E-RESOURCE

UDAI PRATAP COLLEGE, VARANASI

Programme/Class: Diploma in Plant Identification, Utilization & Ethnomedicine
UG, Year: II, Semester: III, Paper: I, Subject: Botany
Course code: B040301T (NPE-2020); **Course Title:** Flowering Plants Identification & Aesthetic Characteristics
Credits: 4, Course compulsory, Max Marks: 25+75

Class: B.Sc. –Botany
Year: II, Paper: I, UNIT: II
Topic: Molecular Systematics-Protein and Nucleic acid homology
Name: Prof. Ajai Kumar Singh, Department of Botany, Faculty of Science,
Mobile No. 9450538149, E-mail: ajaiupcollege@gmail.com

MOLECULAR SYSTEMATICS (MS)

Molecular Systematics (MS) is a branch of biology that uses molecular data to study the evolutionary relationships among organisms. It has revolutionized the field of evolutionary biology by providing a powerful tool for reconstructing the tree of life and understanding genetic basis of biodiversity.

Principles of MS:

MS relies on the principles of molecular evolution, which include the assumption that DNA sequences changes over time at a relatively constant rate. By comparing these changes in DNA sequences, it is possible to infer the degree of relatedness between different species.

Molecular Markers:

One of the key components of MS is the use of molecular markers, such as DNA sequences, to infer evolutionary relationships. These markers can include mitochondrial

DNA, ribosomal DNA, and protein-coding genes. By analyzing these markers, we can determine the genetic similarities and differences between organisms.

Phylogenetic Analysis:

Phylogenetic analysis is a fundamental tool in MS. It involves constructing phylogenetic trees that depict the evolutionary relationships among different species or groups of organisms. These trees are based on the similarities and differences in molecular markers and provide insights into the patterns of evolution and diversification.

Application of MS:

MS has diverse applications in various fields, including **taxonomy**, **conservationbiology**, and **biogeography**. It has been instrumental in *clarifying the classification of organisms, identifying cryptic species, and understanding the genetic basis of adaptation and speciation.*

Advantages and Limitations:

One of the major advantages of MS is its ability to provide robust and objective evidence for evolutionary relationships. However, it also has limitations, such as potential biases in sequence data, incomplete lineage sorting, and horizontal gene transfer.

Future Directions:

The field of MS continues to advance with technological innovations in DNA sequencing and bioinformatics. ***High-throughput*** sequencing technologies and computational methods are enhancing our ability to analyse large-scale molecular datasets and unravel complex evolutionary patterns.

> **'HIGH-THROUGHPUT'** IN SEQUENCING REFERS TO THE AMOUNT OF DNA MOLECULES READ AT THE SAME TIME. TECHNOLOGIES ARE NOW CAPABLE OF SEQUENCING MANY FRAGMENTS OF DNA IN PARALLEL. THIS ENABLES RESEARCHERS TO READ HUNDREDS OF MILLIONS OF DNA FRAGMENTS AND GENERATE MORE DATA, WITH LESS TIME THAN EVER BEFORE.

Overall, MS plays a pivotal role in elucidating the history of life on Earth and provides valuable insights into the processes that have shaped biological diversity.

- MS requires phylogenetic analysis as a tool for studying phylogenetic relationships of living organisms.
- In general, MS provides a powerful statistical framework for hypothesis testing and the estimation of evolutionary processes.
- MS is the use of molecular genetics to study the evolution of relationships among individuals and species.
- MS is the discipline of classifying organisms based on differences in protein and DNA in order to make fine taxonomic classifications that are not solely based on morphology.

MONOPHYLETIC, PARAPHYLETIC and **POLYPHYLETIC** are terms used in the classification of organisms to describe the evolutionary relationships between different groups of organisms.

MONOPHYLETIC GROUPS are those that include all descendants of a common ancestor.	PARAPHYLETIC GROUPS include some, but not all descendants of a common ancestor.	POLYPHYLETIC GROUPS are made up of organisms that have converged on a similar characteristic but do not share a common ancestor.

Understanding the difference between these three terms is essential for the classification and understanding of evolutionary relationships between organisms.

MONOPHYLETIC GROUP

A monophyletic group is a group of organisms includes an ancestral species and all of its descendants. In other words, it is a group that has a *single common ancestor* and is considered a natural group or clade. This can be contrasted with a polyphyletic group, which includes organisms from multiple evolutionary lineages that do not share a common ancestor. *In taxonomy, monophyletic groups are considered to be the most scientifically valid grouping of organisms, as they reflect evolutionary relationship.* A monophyletic group is considered a true group because it reflects evolutionary relationships and the natural classification of organisms.

The word "monophyletic" comes from the Greek words "monos," meaning "single," and "phyton," meaning "race" or "tribe." The suffix "-etic" is used to indicate that something belongs to or is characteristic of certain group or category.

