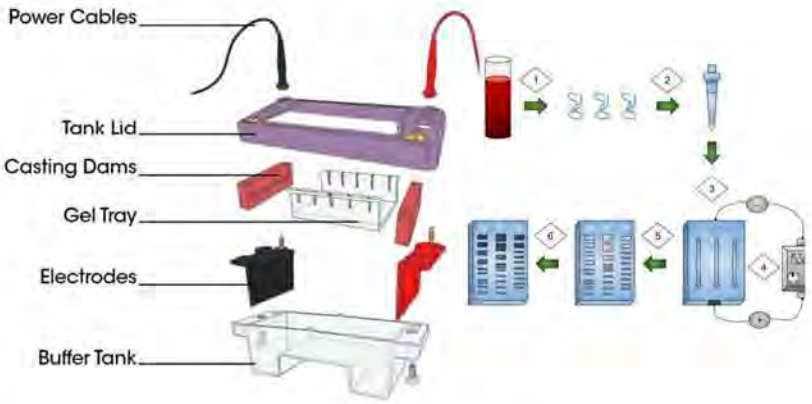
Apparatus & Materials:
- DNA sample (extracted or PCR product)
- Agarose powder
- 1X TAE (Tris-Acetate-EDTA) or TBE (Tris-Borate-EDTA) buffer
- Ethidium bromide (EtBr) or SYBR Green (for DNA staining)

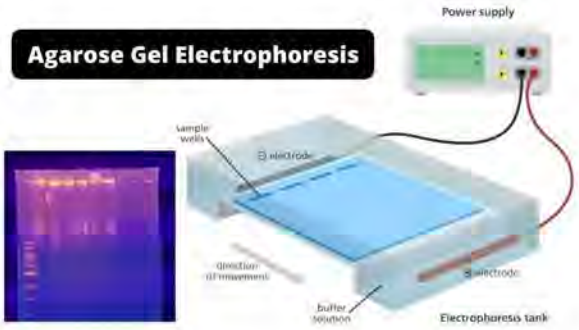
- Loading dye (e.g., bromophenol blue)
- DNA ladder (molecular weight marker)
- Gel casting tray and comb
- Electrophoresis chamber and power supply
- Micropipettes and tips
- UV transilluminator or gel documentation system

Theory:
Gel electrophoresis is a technique used to separate and analyze DNA fragments based on their size. DNA molecules are negatively charged due to their phosphate backbone and migrate towards the positive electrode when an electric field is applied.

- **Agarose Gel Concentration:** Affects resolution; lower % gels (0.8-1%) separate larger fragments, while higher % gels (2-3%) separate smaller fragments.
- **DNA Migration:** Smaller DNA fragments move faster and farther, while larger fragments move slower.
- **Staining:** DNA is visualized under UV light after binding with EtBr or SYBR Green.
- **DNA Quality Indicators:**
  - Intact DNA appears as a high-molecular-weight band.
  - Degraded DNA appears as a smear.
  - Contaminated DNA may show additional unexpected bands.

Diagram:





Procedure:

1. **Prepare the Gel:**

- Dissolve **0.8-2% agarose** in **1X TAE/TBE buffer** by heating.
- Cool slightly, add **EtBr/SYBR Green**, and pour into a gel tray with a comb.
- Allow the gel to solidify (20-30 minutes).

2. **Prepare DNA Samples:**
   - Mix **5 µL of DNA sample** with **1 µL of loading dye**.
   - Prepare a **DNA ladder** in a similar way.
3. **Load the Gel:**
   - Place the solidified gel in the **electrophoresis chamber** and cover it with **1X TAE/TBE buffer**.
   - Load the **DNA samples and ladder** into separate wells using a micropipette.
4. **Run the Electrophoresis:**
   - Connect the electrodes (negative at the well side, positive at the opposite end).
   - Set voltage to **80-120V** and run for **30-45 minutes** (depending on gel size).
5. **Visualize DNA:**
   - Place the gel on a **UV transilluminator** or **gel documentation system**.
   - Capture an image to analyze the bands.

Observations Table:

Sample	Observation under UV Light	Interpretation
High-quality DNA	Single, sharp, high-molecular-weight band	Good integrity
Degraded DNA	Smear instead of distinct bands	DNA degradation
Contaminated DNA	Additional unexpected bands	RNA/protein contamination
DNA ladder	Clear bands at expected sizes	Gel running properly

Result:
Based on the gel electrophoresis results, the DNA quality was assessed. High-quality DNA appeared as intact bands, while degraded or contaminated DNA showed smearing or extra bands.

Conclusion:
Gel electrophoresis is an effective method for evaluating DNA quality. The integrity and purity of DNA can be determined by analyzing band patterns. Proper DNA extraction and storage methods help in obtaining high-quality DNA for further experiments.

Precautions:
- Use gloves and goggles when handling **EtBr** or **SYBR Green** (carcinogenic substances).
- Ensure the gel is properly set before loading samples to prevent leakage.
- Load samples carefully to **avoid cross-contamination** between wells.
- Do not run electrophoresis at excessively high voltage, as it may cause **DNA degradation**.
- Dispose of used gels and chemicals **as per lab safety protocols**.

Lab Activity No 04
Title of the Activity- Sanger Sequencing

Learning Objectives:

- Understand the principle of Sanger sequencing and its role in DNA sequencing.
- Learn how chain termination using dideoxynucleotides (ddNTPs) helps determine DNA sequences.
- Gain knowledge of the components, procedure, and interpretation of Sanger sequencing results.

Aim:

To determine the nucleotide sequence of a given DNA sample using the Sanger sequencing method.

Apparatus & Materials:

- **DNA Template:** The single-stranded DNA to be sequenced.
- **DNA Polymerase:** Taq or another thermostable enzyme for DNA synthesis.
- **Primers:** Short oligonucleotide sequences that initiate DNA synthesis.
- **dNTPs (Deoxynucleotide Triphosphates):** Standard nucleotides (A, T, G, C) for chain elongation.
- **ddNTPs (Dideoxynucleotide Triphosphates):** Modified nucleotides that terminate DNA synthesis.
- **PCR Tubes:** To set up the sequencing reactions.
- **Thermal Cycler:** For amplification and chain termination reactions.
- **Capillary Electrophoresis System or Polyacrylamide Gel:** For fragment separation.
- **Fluorescent Labeling Dye or Radioactive Labeling:** For visualization of the DNA sequence.
- **Computational Software:** For automated sequence analysis.

Theory:

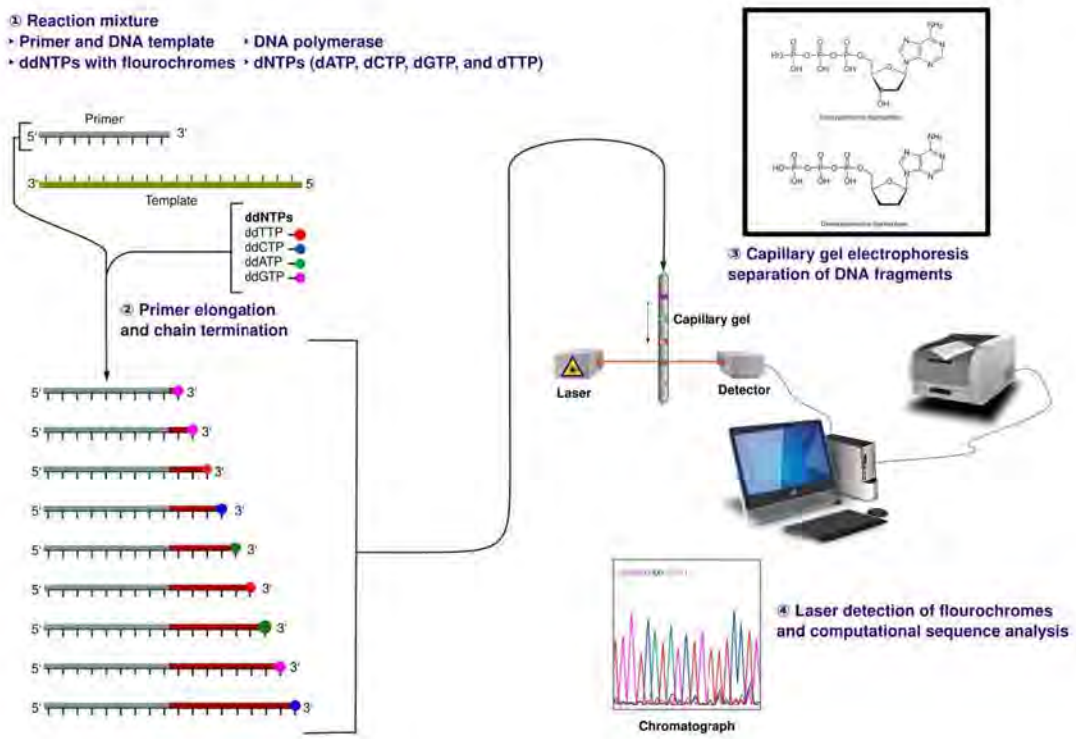
Sanger sequencing, also called the **chain termination method**, is a DNA sequencing technique developed by Frederick Sanger. The process involves DNA synthesis using a polymerase enzyme in the presence of both **normal dNTPs** (which extend the DNA strand) and **ddNTPs** (which cause termination). Since ddNTPs lack a 3'-OH group, their incorporation results in **premature termination** of strand elongation.

By running separate reactions for each nucleotide (A, T, G, C) or using **fluorescently labeled ddNTPs**, the resulting DNA fragments of different lengths can be separated via **capillary electrophoresis or gel electrophoresis** to determine the sequence.

