

Unit - 4

Molecular Evolution

- Molecular Evolution is the process of change in the sequence composition of cellular molecules such as DNA, RNA and protein across generations.
- Molecular evolution uses principles of evolutionary biology and population genetics to explain patterns in these changes.

History of molecular evolution

- It started early 20th century with "Comparative biochemistry"
- molecular evolution came - 1960s & 1970s.
- Advent of protein sequencing allowed molecular biologists to create phylogenies based on sequence comparison and to use the differences b/w homologous sequences as a molecular clock to estimate the time since the

last common ancestor

→ In the late 1960s, the Neutral Theory of molecular evolution provided a theoretical basis for the molecular clock.

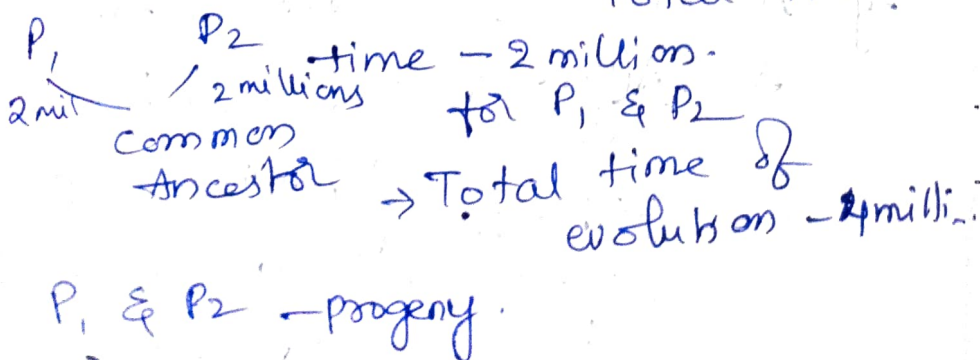
→ After 1970s, nucleic acid sequencing allowed molecular evolution to reach beyond proteins to highly conserved ribosomal RNA sequences, the foundation of a reconceptualization of early history of life.

Rate of molecular evolution

- The rate of evolution is estimated from the amino acid sequences of a protein & nucleotide sequences in a region of DNA in two or more species

$$\text{rate of molecular evolution} = \frac{\text{No of amino acids} + \text{No of nucleotide substitutions}}{\text{Total time}}$$

eg:- Humans



phylogenetic trees

Scaled
Trees

Branched lengths
are equal to
the magnitude
of change
in the nodes.

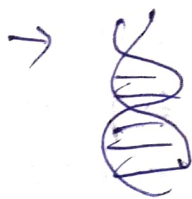
unscaled Trees

— only representing
the relationship
B/w sequences



cellular memory
unit

→ Translated
product
(protein)



DNA

→ modified

mutation
Substitution
Insertion
Deletion

molecular
evolution

→ Mutation: Sudden change in DNA or RNA

→ Substitution: when DNA sequences diverge, they begin to

eg: A G T C - seq ① collect mutations.

A G A (G) - seq ②

↓
Substitution
c-replaced
by G.

The number of
Substitutions (P) found
in an alignment is
widely used in molecular
evolution analysis.

→ Insertion :- Addition of nucleotide or addition of aminoacid.

→ Deletion :- Removing of nucleotide chain or removing of aminoacid.

→ Neutral mutation :- mutation has no fitness effects.

Neutral theory

→ Motoo Kimura, Thomas Jukes
(1968, 69)

→ concept of molecular clock

→ Most of the aminoacid changes & polymorphisms are neutral and created by genetic drift & mutation.

evidences of neutral evolution

- ① Pseudogenes (dead genes that have no function and no fitness effect) evolve very fast.
- ② Genes that have important functions should evolve slower.
- ③ Synonymous codon positions (3-fold, 4 fold degenerate sites) evolve faster than non-synonymous sites should evolve with constant rate.

phylogenetic tree :-

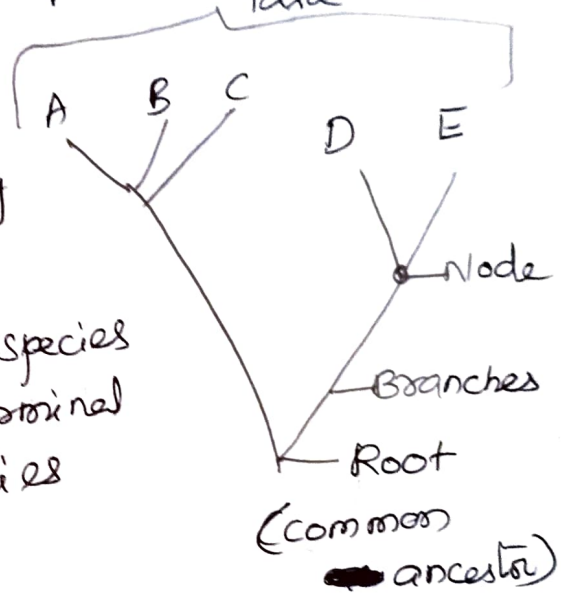
phylogenetic tree is a diagrammatic representation of evolutionary relationships among living organisms