Together, monophyletic means **"belonging to a single tribe"** or **"descended from a single ancestor."**

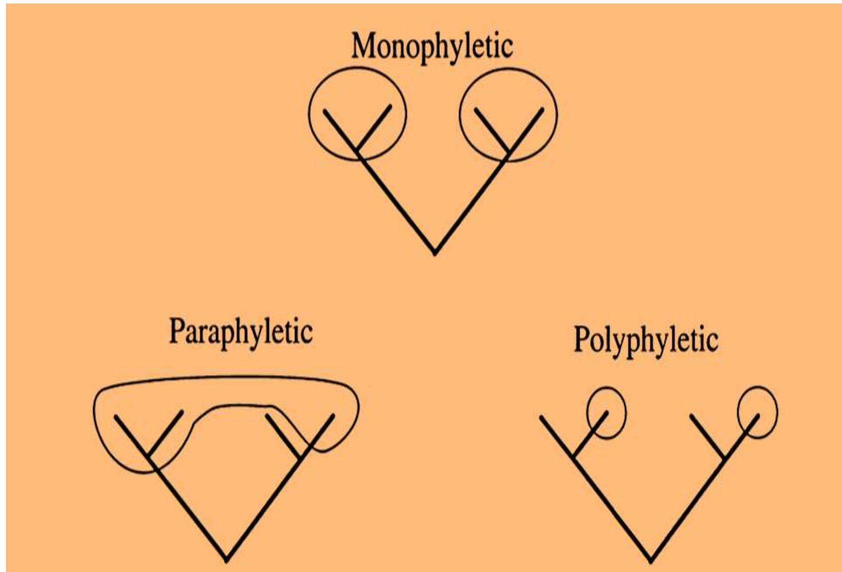
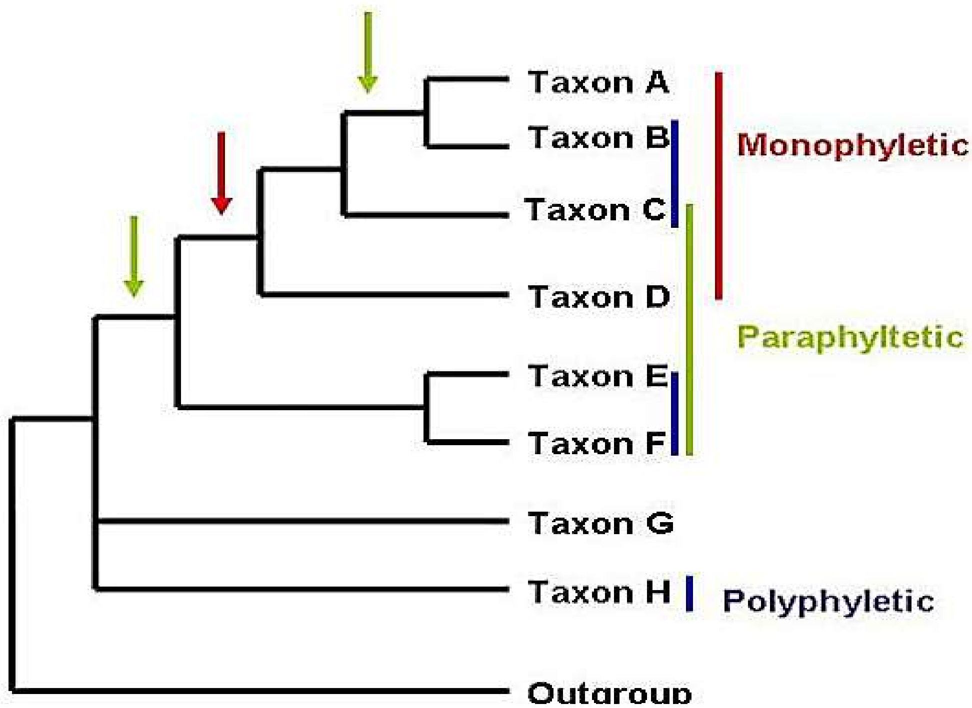


Fig. Cladistic relationships relative to cladograms.



A **MONOPHYLETIC GROUP**, also known as a clade or a natural group, is a taxonomic group that includes a common ancestor and all of its descendants. This means that all the organisms within the group share a common evolutionary history and are more closely related to each other than to any organisms outside the group.

A monophyletic group is defined by its unique evolutionary history, where it represents a single branch on the tree of life. It is characterized by three main features:

1. **Common Ancestor:** A monophyletic group includes an ancestral species from which all other members of the group have descended. This common ancestor can be traced back through a series of shared genetic traits or morphological characteristics.
2. **Shared Derived Traits:** Members of a monophyletic group share specific traits or characteristics that are derived from their common ancestor. These traits are known as SYNAPOMORPHIES and can include genetic mutations, anatomical structures, or behavioral patterns.
3. **Exclusive Membership:** A monophyletic group consists only of the common ancestor and all of its descendants. It does not include any organisms that do not share the same evolutionary lineage. This exclusivity ensures that all members of the group are more closely related to each other than to any organisms outside of the group.

Importance and Significance of Monophyletic Group: Monophyletic groups are essential in understanding evolutionary relationships and classifying organisms into hierarchical taxonomic categories. They provide insights into the diversification and evolution of species over time. By identifying monophyletic groups, it is possible to reconstruct phylogenetic trees that depict the evolutionary history and relatedness among different species.

The concept of monophyly also has practical applications in various fields of biology, including:

A. **Taxonomy**: Monophyletic groups serve as the foundation for modern taxonomic classifications. By organizing organisms into monophyletic groups, taxonomists can create a hierarchical system that reflects the evolutionary relationships among species.

B. **Conservation Biology**: Monophyletic groups help identify units of biodiversity that should be prioritized for conservation efforts. Protecting entire monophyletic groups ensures the preservation of unique evolutionary lineages and prevents the loss of genetic diversity.

C. **Biogeography**: Studying the distribution patterns of monophyletic groups can provide insights into historical biogeographic events such as continental drift, speciation, and dispersal. Understanding the geographic distribution of monophyletic groups helps explain how different lineages have evolved and spread across different regions.