Diagram:

Procedure:

1. **Reaction Setup:**
    - Prepare four separate reaction tubes, each containing:
        - Template DNA
        - DNA polymerase
        - Primers
        - dNTPs
        - A small amount of one of the four **ddNTPs** (ddATP, ddTTP, ddGTP, or ddCTP)



2. **DNA Amplification via PCR:**
   - Run a thermal cycling program:
     - **Denaturation (95°C, 30 sec):** DNA strands separate.
     - **Annealing (50-60°C, 30 sec):** Primers bind to the template.
     - **Extension (72°C, 1 min):** DNA polymerase incorporates nucleotides until a **ddNTP** is added, terminating the strand.
     - Repeat for 25-30 cycles.
3. **Fragment Separation:**
   - Load the amplified DNA into a **capillary electrophoresis system** or **polyacrylamide gel electrophoresis** to separate fragments by size.
4. **Detection and Analysis:**
   - If using **radioactive or fluorescently labeled ddNTPs**, detect fragment sizes using **autoradiography** (for radioactive sequencing) or **laser detection in a sequencer** (for fluorescent sequencing).
   - The banding pattern in electrophoresis corresponds to the **complementary DNA sequence**.

Observations Table:

Step	Observation	Interpretation
DNA denaturation	Single-stranded DNA formation	Ready for primer binding
Primer annealing	Primers attach to the template	DNA polymerase can initiate synthesis
DNA extension with dNTPs & ddNTPs	DNA fragments of varying lengths	Chain termination occurring
Electrophoresis result	DNA bands or peaks of	DNA sequence can be read

	different sizes	

Result:
The sequence of the given DNA template is determined based on the band pattern or **fluorescent peak readings** obtained after electrophoresis. The **shortest fragments represent the start of the sequence**, and the **longest fragments represent the end**.

Conclusion:
Sanger sequencing is a reliable method for determining DNA sequences. It uses **ddNTP-mediated termination**, and by separating the fragments based on size, the **order of nucleotides in the DNA strand can be deduced**. The method is widely used in **genetic analysis, mutation detection, and genome sequencing**.

Precautions:
- Ensure **accurate pipetting** of dNTPs and ddNTPs to prevent sequencing errors.
- Maintain **sterile conditions** to prevent contamination.
- Use **high-purity DNA** for accurate sequencing results.
- Optimize **PCR cycling conditions** to ensure proper amplification.
- Handle **fluorescent dyes and radioactive materials** with proper safety measures.

Lab Activity No 05
Title of the Activity- Retrieval of DNA sequence data from NCBI

Learning Objectives:
- Understand how to access and retrieve DNA sequence data from the **National Center for Biotechnology Information (NCBI)** database.
- Learn the use of **GenBank, FASTA format, and BLAST (Basic Local Alignment Search Tool)** for sequence analysis.
- Gain experience in navigating the **NCBI website** for DNA sequence retrieval.

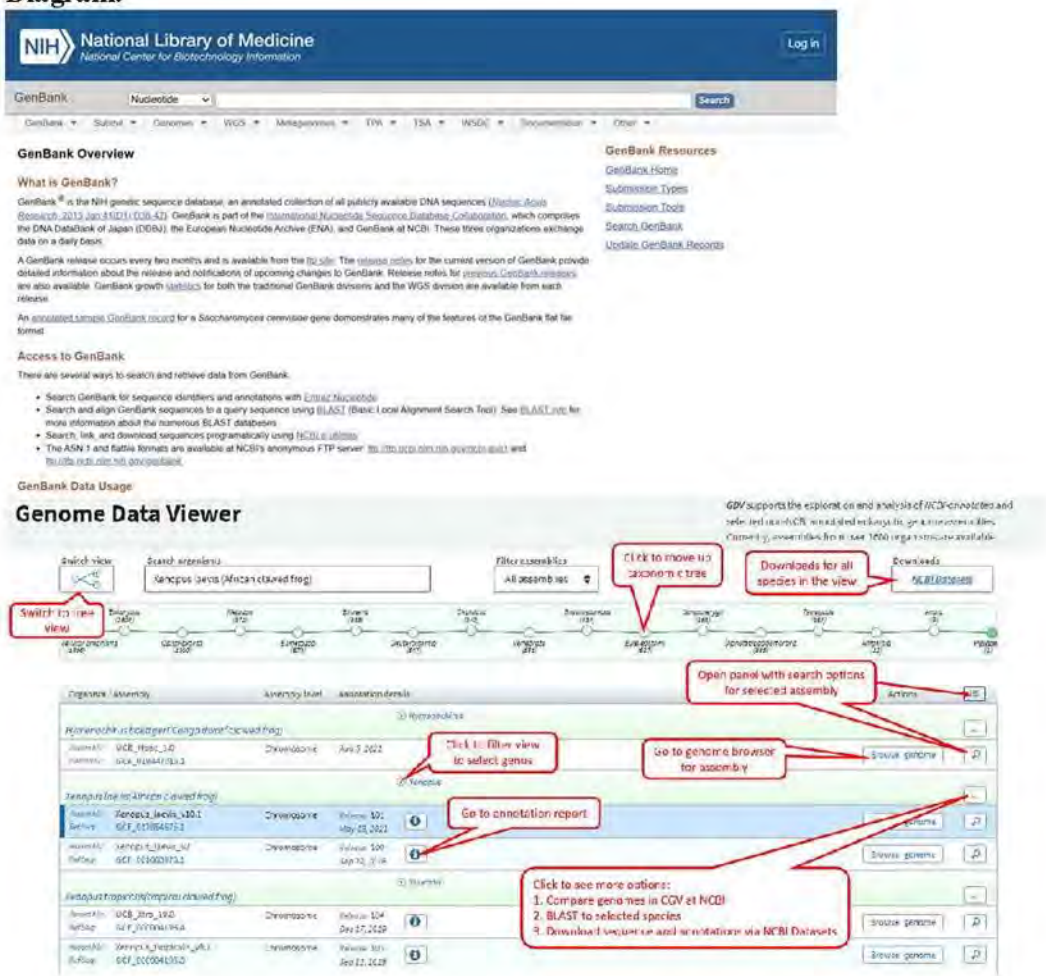
Aim:
To retrieve and analyze a DNA sequence from the **NCBI GenBank database** using an accession number or a keyword search.

Apparatus & Materials:
- **Computer with Internet access**
- **NCBI website (https://www.ncbi.nlm.nih.gov/)**
- **Web browser (Google Chrome, Mozilla Firefox, etc.)**
- **Accession number or search term**
- **Text editor (Notepad, MS Word, or BioEdit for FASTA files)**

Theory:
The **National Center for Biotechnology Information (NCBI)** provides a public database containing DNA, RNA, and protein sequences. The key features include:
1. **GenBank Database:** A comprehensive collection of publicly available genetic sequences.
2. **FASTA Format:** A standard format for representing DNA and protein sequences.
3. **Accession Number:** A unique identifier for each sequence entry in GenBank.
4. **BLAST (Basic Local Alignment Search Tool):** A tool to compare a query sequence with known sequences in the database.

NCBI allows users to **search, retrieve, and analyze** DNA sequences, making it a vital resource for bioinformatics research.

Diagram:



Procedure:

Step 1: Access NCBI Website

- Open a web browser and go to NCBI GenBank.

Step 2: Search for a DNA Sequence

- In the search bar, enter a **gene name, organism name, or an accession number** (e.g., **NM_001301717.2** for the human BRCA1 gene).
- Click **Search** to retrieve matching records.

Step 3: Select a Sequence

- Click on the desired entry from the search results.
- The sequence information, including **source organism, gene location, coding region, and references**, will be displayed.

Step 4: Retrieve the DNA Sequence

- Scroll down to find the **FASTA format** option.
- Click **"FASTA"** to display the sequence in a text format.
- Copy the sequence for further analysis.

Step 5: Download the Sequence File

- Click **"Send to" → "File" → Select FASTA format → Download**.

Step 6: Analyze the Sequence using BLAST (Optional)

- Go to the **BLAST tool** (NCBI BLAST).
- Paste the copied sequence into the input box.
- Click **"BLAST"** to find similar sequences in the database.

Observations Table:

Step	Observation	Interpretation
Search results	Multiple sequence entries appear	NCBI contains many related sequences
Selection of entry	Detailed sequence and metadata displayed	Correct sequence identified
FASTA sequence retrieved	A text file containing A, T, G, C bases	DNA sequence successfully extracted
BLAST result (if used)	Similar sequences with % identity appear	Sequence similarity analysis completed

Result:

A DNA sequence was successfully retrieved from **NCBI GenBank**, displayed in **FASTA format**, and stored for further analysis.

Conclusion:

The NCBI database is an essential tool for retrieving and analyzing DNA sequences. By using search queries and accession numbers, researchers can access genetic information for various applications in **bioinformatics, evolutionary studies, and genetic research**.

Precautions:

- Ensure correct spelling and scientific names while searching.
- Use **official accession numbers** for accurate sequence retrieval.
- Check the **date and version** of the sequence to get the latest update.
- Verify the source organism before using the data in further analysis.
- Avoid **editing** the retrieved sequence accidentally before saving.

Lab Activity No 06
Title of the Activity- Multiple Sequence Alignment (MSA) Using ClustalX

Learning Objectives:

- Understand the principle and importance of Multiple Sequence Alignment (MSA).
- Learn how to perform MSA using **ClustalX**, a widely used bioinformatics tool.
- Interpret sequence alignment results for evolutionary and functional analysis.