PARAPHYLETIC GROUP

The term "Paraphyletic" is used to define a group of creatures that has a common ancestor and some of the descendants, but not all of them. In other words, it refers to a taxonomic group that does not contain all the descendants of a common ancestor. When a group of creatures is referred to as a paraphyletic group, it suggests that some of the natural group's members have been moved into another group for whatever reason. There are several potential causes for this.

The word “paraphyletic” comes from the Greek words “para,” meaning “beside” or “near.” And “phyton,” meaning “race,” or “tribe.” The suffix “-etic” is used to indicate that something belongs to or is characteristic of a certain group or category. Together, paraphyletic means “near a tribe” or “related to a tribe,” but not including all its descendants.

This group is characterized by the exclusion of one or more monophyletic groups, which are groups that include all the descendants of a common ancestor. This means that in a paraphyletic group, some descendants are excluded from the group based on certain characteristics and traits.

Paraphyletic groups are also known as **“partial groups” /“inclusive groups”** because they are based on evolutionary relationships but do not include all the descendants of the common ancestor. This type of classification is considered more useful than polyphyletic groups but less useful than monophyletic groups.

The exclusion of certain descendants from a group can lead to confusion and misrepresentation of evolutionary relationships. Paraphyletic groups can give the impression that certain organisms are more closely related than they actually are. Therefore, many modern taxonomists strive to avoid paraphyly and aim for monophyletic classifications that reflect the true evolutionary history of organisms. The shift towards cladistics and the recognition of monophyletic groups has led to a greater emphasis on accurate classification based on shared derived characteristics.

A Paraphyletic group is characterized by convergent evolution, where different species independently evolve similar traits due to adaptation to similar environments or ecological niches. These similarities can lead to the mistaken classification of these organisms as belonging to a single group. However, upon closer examination of their evolutionary history and genetic relationships, it becomes clear that they do not share a recent common ancestor.

Paraphyletic groups often arise when certain characteristics or traits have evolved multiple times in different lineages through independent evolutionary events.

Succulent plants are have adapted to arid environments by storing water in their leaves/stems. This trait has evolved independently in various plant families such as cacti (Cactaceae), and ice plants (Aizoaceae). While they all possess the ability to store water, these plant families have different evolutionary histories.

Identifying polyphyletic groups is crucial for accurate classification and understanding of evolutionary relationships among organisms. When organisms are mistakenly grouped together based on convergent traits, it can lead to misleading interpretations of their evolutionary history and relationships.

To ensure accurate classification, taxonomists rely on phylogenetic analysis, which involves examining the genetic relationships and evolutionary history of organisms. By using molecular data, such as DNA sequences, it is possible to reconstruct the evolutionary tree of life and determine the true relationship between organisms.

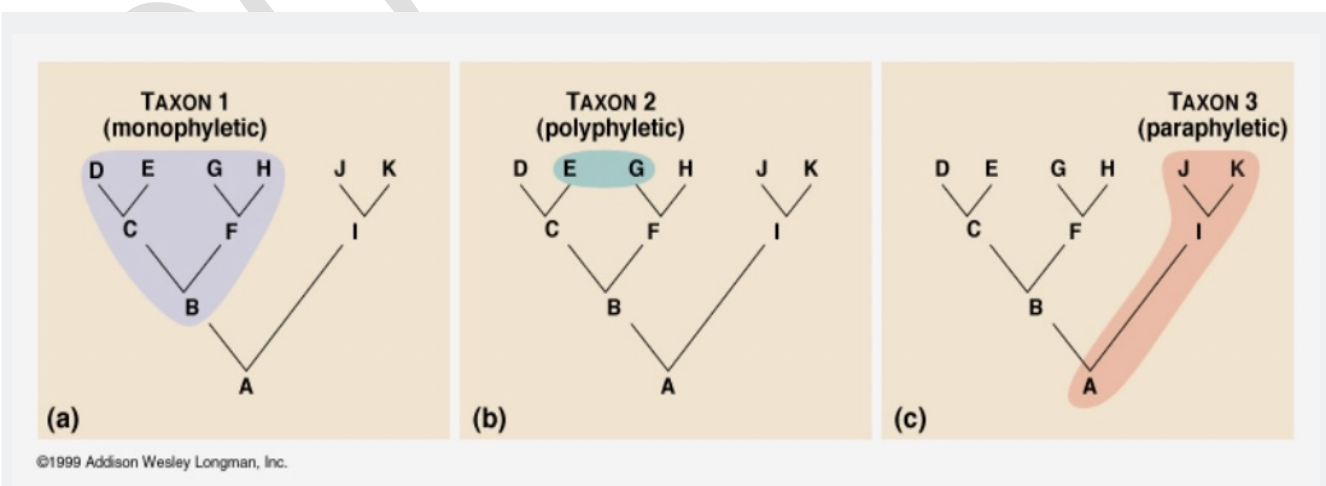
POLYPHYLETIC GROUP

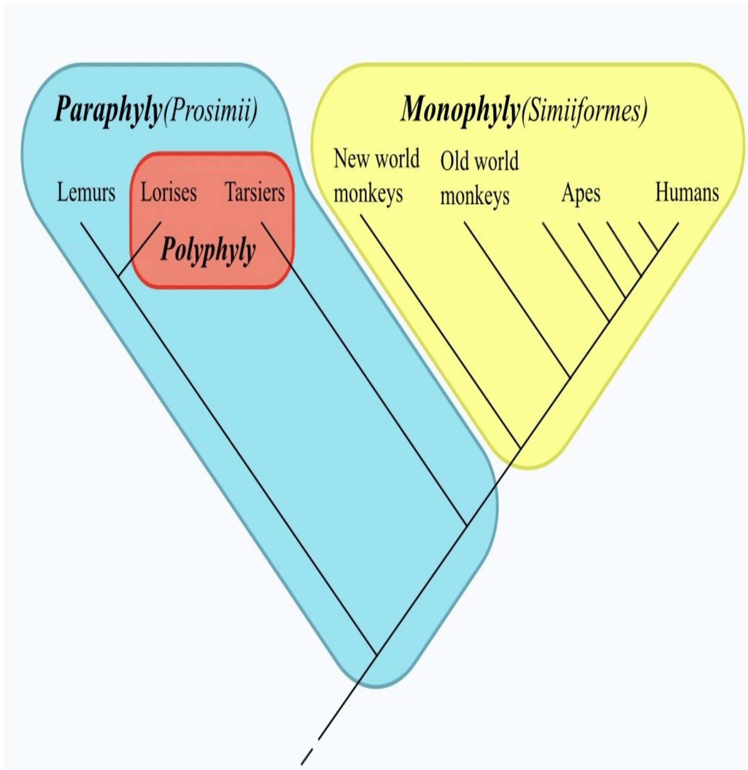
A group of creatures with no one common ancestor is referred to as a polyphyletic taxon. [**Share a common characteristics but do not share a common ancestor**] The polyphyletic group is made up of unrelated creatures that have several common ancestors. It's somewhat of an odd collection of creatures. A polyphyletic taxon is often reclassified when it is discovered since it is an entirely artificial assemblage.

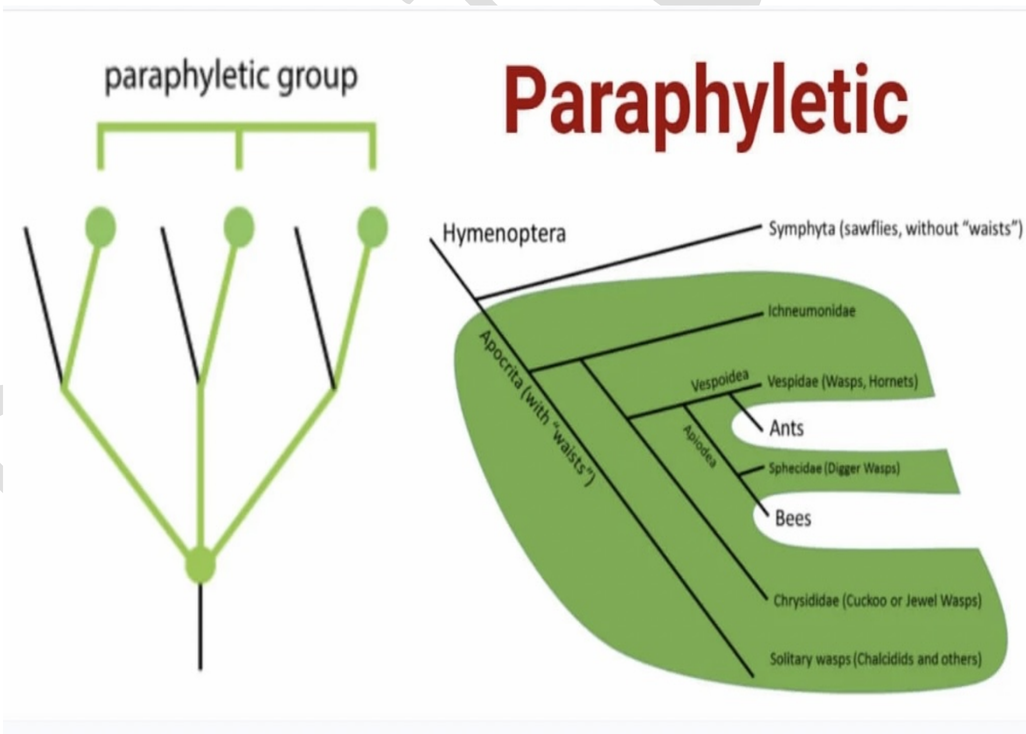
The word "polyphyletic" comes from the Greeks word "poly," meaning "many," and "phylon," meaning "race" or "tribe." The suffix "-etic"is used to indicate that something belongs to or is characteristic of a certain group or category. Together, polyphyletic means "belonging to many tribes" or "descended from many ancestors."

Polyphyletic groups are also known as **"artificial groups" /"convenience groups"** because they are based on superficial similarities rather than evolutionary relationships. This type of classification can be confusing and is not considered to be as useful as monophyletic or paraphyletic.

Recognizing polyphyletic groups help refine our understanding of biodiversity and provide insights into the process of evolution.




Paraphyly(Prosimii)
Monophyly(Simiiformes)
Lemurs
Lorises
Tarsiers
Polyphyly
New world monkeys
Old world monkeys
Apes
Humans


paraphyletic group
Paraphyletic
Hymenoptera
Symphyta (sawflies, without "waists")
Apocrita (with "waists")
Ichneumonidae
Vespoidea
Vespidae (Wasps, Hornets)
Ants
Apiodea
Sphecidae (Digger Wasps)
Bees
Chrysididae (Cuckoo or Jewel Wasps)
Solitary wasps (Chalcidids and others)

HOMOLOGY

Homology is the similarity in structure between different organisms based on their descent from a common, evolutionary ancestor.