.

Aim:

To perform **Multiple Sequence Alignment (MSA)** of DNA or protein sequences using **ClustalX** and analyze sequence conservation and evolutionary relationships.

Apparatus & Materials:

- **Computer with Windows/Linux/Mac OS**
- **ClustalX software (Download from: EBI Website)**
- **FASTA format sequence files**
- **Internet access (optional, for database searches and downloads)**

Theory:

Multiple Sequence Alignment (MSA) is a method used to align three or more biological sequences (DNA, RNA, or proteins) to identify regions of similarity. These similarities may indicate **evolutionary relationships, functional regions, or conserved motifs**.

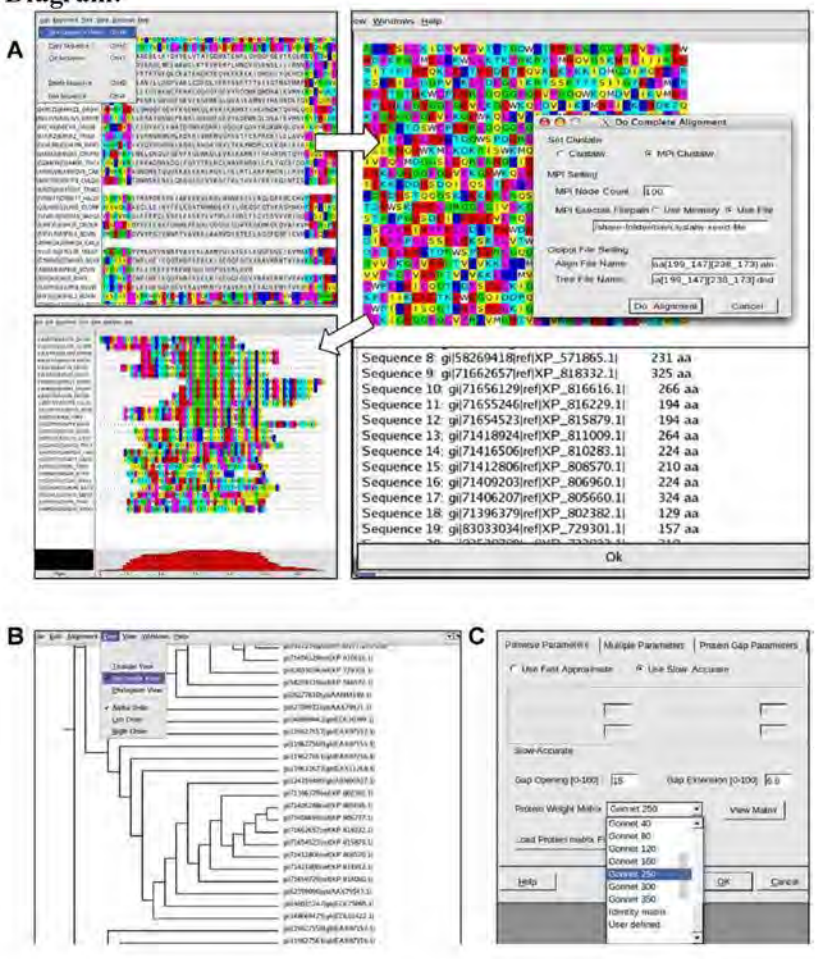
ClustalX is a graphical version of the Clustal algorithm that performs MSA using:

1. **Progressive Alignment Method** – Sequences are aligned stepwise based on similarity scores.
2. **Scoring Matrices** – Used to evaluate alignment quality (e.g., BLOSUM for proteins, PAM for DNA).
3. **Phylogenetic Tree Construction** – Helps understand evolutionary relationships.

Applications of MSA:

- Identifying **conserved domains** in genes and proteins.
- Constructing **phylogenetic trees** for evolutionary studies.
- Comparing **functional sites** in homologous sequences.

Diagram:



Procedure:

Step 1: Install ClustalX

- Download **ClustalX** from the EBI website and install it on your system.

Step 2: Prepare the Input Sequences

- Obtain DNA/protein sequences in **FASTA format**.
- Save all sequences in a **single text file** with .fasta or .txt extension.

Step 3: Load Sequences into ClustalX

- Open **ClustalX** and click **File → Load Sequences**.
- Select the prepared **FASTA file** and open it.

Step 4: Perform Multiple Sequence Alignment

- Click **Alignment → Do Complete Alignment**.
- Adjust parameters if needed (gap penalties, scoring matrix, etc.).
- Wait for the software to process the alignment.

Step 5: View and Analyze the Alignment

- The aligned sequences will be displayed with **conserved regions marked**.
- Save the results by clicking **File → Save Alignment As**.

Step 6: Construct a Phylogenetic Tree (Optional)

- Click **Trees → Draw Tree** to generate a phylogenetic tree.
- Save the tree as an image or text file for analysis.

Observations Table:

Step	Observation	Interpretation
Loading sequences	All sequences appear in ClustalX	Proper sequence input
Alignment process	Aligned sequences with gaps appear	Successful MSA
Conserved regions	Identified by asterisks (*) in the alignment	Highly conserved sequences
Phylogenetic tree	Dendrogram/tree displayed	Evolutionary relationships inferred

Result:

The given DNA/protein sequences were successfully aligned using **ClustalX**, and **conserved regions** were identified. The **phylogenetic tree** showed relationships between sequences.

Conclusion:

ClustalX is an effective tool for performing **Multiple Sequence Alignment (MSA)**, revealing conserved regions and evolutionary relationships among DNA/protein sequences. This alignment can be used for **functional analysis, comparative genomics, and phylogenetics**.

Precautions:

- Ensure **FASTA format** is correctly maintained to avoid errors.
- Choose the **appropriate scoring matrix** (BLOSUM for proteins, PAM for DNA).
- Set **gap penalties correctly** to prevent misalignment.
- Verify **sequence quality** before running MSA.
- Save results **frequently** to prevent data loss.

Lab Activity No 07
Title of the Activity- NJ Phylogenetic Tree Construction Using MEGA

Learning Objectives:

- Understand the **Neighbor-Joining (NJ) method** for constructing phylogenetic trees.

- Learn how to use **MEGA (Molecular Evolutionary Genetics Analysis)** software for tree construction.
- Analyze the evolutionary relationships between different DNA/protein sequences.

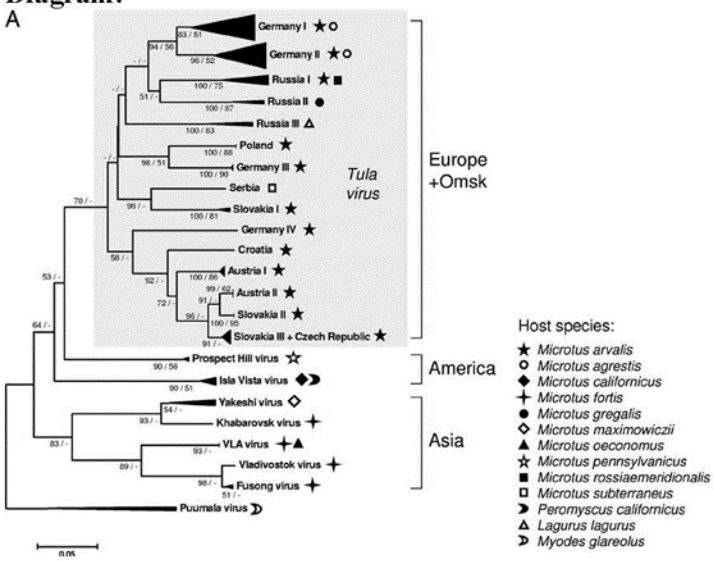
Aim:
To construct a **Neighbor-Joining (NJ) phylogenetic tree** using **MEGA** software to study the evolutionary relationships between sequences.

Apparatus & Materials:
- **Computer with Windows/Linux/Mac OS**
- **MEGA software (Download from: https://www.megasoftware.net/)**
- **FASTA format sequence file**
- **Internet access (optional, for sequence retrieval from NCBI)**

Theory:
A **phylogenetic tree** represents the evolutionary relationships among species or genes. The **Neighbor-Joining (NJ) method**, developed by Saitou and Nei (1987), is a widely used algorithm for tree construction.
Principle of Neighbor-Joining (NJ) Method:
1. **Calculate a distance matrix** based on sequence similarity.
2. **Iteratively join the two closest taxa** (sequences) into a common ancestor.
3. **Reduce the distance matrix** and repeat until a single tree is formed.

Applications of NJ Trees:
- Understanding **evolutionary relationships** among species.
- Identifying **common ancestors** and divergence points.
- Analyzing **genetic similarities** in different populations.

Diagram:
A



Procedure:
Step 1: Install and Open MEGA Software
- Download and install **MEGA** from the official website.
- Open the software and select **Phylogeny → Construct/Analyze Phylogenetic Tree**.

Step 2: Prepare the Input Sequences

- Obtain DNA or protein sequences in **FASTA format**.
- Open **MEGA → Align → Edit/Build Alignment** to import sequences.
- Save the alignment as a **MEGA (.meg) file**.

Step 3: Create a Distance Matrix

- Click **Phylogeny → Compute Pairwise Distance**.
- Choose an appropriate substitution model (e.g., **Kimura 2-parameter** for DNA).

Step 4: Construct the NJ Phylogenetic Tree

- Go to **Phylogeny → Construct NJ Tree**.
- Select the **NJ method** and choose the substitution model.
- Click **OK** to generate the tree.