[*Homology is a relationship defined between structures of DNA derived from a common ancestor and illustrates descent from a common ancestor. Analogous structures are physically (but not genetically) similar structures that were not present the last common ancestor*].

Homology (common ancestry and similar structure) can be reliably inferred from statistically significant similarity in a BLAST, FASTA, SSEARCH, or HMMER search, but to infer that two proteins are homologous does not that every part of one protein has a homolog in the other).

A homologous trait is often called a homolog (also spelled homologue). In genetics, the term "homolog" is used both to refer to a homologous protein and to the gene (DNA sequence/The sequential correspondence in one nucleic acid molecule) encoding it. As with anatomical structures, homology between protein or DNA sequences is defined in terms of ***shared ancestry***.

As far as Chromosomes are concerned homology is possession by two or more species of a character state derived, with or without modification, from their common ancestor. Homologous chromosomes are those members of a chromosome complement that bear the same genes.

At least 1.7 million species of living organisms have been discovered, and the list grows longer every year. How are they to be classified? Ideally, classification should be based on **homology**; that is, **shared characteristics that have been inherited from a common ancestor.**

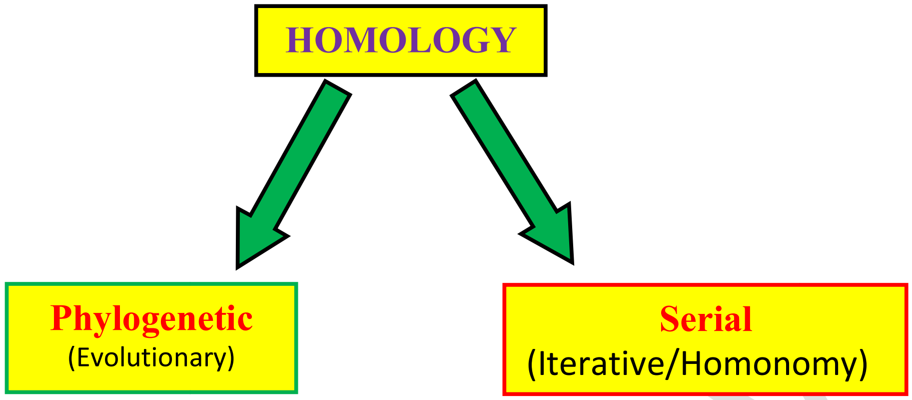
The more recently two species have shared a common ancestor, the more homologies they share and the more similar these homologies are.

Until recent decades, the study of homologies was limited to anatomical structures and pattern of embryonic development. However, since the birth of molecular biology, homologies can now also be studied at the level of proteins and DNA.

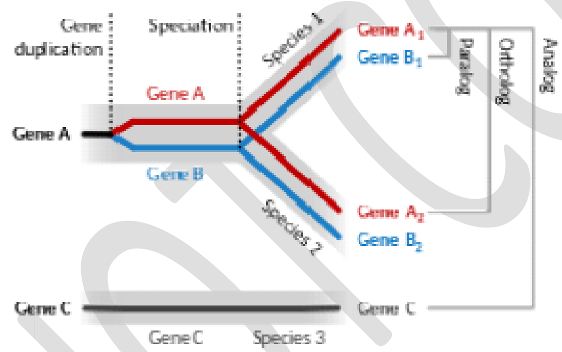
Usually two types of homology are considered,

(I). Phylogenetic/Evolutionary, between species, and

(II). Serial/Iterative/Homonomy, within individuals.



Homology among DNA, RNA, or proteins is *typically inferred from their nucleotide or amino acid sequence similarity*. **Significant similarity is strong evidence that two sequences are related by evolutionary changes from a common ancestral sequence.**



DIFFERENCE BETWEEN SIMILARITY AND HOMOLOGY

Similarity: Degree of likeness between two sequences, usually expressed as a percentage of similar (or identical) residues over a given length of the alignment. Can usually be easily calculated.

Homology: Statement about common evolutionary ancestry of two sequences. Can only be true or false.

PROTEIN HOMOLOGY IN TAXONOMY

Protein homology is often used in taxonomy to assess evolutionary relationships between organisms.

By comparing similarities in protein sequences, taxonomists can infer common ancestry and categorize organisms into taxonomic groups based on their molecular similarities.

Protein homology, assessed through molecular data, is crucial for understanding evolutionary relationships among organisms. Similarity in protein sequences indicate shared ancestry, aiding in taxonomic classification.
High homology suggests a close evolutionary relationship, guiding the placement of species within a taxonomic framework.

Homologous recombination allows for the regulated exchange of genetic information between two different DNA molecules of identical or nearly identical sequence composition, and is a major pathway for the repair of double-stranded DNA breaks. A key facet of homologous recombination is the ability of recombination proteins to perfectly align the damaged DNA with homologous sequence located elsewhere in the genome. This reaction is referred to as the *homology search* and is akin to the target searches conducted by many different DNA-binding proteins.

Homologous Recombination

Homologous Recombination enables the exchange of genetic information between different DNA molecules and is a major driving force in evolution. Homologous recombination contributes to double-strand DNA break (DSB) repair, the rescue of stalled or collapsed replication forks, programmed and aberrant chromosomal rearrangements, horizontal gene transfer, and meiosis.

Understanding Protein Homology

Protein homology refers to the similarity between the **sequences, structures,** and **functions** of two or more proteins. This similarity can arise from a shared ancestry or common evolutionary history, which implies that these proteins have evolved from a common ancestral protein.

The different aspects of protein homology, including its types, methods for determining homology, and its significance in various scientific fields are as below.

Types of Protein Homology

There are two main types of protein homology:

A. **Sequence homology,** and
B. **Structural homology**

A. **Sequence homology**(Sequence homology is an indication of the geneticrelatedness of different organisms and gene function)**:**

This refers to the similarity in the amino acid sequences of two or more proteins. Sequence homology can be further classified into:

a. **Primary homology,**
b. **Secondaryhomology**,
c. **Tertiary homology**

} **All these are based on the level of similarity in the sequences**

- **Primary homology indicates an identical sequence,**
- **Secondary homology refers to the similarity in the arrangement of secondary structural elements such as alpha-helices and beta-strands,** and
- **Tertiary homology indicates a similar three-dimensional folding pattern.**

B. Structural homology: This refers to the similarity in the *three-dimensional structures* of two or more proteins. It is usually inferred from sequence homology, as proteins with similar sequences are likely to have similar structures. Structural homology,based on the level of similarity in the three-dimensional structurescan be further classified into:

a. **Fold homology**indicates a similar overall folding pattern,
b. **Superfamily homology,**refers to a similar arrangement of secondary structural elements,and
c. **Domain homology** indicates the presence of similar functional units within a protein.

Determining Protein Homology

Determining protein homology involves *comparing the sequences, structures, and functions of the proteins in question.* Several methods are used for this purpose, including:

Sequence alignment: This method involves *comparing the amino acid sequences* of two or more proteins to identify regions of similarity. The most common method for sequence alignment is the Basic Local Alignment Search Tool **(BLAST)**.

Structure alignment: This method involves *comparing the three-dimensional structures of two or more proteins* to identify regions of similarity. The most common method for structure alignment is the **Dali server,** which uses a fast and efficient algorithm to align protein structures based on their spatial arrangement.

Functional similarity: This method involves *comparing the biological functions* of two or more proteins to determine if they have similar roles in the cell. This can be done through the analysis of **gene ontology annotations**, **literature reviews, and other functional data sources.**

Significance of Protein Homology

Protein homology has numerous applications and significance in various scientific fields, including:

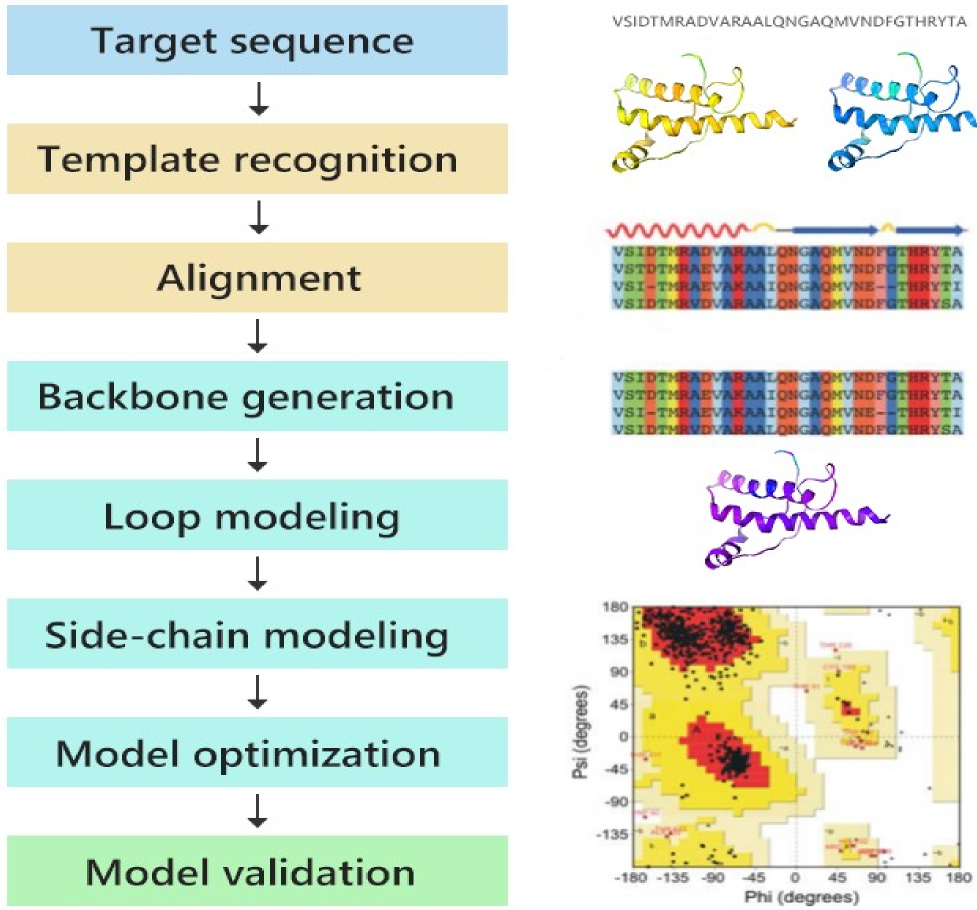
Evolutionary studies: Homology provides insights into the evolutionary relationships between species and helps in understanding the evolutionary history of proteins.

Drug discovery: Homology-based drug discovery involves identifying proteins with similar functions or structures and targeting them with drugs to treat a specific disease or condition.

Molecular modelling and protein engineering: Homology can be used to model the three-dimensional structure of a protein based on the structures of homologous proteins, which can be useful for predicting protein function and designing new proteins with desired properties.