Step 5: View and Interpret the Tree

- The phylogenetic tree will be displayed with **branch lengths and bootstrap values**.
- Save the tree by clicking **File → Export Tree as Image**.

Observations Table:

Step	Observation	Interpretation
Sequence alignment	Aligned sequences displayed in MEGA	Ready for tree construction
Distance matrix calculation	Pairwise distances generated	Sequence relationships quantified
NJ tree construction	Tree structure appears with branches	Evolutionary connections established
Bootstrap values	Support values displayed on branches	Confidence level of relationships

Result:

A **Neighbor-Joining (NJ) phylogenetic tree** was successfully constructed using **MEGA software**, revealing the **evolutionary relationships** among the input sequences.

Conclusion:

The NJ method in MEGA efficiently constructs **phylogenetic trees** based on sequence similarity. It is widely used in **molecular evolution, comparative genomics, and phylogenetics** to study the relationships among species or genes.

Precautions:

- Ensure sequences are **properly aligned** before tree construction.
- Choose the **appropriate substitution model** for accurate distance calculations.
- Use **bootstrap analysis** (≥1000 replications) for statistical reliability.
- Verify sequence quality to avoid misinterpretation.
- Save results and **export the tree image** for documentation.

Lab Activity No 08
Title of the Activity- MP Phylogenetic Tree Construction Using MEGA

Learning Objectives:

- Understand the **Maximum Parsimony (MP) method** for constructing phylogenetic trees.
- Learn how to use **MEGA (Molecular Evolutionary Genetics Analysis) software** for MP tree construction.

- Obtain DNA or protein sequences in **FASTA format**.
- Open **MEGA → Align → Edit/Build Alignment** to import sequences.
- Save the alignment as a **MEGA (.meg) file**.

Step 3: Create a Distance Matrix

- Click **Phylogeny → Compute Pairwise Distance**.
- Choose an appropriate substitution model (e.g., **Kimura 2-parameter** for DNA).

Step 4: Construct the NJ Phylogenetic Tree

- Go to **Phylogeny → Construct NJ Tree**.
- Select the **NJ method** and choose the substitution model.
- Click **OK** to generate the tree.

Step 5: View and Interpret the Tree

- The phylogenetic tree will be displayed with **branch lengths and bootstrap values**.
- Save the tree by clicking **File → Export Tree as Image**.

Observations Table:

Step	Observation	Interpretation
Sequence alignment	Aligned sequences displayed in MEGA	Ready for tree construction
Distance matrix calculation	Pairwise distances generated	Sequence relationships quantified
NJ tree construction	Tree structure appears with branches	Evolutionary connections established
Bootstrap values	Support values displayed on branches	Confidence level of relationships

Result:

A **Neighbor-Joining (NJ) phylogenetic tree** was successfully constructed using **MEGA software**, revealing the **evolutionary relationships** among the input sequences.

Conclusion:

The NJ method in MEGA efficiently constructs **phylogenetic trees** based on sequence similarity. It is widely used in **molecular evolution, comparative genomics, and phylogenetics** to study the relationships among species or genes.

Precautions:

- Ensure sequences are **properly aligned** before tree construction.
- Choose the **appropriate substitution model** for accurate distance calculations.
- Use **bootstrap analysis** (≥1000 replications) for statistical reliability.
- Verify sequence quality to avoid misinterpretation.
- Save results and **export the tree image** for documentation.

Lab Activity No 08
Title of the Activity- MP Phylogenetic Tree Construction Using MEGA

Learning Objectives:

- Understand the **Maximum Parsimony (MP) method** for constructing phylogenetic trees.
- Learn how to use **MEGA (Molecular Evolutionary Genetics Analysis) software** for MP tree construction.

- Analyze **evolutionary relationships** among DNA/protein sequences using the MP approach.

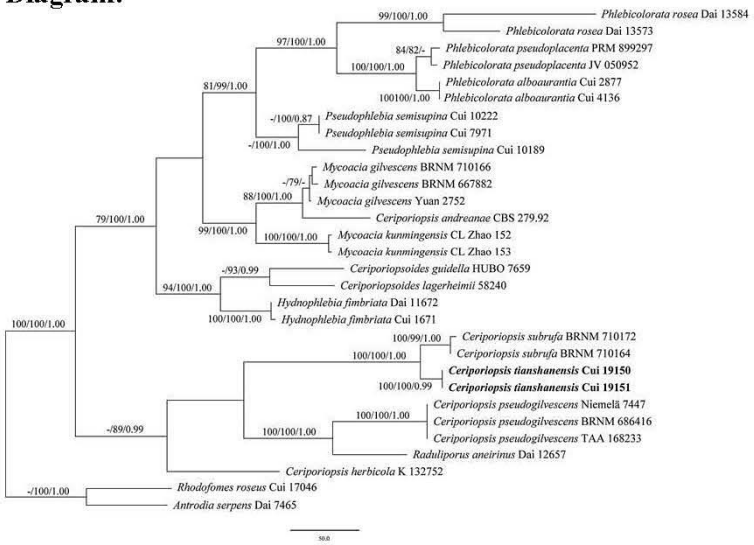
Aim:
To construct a **Maximum Parsimony (MP) phylogenetic tree** using **MEGA software** and analyze sequence relationships.

Apparatus & Materials:
- **Computer with Windows/Linux/Mac OS**
- **MEGA software (Download from: https://www.megasoftware.net/)**
- **FASTA format sequence file**
- **Internet access (optional, for sequence retrieval from NCBI)**

Theory:
A **phylogenetic tree** visually represents evolutionary relationships among species or genes. The **Maximum Parsimony (MP) method** seeks to find the simplest tree (with the fewest evolutionary changes) that explains the given sequence data.
Principles of Maximum Parsimony (MP):
1. **Minimizes total evolutionary changes** (mutations) in the tree.
2. **Uses character states (nucleotide/amino acid positions)** to determine the most likely ancestral relationships.
3. **Ideal for closely related sequences** but computationally intensive for large datasets.

Applications of MP Trees:
- Understanding **evolutionary history** of species or genes.
- Identifying **ancestral traits and lineage divergence**.
- Studying genetic variation and **conserved regions** in sequences.

Diagram:



Procedure:
Step 1: Install and Open MEGA Software
- Download and install **MEGA** from the official website.
- Open the software and select **Phylogeny → Construct/Analyze Phylogenetic Tree**.

Step 2: Load and Align Sequences

- Obtain DNA or protein sequences in **FASTA format**.
- Open **MEGA → Align → Edit/Build Alignment** to import sequences.
- Perform **Multiple Sequence Alignment (MSA)** and save as a **MEGA (.meg) file**.

Step 3: Select Maximum Parsimony Method

- Click **Phylogeny → Construct Maximum Parsimony Tree**.
- Choose **MP search criteria**, such as **Close-Neighbor-Interchange (CNI)** for branch swapping.
- Select the **substitution model** (e.g., Parsimony for nucleotides or proteins).

Step 4: Perform Bootstrap Analysis (Optional for Reliability)

- In the MP settings, enable **Bootstrap Test (≥1000 replicates)**.
- Higher bootstrap values indicate **stronger support** for evolutionary relationships.

Step 5: View and Interpret the Tree

- The MP tree will display **branch lengths and bootstrap values**.
- Save the tree by clicking **File → Export Tree as Image**.

Observations Table:

Step	Observation	Interpretation
Sequence alignment	Aligned sequences loaded into MEGA	Ready for tree construction
MP tree calculation	Tree structure appears with branches	Evolutionary relationships inferred
Bootstrap values	Support values displayed on branches	Confidence in evolutionary history

Result:

A **Maximum Parsimony (MP) phylogenetic tree** was successfully constructed using **MEGA software**, illustrating the **evolutionary relationships** among the input sequences.

Conclusion:

The **Maximum Parsimony (MP) method** provides a simple and effective way to infer evolutionary relationships. It is particularly useful for closely related species but may be computationally challenging for large datasets.

Precautions:

- Ensure sequences are **properly aligned** before tree construction.
- Use **bootstrap analysis** to confirm tree reliability.
- Select appropriate **parsimony criteria (CNI, SPR, TBR)** based on dataset size.
- Verify sequence quality to prevent misinterpretation.
- Save results and **export the tree image** for documentation.

Lab Activity No 09
Title of the Activity- ML Phylogenetic Tree Construction Using MEGA

Learning Objectives:

- Understand the **Maximum Likelihood (ML) method** for phylogenetic tree construction.
- Learn how to use **MEGA (Molecular Evolutionary Genetics Analysis)** software to build ML trees.
- Analyze **evolutionary relationships** between DNA/protein sequences using the ML approach.

Aim:
To construct a **Maximum Likelihood (ML) phylogenetic tree** using **MEGA software** and interpret the evolutionary relationships among sequences.

Apparatus & Materials:

- **Computer with Windows/Linux/Mac OS**
- **MEGA software (Download from: https://www.megasoftware.net/)**
- **FASTA format sequence file**
- **Internet access (optional, for sequence retrieval from NCBI)**

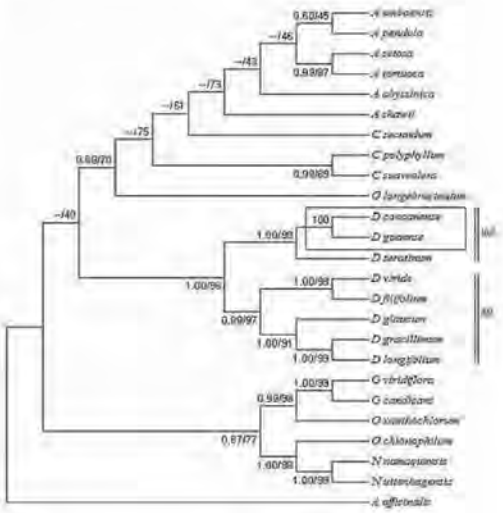
Theory:
A **phylogenetic tree** is a diagram representing the evolutionary relationships among organisms or genes. The **Maximum Likelihood (ML) method** is a statistical approach that finds the **best tree** by **calculating probabilities of different evolutionary events**.
Principles of Maximum Likelihood (ML) Method:

1. Uses a **statistical model of evolution** (e.g., Jukes-Cantor, Kimura 2-parameter).
2. Evaluates multiple tree topologies to **find the most likely tree**.
3. Computes **branch lengths** based on **likelihood scores**.
4. Computationally intensive but **more accurate than NJ or MP**.

Applications of ML Trees:

- Studying **species evolution and genetic divergence**.
- Inferring **ancestral relationships** in complex datasets.
- Comparing **mutation rates and selection pressure** in genes.

Diagram:



Procedure:
Step 1: Install and Open MEGA Software

- Download and install **MEGA** from the official website.
- Open the software and select **Phylogeny → Construct/Analyze Phylogenetic Tree**.

Step 2: Load and Align Sequences

- Obtain DNA or protein sequences in **FASTA format**.
- Open **MEGA → Align → Edit/Build Alignment** to import sequences.
- Perform **Multiple Sequence Alignment (MSA)** and save as a **MEGA (.meg) file**.

Step 3: Select Maximum Likelihood Method

- Click **Phylogeny → Construct Maximum Likelihood Tree**.
- Choose an **evolutionary model** (e.g., **Kimura 2-parameter, JTT for proteins**).

Step 4: Perform Bootstrap Analysis (Optional for Reliability)

- In the ML settings, enable **Bootstrap Test (≥1000 replications)**.
- Higher bootstrap values indicate **stronger support** for evolutionary relationships.

Step 5: View and Interpret the Tree

- The ML tree will display **branch lengths and bootstrap values**.
- Save the tree by clicking **File → Export Tree as Image**.

Observations Table:

Step	Observation	Interpretation
Sequence alignment	Aligned sequences displayed in MEGA	Ready for tree construction
ML tree calculation	Tree structure appears with branches	Evolutionary relationships inferred
Bootstrap values	Support values displayed on branches	Confidence in evolutionary history

Result:

A **Maximum Likelihood (ML) phylogenetic tree** was successfully constructed using **MEGA software**, revealing **evolutionary relationships** among the input sequences.

Conclusion:

The **Maximum Likelihood (ML) method** is a powerful approach for constructing accurate phylogenetic trees. It provides **statistically robust evolutionary relationships** based on sequence data.

Precautions:

- Ensure sequences are **properly aligned** before tree construction.
- Choose an **appropriate substitution model** for accurate distance calculations.
- Use **bootstrap analysis** to confirm tree reliability.
- Verify sequence quality to avoid misinterpretation.
- Save results and **export the tree image** for documentation.

Lab Activity No 10
Title of the Activity- Construction of chloroplast genome map using GeSeq

Learning Objectives:

- Understand the process of **chloroplast genome annotation** using **GeSeq (Gene Sequence Annotation Server)**.
- Learn how to use **GeSeq** to predict and visualize **chloroplast genome features** such as protein-coding genes, rRNAs, and tRNAs.
- Construct a **chloroplast genome map** and analyze its structure.

Aim:
To annotate and construct a **chloroplast genome map** using **GeSeq**, identifying genes and structural elements within the chloroplast genome.

Apparatus & Materials:
- **Computer with Internet access**
- **GeSeq (Available at: https://chlorobox.mpimp-golm.mpg.de/geseq.html)**
- **FASTA file of chloroplast genome sequence**
- **NCBI or other sequence repositories for reference genomes**

Theory:
The **chloroplast genome** is a circular, double-stranded DNA molecule found in plant cells, primarily responsible for **photosynthesis and other essential metabolic processes**.
Annotation of the chloroplast genome is crucial for studying **gene functions, evolutionary relationships, and genetic engineering applications**.
What is GeSeq?
- **GeSeq (Gene Sequence Annotation Server)** is an online tool used for **annotating chloroplast genomes**.
- It identifies **protein-coding genes, rRNAs, tRNAs, and conserved regions** based on sequence comparison with reference databases.
- It also generates **graphical genome maps**, which are useful for **genome visualization and comparative genomics**.

Applications of Chloroplast Genome Annotation:
- Understanding **chloroplast gene functions**.
- Studying **plant phylogenetics and evolution**.
- Assisting in **genetic modification and synthetic biology**.
- Detecting **mutations and structural variations** in chloroplast genomes.

Diagram:

Procedure:
Step 1: Access GeSeq Online
- Open **GeSeq** by visiting: https://chlorobox.mpimp-golm.mpg.de/geseq.html.

Step 2: Upload the Chloroplast Genome Sequence
- Click **"Upload FASTA file"** and select the chloroplast genome sequence file.
- Alternatively, enter the **NCBI accession number** to fetch an existing sequence.

Step 3: Select Annotation Parameters

- Choose databases for comparison: **NCBI RefSeq, EMBL, or BLAST against existing chloroplast genomes**.
- Enable **tRNA and rRNA annotation** using **tRNAscan-SE and RNA-BLAST**.
- Set the **annotation output format** (e.g., **GFF, GenBank, or graphical genome map**).

Step 4: Run the Annotation Process

- Click **"Start Annotation"** and wait for the tool to process the genome.
- The process may take a few minutes, depending on genome size.

Step 5: View and Download Annotated Genome

- Check the **list of identified genes (protein-coding, tRNA, rRNA, and non-coding regions)**.
- Click **"Generate Genome Map"** to visualize the **circular chloroplast genome structure**.
- Download results in **GenBank format** for further analysis.

Observations Table:

Step	Observation	Interpretation
Genome upload	FASTA file successfully loaded	Ready for annotation
Annotation process	Genes, rRNAs, and tRNAs identified	Chloroplast genome elements detected
Genome map generation	Circular genome map displayed	Visualization of genome structure
Exported annotation	Gene list and sequence data available	Data ready for comparative analysis

Result:

A **chloroplast genome map** was successfully constructed using **GeSeq**, identifying key genetic elements such as **protein-coding genes, rRNAs, and tRNAs**.

Conclusion:

The **GeSeq tool** effectively annotates **chloroplast genomes** by comparing sequences with reference databases. This process provides a comprehensive view of **chloroplast gene organization**, useful for **phylogenetic studies, evolutionary biology, and genetic research**.

Precautions:

- Ensure the **input sequence is high quality** and in **FASTA format**.
- Select appropriate **reference databases** for accurate annotation.
- Verify **gene predictions** with other annotation tools if needed.
- Use **graphical visualization tools** to confirm gene positions.
- Save and backup **annotation results and genome maps** for further analysis

THANKS

MODEL QUESTIONS

Multiple choice question

Who is known as the father of taxonomy?
a) Aristotle
b) Carl Linnaeus
c) Charles Darwin
d) Gregor Mendel
Answer: b) Carl Linnaeus
Which of the following plant classification systems is primarily molecular-based?
a) Bentham and Hooker System
b) Linnaean System
c) APG System
d) Artificial System
Answer: c) APG System
Binomial nomenclature consists of how many Latin words?
a) One
b) Two
c) Three
d) Four
Answer: b) Two
What is the lowest major taxonomic group in plant classification?
a) Order
b) Family
c) Genus
d) Species
Answer: d) Species

Fill in the blanks

Scientific names are treated as_____, regardless of derivation.
Answer: Latin

The Bentham and Hooker system of classification is a _____system.
Answer: **natural**

Mark the True / False

Taxonomy is independent of zoological nomenclature. **(True)**
Linnaeus classified plants based on their vegetative characteristics. **(False)**

Fill in the blanks

_____ observed that species are widely understood but difficult to define.

Answer: Charles Darwin

The suffix "-aceae" is used for naming _____ in taxonomy.

Answer: Families

The species **Rosa alba** belongs to the genus _____.

Answer: Rosa

In taxonomic hierarchy, _____ is the rank above Family.

Answer: Order

The nucleus contains genetic material in the form of _____ or _____.

Answer: chromatin, chromosomes

_____ are the units of heredity made of DNA.

Answer: Genes

Mark the True / False (

Species is the highest taxonomic rank. **(False)**

A taxon can belong to multiple categories. **(False)**

A dendrogram represents evolutionary relationships. **(True)**

The family Poaceae includes genera like Poa. **(True)**

DNA is composed of three strands.

Answer: False

Watson and Crick proposed the double-helix model of DNA.