Genetic studies: Homology is used in genetic studies to identify genes with similar functions, which can help in understanding gene regulation, gene expression, and genetic disorders.

In conclusion, protein homology is an essential concept in understanding the evolution, function, and structure of proteins.*By comparing the sequences, structures, and functions of proteins, we can gain insights into the evolutionary relationships between species, develop new drugs, and advance our understanding of molecular biology.*



NUCLEIC ACID HOMOLOGY (NAH)

Nucleic acid homology refers to the similarity between the nucleotide sequences of DNA or RNA from different organisms. This similarity can provide insights into evolutionary relationships, genetic relatedness, and functional similarities between species. *The comparison of nucleic acid sequences is a fundamental tool in molecular biology and genetics, enabling researchers to study the relatedness of genes and genomes across different organisms.*

Importance of Nucleic Acid Homology

Nucleic acid homology is crucial for understanding the evolutionary history of organisms and the genetic basis of traits. By comparing nucleic acid sequences, scientists can infer common ancestry and identify conserved regions that are essential for specific biological functions. This information is valuable for various fields, including evolutionary biology, phylogenetics, and biomedical research.

Methods for Analyzing Nucleic Acid Homology

Several methods are used to analyze nucleic acid homology, including sequence alignment, phylogenetic analysis, and hybridization techniques.

(i). *Sequence alignment* involves comparing nucleotide sequences to identify regions of similarity and divergence.

(ii). *Phylogenetic analysis*uses these sequence similarities to construct evolutionary trees and infer relationships between species.

(iii). *Hybridization techniques*, such as DNA-DNA hybridization and Southern blotting, allow for the direct comparison of nucleic acid sequences between different organisms.

Applications of Nucleic Acid Homology

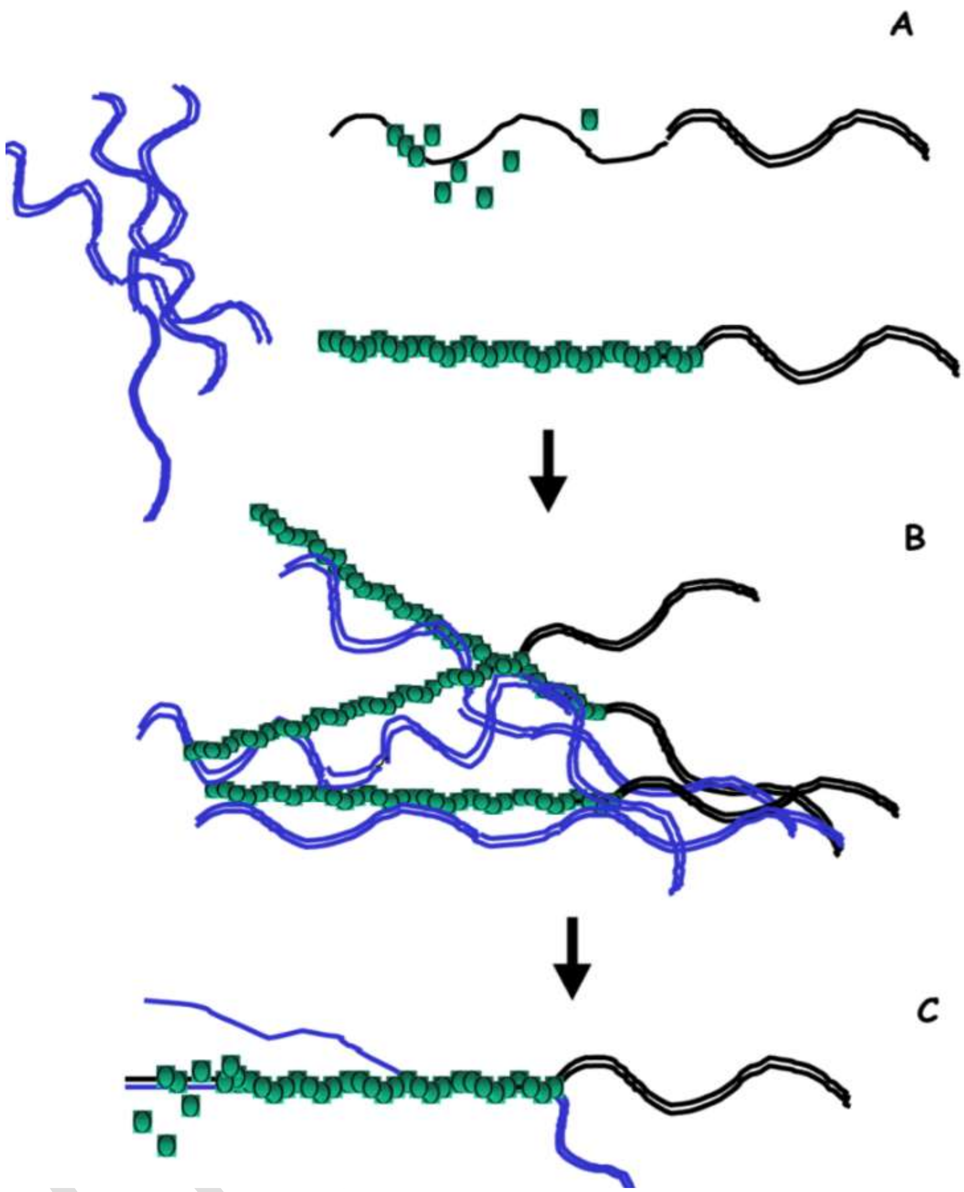
Nucleic acid homology has diverse applications in research and practical fields. In evolutionary biology, it helps in reconstructing the evolutionary history of species and understanding genetic diversity. In genetics, it aids in identifying genes associated with specific traits or diseases by comparing sequences across different individuals or populations. In biotechnology, it is used to design probes for gene expression studies, identify genetic markers, and assess genetic relatedness in breeding programs.