Answer: True

The mitochondrial genome in plants is mainly responsible for _____.
Answer: energy production
The _____ region is a commonly used intergenic spacer in plant DNA studies.
Answer: trnH-psbA
The nuclear genome consists of both _____ and _____ DNA.
Answer: coding, non-coding
_____ sequencing is used for high-throughput DNA analysis.
Answer: Next-Generation Sequencing (NGS)

The nuclear genome is linear in structure. **(True)**
The chloroplast genome follows biparental inheritance. **(False)**
The mitochondrial genome in plants is circular. **(True)**
rbcL is a nuclear gene used in phylogenetics. **(False)**

What is the most inclusive taxonomic category in the classification of vascular plants?
a) Family
b) Genus
c) Species
d) Division
Answer: d) Division
Which of the following taxonomic ranks directly precedes "Family" in the hierarchy?
a) Species
b) Order
c) Genus
d) Class
Answer: b) Order
The naming conventions in plant classification follow which system?
a) Binomial nomenclature
b) Trinomial nomenclature
c) Phylogenetic nomenclature
d) Descriptive nomenclature
Answer: a) Binomial nomenclature
What does a dendrogram primarily illustrate?
a) Evolutionary relationships
b) Genetic modifications
c) Environmental adaptations
d) Growth patterns of plants
Answer: a) Evolutionary relationships
Which category includes multiple genera but is below the order level?
a) Species
b) Class
c) Family
d) Division
Answer: c) Family
Which of the following is a monocot characteristic?
a) Reticulate venation
b) Pentamerous flowers
c) Trimerous flowers
d) Two cotyledons
Answer: c) Trimerous flowers
The term "-aceae" in plant classification refers to which taxonomic rank?
a) Species
b) Family
c) Order
d) Genus
Answer: b) Family
The **Nominalistic Species Concept** primarily focuses on:
a) Evolutionary relationships
b) Logical classification without biological significance
c) Genetic variation within species
d) Hybridization of plants
Answer: b) Logical classification without biological significance

What is the basic unit of life?
a) Nucleus
b) DNA
c) Cell
d) Chromosome
Answer: c) Cell

Who discovered the double-helix structure of DNA?
a) Watson & Crick
b) Rosalind Franklin
c) Maurice Wilkins
d) Gregor Mendel
Answer: a) Watson & Crick

Which nitrogenous base pairs with thymine (T) in DNA?
a) Cytosine (C)
b) Guanine (G)
c) Adenine (A)
d) Uracil (U)
Answer: c) Adenine (A)

What is the function of DNA polymerase in DNA replication?
a) Breaks down proteins
b) Synthesizes new DNA strands
c) Transports genetic material
d) Converts DNA to RNA
Answer: b) Synthesizes new DNA strands

What is the primary goal of molecular systematics?
a) Study of fossils
b) Classification of plants using molecular data
c) Identification of soil types
d) Study of weather patterns
Answer: b) Classification of plants using molecular data
Which molecule is commonly used in plant phylogenetic studies?
a) Hemoglobin
b) Chlorophyll
c) DNA
d) ATP
Answer: c) DNA
Which molecular marker is widely used in phylogenetic analysis?
a) rbcL
b) Myoglobin
c) Insulin
d) Actin
Answer: a) rbcL
What is the function of the chloroplast genome?
a) Control cell division
b) Regulate photosynthesis
c) Digest cellular waste
d) Store energy
Answer: b) Regulate photosynthesis
The mitochondrial genome in plants is mainly responsible for:
a) Oxygen transport
b) Energy production
c) Photosynthesis
d) DNA replication
Answer: b) Energy production
The nuclear genome contains:
a) Only protein-coding genes
b) Both coding and non-coding DNA
c) Only repetitive sequences
d) Only tRNA genes
Answer: b) Both coding and non-coding DNA
The atpB gene is part of which genome?
a) Mitochondrial
b) Chloroplast
c) Nuclear
d) Plasmid
Answer: b) Chloroplast
What is the structure of the chloroplast genome?
a) Linear
b) Circular
c) Single-stranded
d) Fragmented
Answer: b) Circular

Figure based question

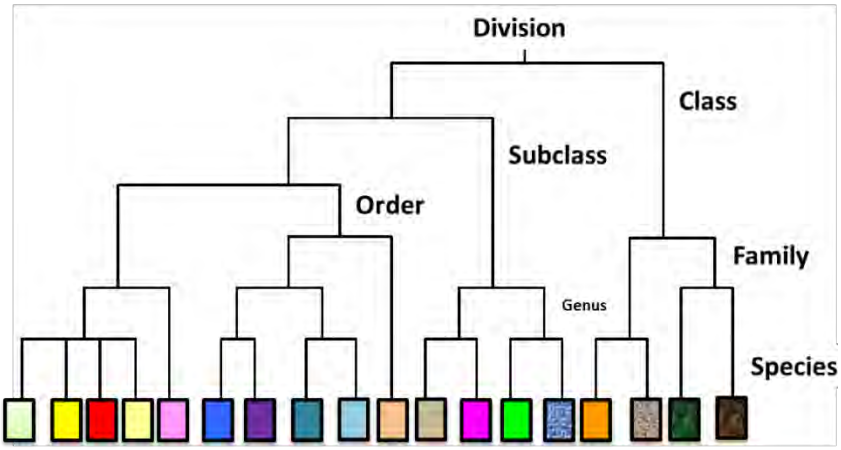
Look at the image of a herbarium sheet with a pressed plant specimen. Which part of the specimen is the most important for identification?

Question: What is this and why is it important in taxonomy?

Answer: This is a herbarium specimen, used for plant identification, documentation, and reference in taxonomy.

Figure based question

1.Observe an image of different plant taxa in a branching diagram. What does it represent?



Answer: Dendrogram

Short answer question

What is binomial nomenclature?
Answer: Binomial nomenclature is a system of naming species using two Latin words: the first for the genus and the second for the species (e.g., *Mangifera indica*).

What is a taxonomic hierarchy?
Answer: It is an organized ranking system where organisms are classified into different levels, such as species, genus, family, order, and so on.

Explain the purpose of a dendrogram.
Answer: A dendrogram is a tree-like diagram used to visually represent evolutionary relationships and classification hierarchies among organisms.

What are chromosomes made of?
Answer: Chromosomes are made of DNA and proteins called histones.

What is molecular systematics?
Answer: It is the study of evolutionary relationships using molecular data such as DNA and proteins.
What is the role of rbcL in plant systematics?
Answer: rbcL is a chloroplast gene commonly used for phylogenetic studies and DNA barcoding in plants.

What is an outgroup in a phylogenetic tree?
A species or group used as a reference that is closely related but not part of the main study group.
Define a clade in a phylogenetic tree.
A group consisting of a common ancestor and all its descendants.
What is the difference between a cladogram and a phylogram?
A cladogram shows only branching patterns, whereas a phylogram represents genetic changes using branch lengths.
What is the primary basis of Phenetics?
Overall similarity among organisms, not evolutionary relationships.

What is the primary reason genetic diversity is important?
a) It increases an organism's size
b) It enhances adaptability to environmental changes
c) It reduces reproduction rates
d) It eliminates mutations
Answer: b) It enhances adaptability to environmental changes.
Which of the following factors does NOT affect genetic diversity?
a) Natural selection
b) Genetic drift
c) Artificial intelligence
d) Gene flow
Answer: c) Artificial intelligence.
Which method is commonly used for measuring genetic diversity?
a) Microsatellites
b) SNPs
c) Whole genome sequencing
d) All of the above
Answer: d) All of the above.
Which gene is primarily used for DNA barcoding in animals?
a) COI
b) rbcL
c) matK
d) ITS
Answer: a) COI.
What is the main purpose of PCR in DNA barcoding?
a) Breaking down DNA
b) Amplifying DNA
c) Coloring DNA
d) Storing DNA
Answer: b) Amplifying DNA.
Which of the following is a DNA barcode database?
a) Google
b) Wikipedia
c) BOLD
d) PubMed
Answer: c) BOLD.
In plants, which genes are commonly used for DNA barcoding?
a) COI
b) rbcL and matK
c) ITS
d) 16S rRNA
Answer: b) rbcL and matK.
What is the main advantage of DNA barcoding?
a) It eliminates mutations
b) It allows for quick and accurate species identification
c) It modifies genetic traits
d) It replaces taxonomy
Answer: b) It allows for quick and accurate species identification.
Which step comes immediately after DNA sequencing in barcoding?
a) PCR amplification
b) Comparison with barcode databases
c) DNA extraction
d) Sample collection
Answer: b) Comparison with barcode databases.
What does SNP stand for in genetic diversity studies?
a) Single Nucleotide Polymorphism
b) Sequence Normalized Protein
c) Short Nucleotide Pair
d) Single Nucleotide Pathway
Answer: a) Single Nucleotide Polymorphism.

What does a phylogenetic tree represent?
a) Geological changes
b) Evolutionary relationships
c) Physical structures of organisms
d) Genetic mutations
Answer: b) Evolutionary relationships
What is the main function of an outgroup in a phylogenetic tree?
a) To represent an ancestor
b) To serve as a reference for comparison
c) To indicate a species that evolved first
d) To classify extinct species
Answer: b) To serve as a reference for comparison
Which type of phylogenetic tree does not show a common ancestor?
a) Rooted tree
b) Phylogram
c) Unrooted tree
d) Chronogram
Answer: c) Unrooted tree
What is the primary basis for classifying organisms in Phenetics?
a) Evolutionary relationships
b) Genetic mutations
c) Overall similarity
d) Common ancestry
Answer: c) Overall similarity
Which statistical method is used in Phenetics for similarity measurement?
a) Bayesian Inference
b) Euclidean Distance
c) Cladistics
d) Homoplasy Analysis
Answer: b) Euclidean Distance
What does a cladogram show?
a) Evolutionary time scale
b) Genetic changes in DNA
c) Only branching patterns
d) Fossil records
Answer: c) Only branching patterns
What is the fundamental unit of classification in Phenetics?
a) Clade
b) Phylogram
c) Operational Taxonomic Unit (OTU)
d) Node
Answer: c) Operational Taxonomic Unit (OTU)
Which of the following is an example of discontinuous variation?
a) Height of a plant
b) Leaf size in trees
c) Flower color in pea plants
d) Weight of mango fruits
Answer: c) Flower color in pea plants
What method is used for tree construction in Phenetics?
a) UPGMA
b) Maximum Parsimony
c) Maximum Likelihood
d) Bayesian Inference
Answer: a) UPGMA
What does branch length represent in a phylogram?
a) Number of genetic changes
b) Evolutionary relationships
c) Physical characteristics
d) Common ancestors
Answer: a) Number of genetic changes

_____ is the primary molecular marker used for DNA barcoding in animals.
Answer: COI
_____ introduces new genetic variations into a population.
Answer: Mutation
The _____ effect occurs when a population undergoes a dramatic reduction in size.
Answer: Bottleneck
Microsatellites are also known as _____.
Answer: Short Tandem Repeats (STRs)
SNPs stand for _____.
Answer: Single Nucleotide Polymorphisms
In DNA barcoding, the _____ gene is used for animals.
Answer: COI
The most ancestral node in a phylogenetic tree is called the _____.
Answer: root
A _____ includes a common ancestor and all its descendants.
Answer: clade
_____ is used as a reference in a phylogenetic tree.
Answer: Outgroup
In Phenetics, organisms are classified based on overall _____.
Answer: similarity
A _____ represents similarity but does not show evolutionary history.
Answer: phenogram
_____ is a statistical method used in phenetics.
Answer: Euclidean distance
In cladistics, a _____ is a shared derived trait.
Answer: synapomorphy
The database BOLD stands for _____.
Answer: Barcode of Life Data System

DNA barcoding is used only for animals. **(False)**
PCR is used to amplify DNA sequences in barcoding. **(True)**
The *COI* gene is used for DNA barcoding in plants. **(False)**
Genetic drift is a form of natural selection. **(False)**
SNPs are variations at a single nucleotide position. **(True)**
BOLD and GenBank are databases used in DNA barcoding. **(True)**
Whole genome sequencing provides less genetic detail than microsatellites. **(False)**
DNA barcoding can be used to detect food fraud. **(True)**
A phylogenetic tree always includes a root. **(False, unrooted trees do not have a root.)**
Cladistics groups organisms based on overall similarity. **(False, it groups based on common ancestry.)**
Phenetics does not consider evolutionary history. **(True)**
Outgroups help determine the evolutionary direction of a phylogenetic tree. **(True)**
A phenogram represents evolutionary history. **(False, it represents similarity only.)**
Character weighting is used to give all traits equal importance. **(False, it assigns importance to selected traits.)**
Jaccard's Coefficient is used for similarity measurement in Phenetics. **(True)**
Cladistics always uses genetic data for classification. **(False, it can also use morphological data.)**

Match the following)

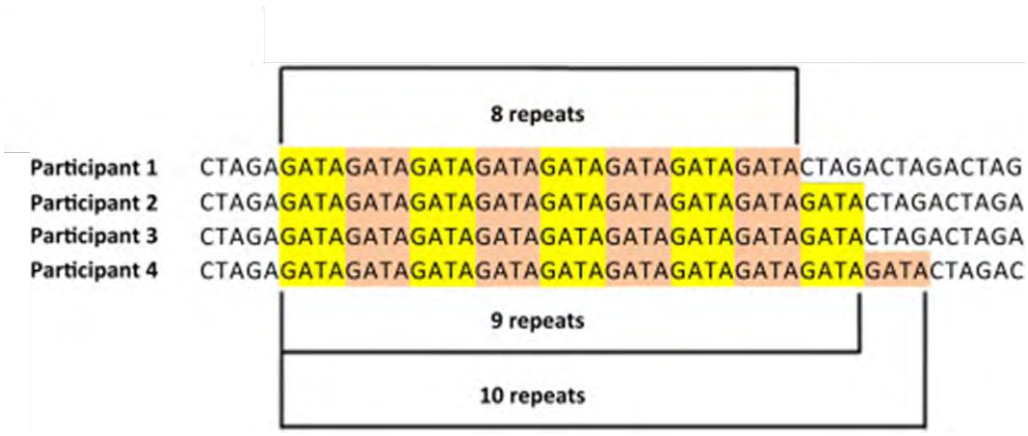
Question

Column A	Column B
UPGMA	Clustering method in phenetics
rbcL and matK	Plants
Phylogenetic Tree	Evolutionary relationships
Phenetics	Overall similarity
PCR	DNA Amplification
OTU	Basic unit of classification
Microsatellites	STRs
ITS Region	Fungi
DNA Barcoding	Species Identification
COI Gene	Animals
Cladogram	Branching patterns only
Cladistics	Common ancestry

Answers

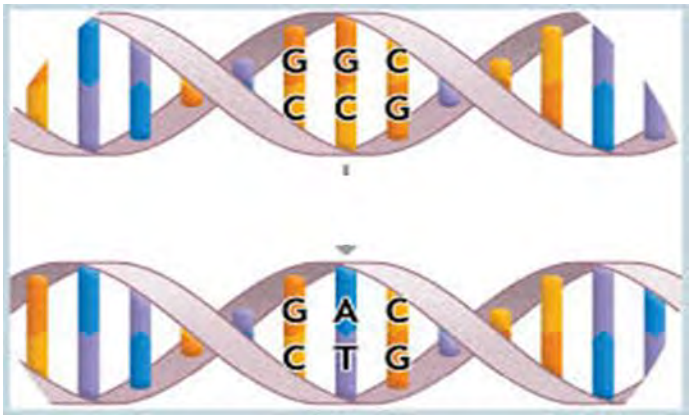
Column A	Column B
DNA Barcoding	Species Identification
COI Gene	Animals
rbcL and matK	Plants
ITS Region	Fungi
PCR	DNA Amplification
Microsatellites	STRs
Phylogenetic Tree	Evolutionary relationships
Cladogram	Branching patterns only
Phenetics	Overall similarity
Cladistics	Common ancestry
UPGMA	Clustering method in phenetics
OTU	Basic unit of classification

Observe the image and identify the repeating sequences present..



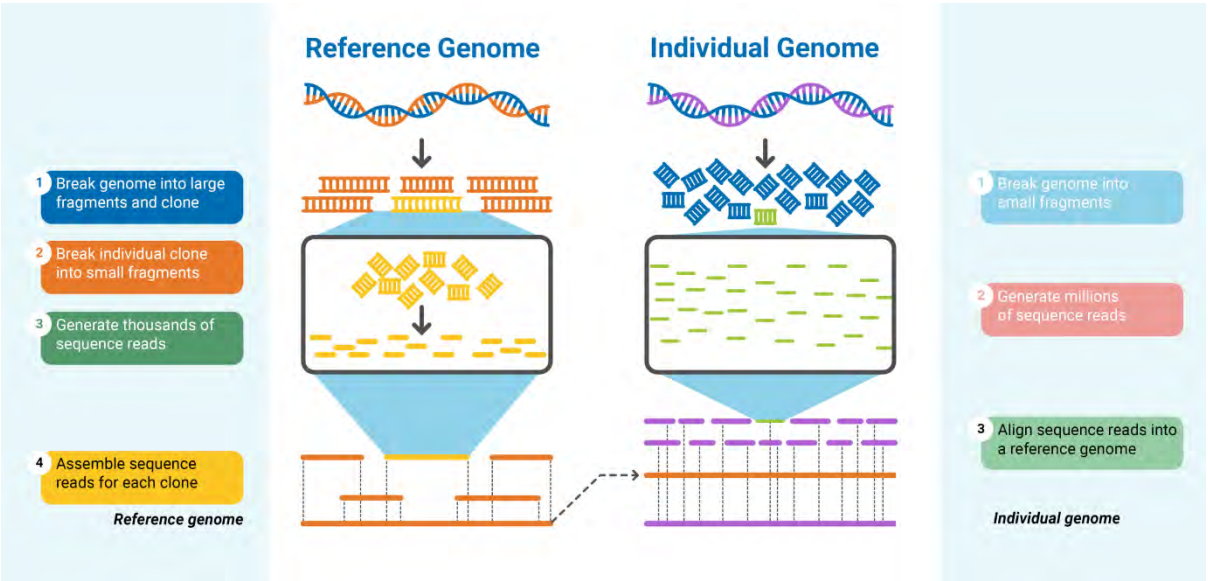
Answer; Short Tandem Repeats

Observe the image and identify the repeating sequences present..



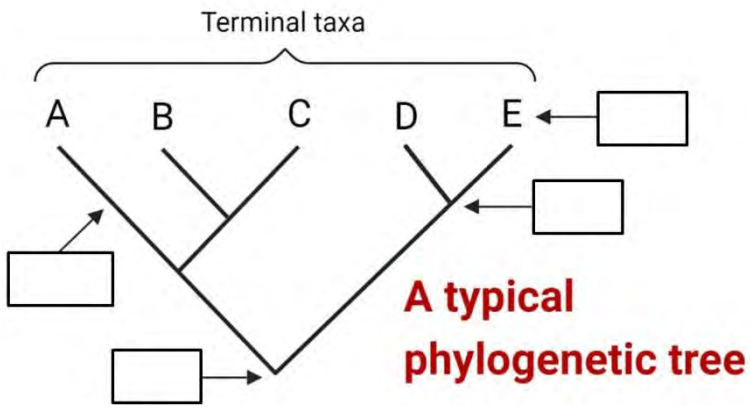
Answer: Single Nucleotide Polymorphisms

Observe the image and identify the genome type.