Challenges in Nucleic Acid Homology Analysis

Despite its significance, analyzing nucleic acid homology comes with challenges. The vast amount of genomic data available requires sophisticated computational tools for sequence alignment and phylogenetic analysis. Additionally, factors such as gene duplication, horizontal gene transfer, and genome rearrangements can complicate the interpretation of homologous sequences.

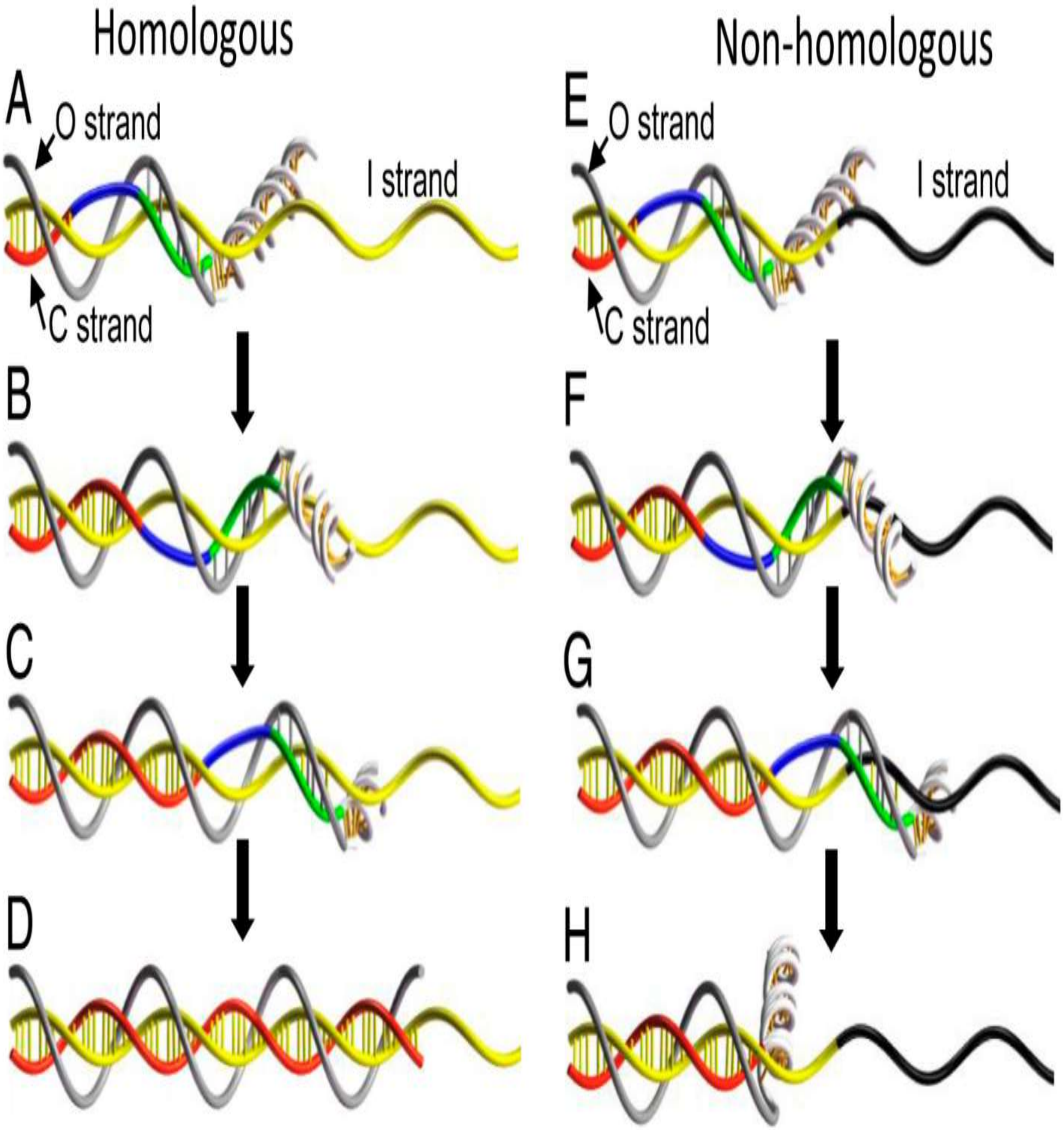
Conclusion

Nucleic acid homology provides a powerful means to study genetic relationships, evolutionary history, and functional genomics across diverse organisms. Its applications span from basic research to applied fields such as medicine and agriculture, making it an indispensable tool in modern biological sciences.



An overview of the early steps of homologous Recombination promoted by RecA protein.

A. Formation of the nucleoprotein filaments.
B. Search for homology.
C. Homologous pairing and strand exchange.



***Stepwise progression in Homologous Recombination* (HR).**

A-D: HR with fully homologous DNA. The C strand is divided into three parts.

(i). Post-strand exchanged (red),

(ii). homology testing (green), and

(iii). Transitional connection (blue).

E-H:HR blocked by the nonhomologous segment.The homologous I strand is coloured yellow, and the nonhomologous I strand is coloured black.

DECLARATION

THANX