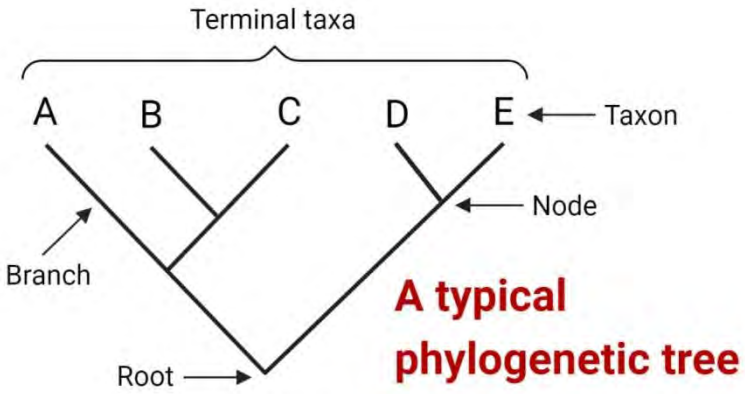


Answer: Whole Genome Sequencing (WGS)

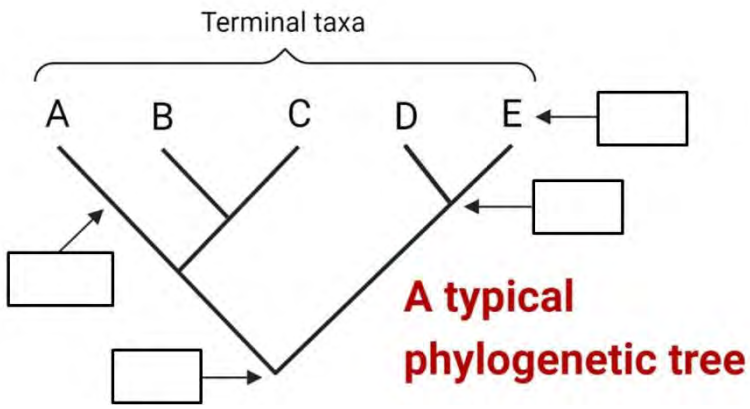
Label the **nodes** of the given tree.



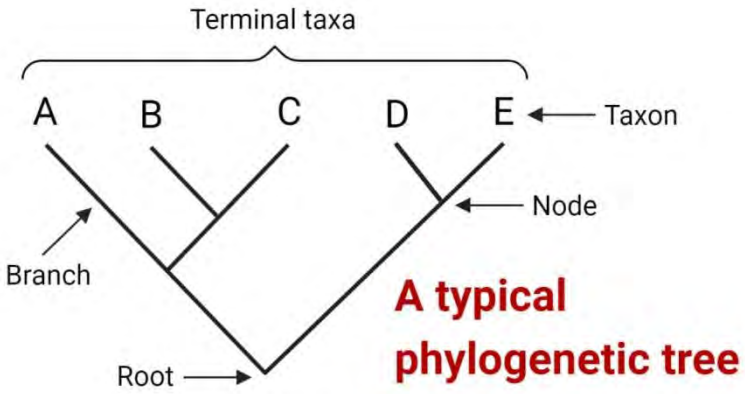
Answer;



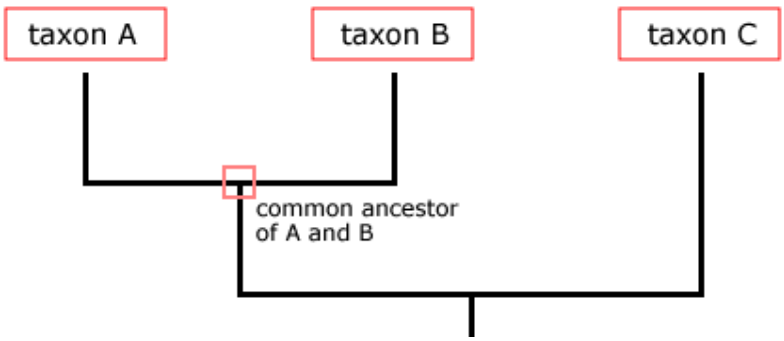
9 Label the **Taxon** of the given tree.



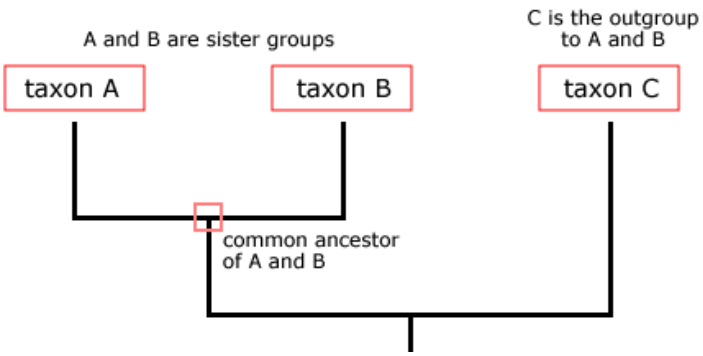
Answer;



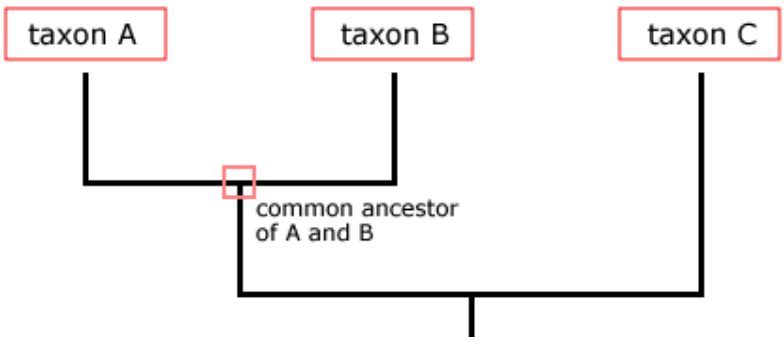
In the provided phylogenetic tree, mark the outgroup



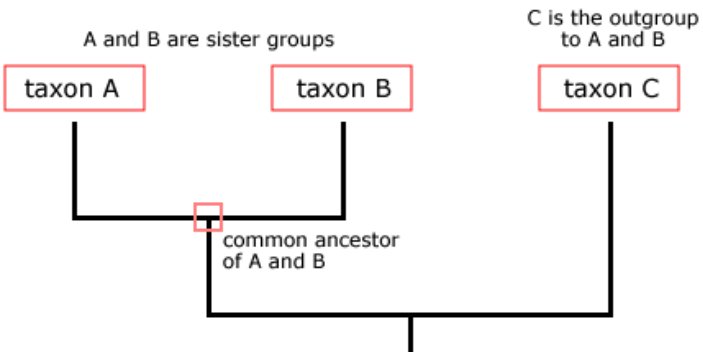
Answer:



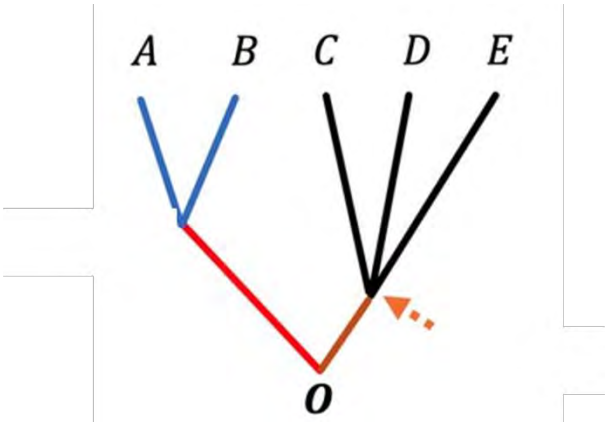
In the provided phylogenetic tree, mark the sister taxa



Answer: Taxon A and B

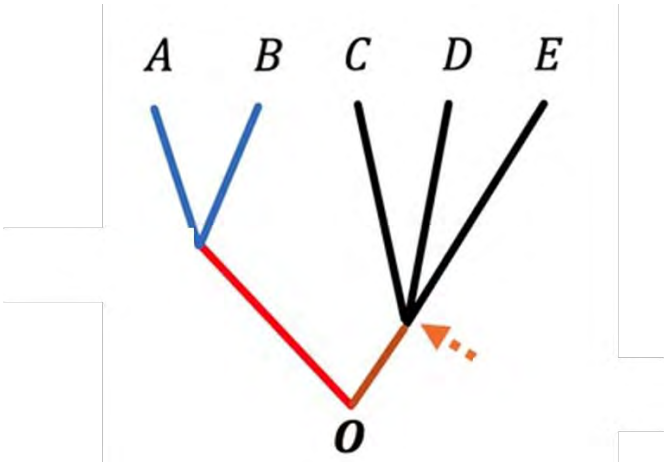


. In the provided phylogenetic tree, taxon A and B represents?



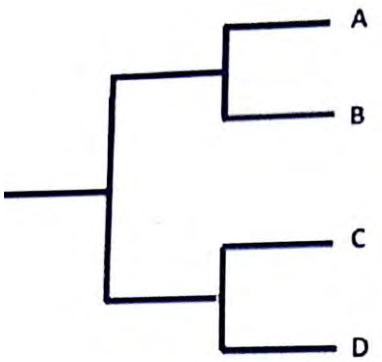
Answer: Dichotomy

In the provided phylogenetic tree, taxon C, D and E represents?



Answer: Polytomy

Find out the type of phylogenetic tree?



Answer: Cladogram