→ It represents how different species evolved from a series of common ancestors

→ The phylogenetic tree is also called the "Tree of Life" or "Dendrogram"

Node = ancestor of present day species

Taxa = present day species
~~off~~ terminal species



Types of phylogenetic trees

① Rooted tree

② Unrooted tree

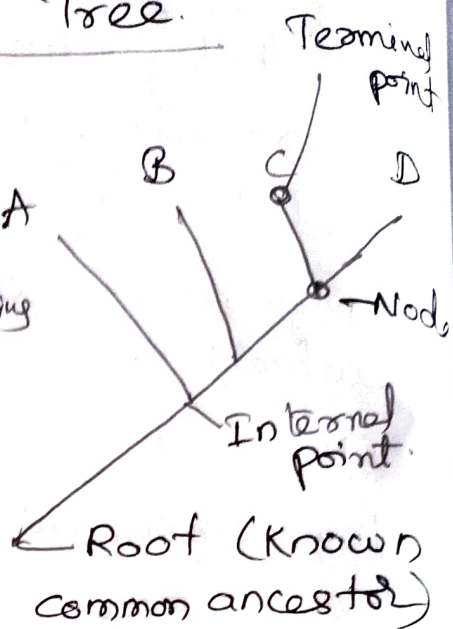
① Bifurcating tree ② multifurcating trees

③ Bifurcating tree ④ multifurcating trees

① Rooted phylogenetic Tree.

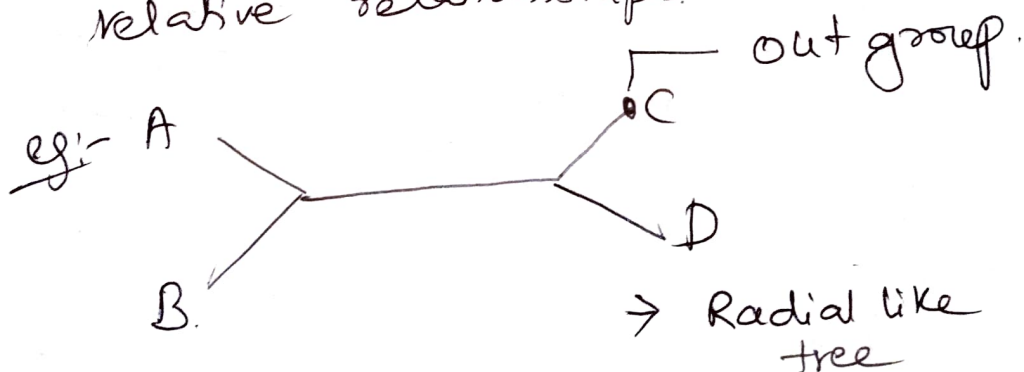
→ It is directed tree with a unique node A

- the root - corresponding to the most recent common ancestor of all the entities at the leaves of tree.



② Unrooted phylogenetic tree.

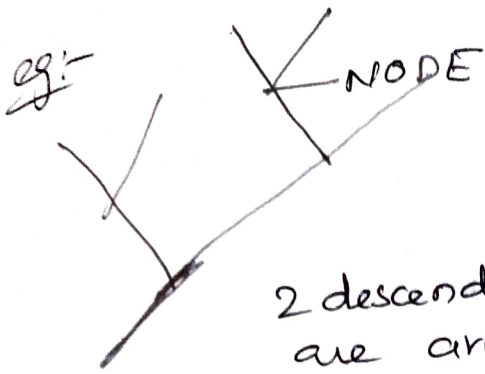
- An unrooted phylogenetic tree does not give the information of a common ancestor, but only positions the taxa to show their relative relationships.



a) Bifurcating phylogenetic Tree.

- A bifurcating tree has exactly two descendants arising from each interior node. Both rooted and

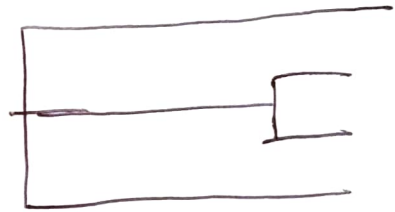
unrooted trees can be bifurcating.



2 descending
are arising - bifurcating

b) multifurcating phylogenetic tree

- multifurcating tree has multiple descendents arising from each of the interior node. Both rooted and unrooted trees can be multifurcating



Importance of phylogenetic tree.

- To study evolutionary histories
→ How the species were spread geographically.

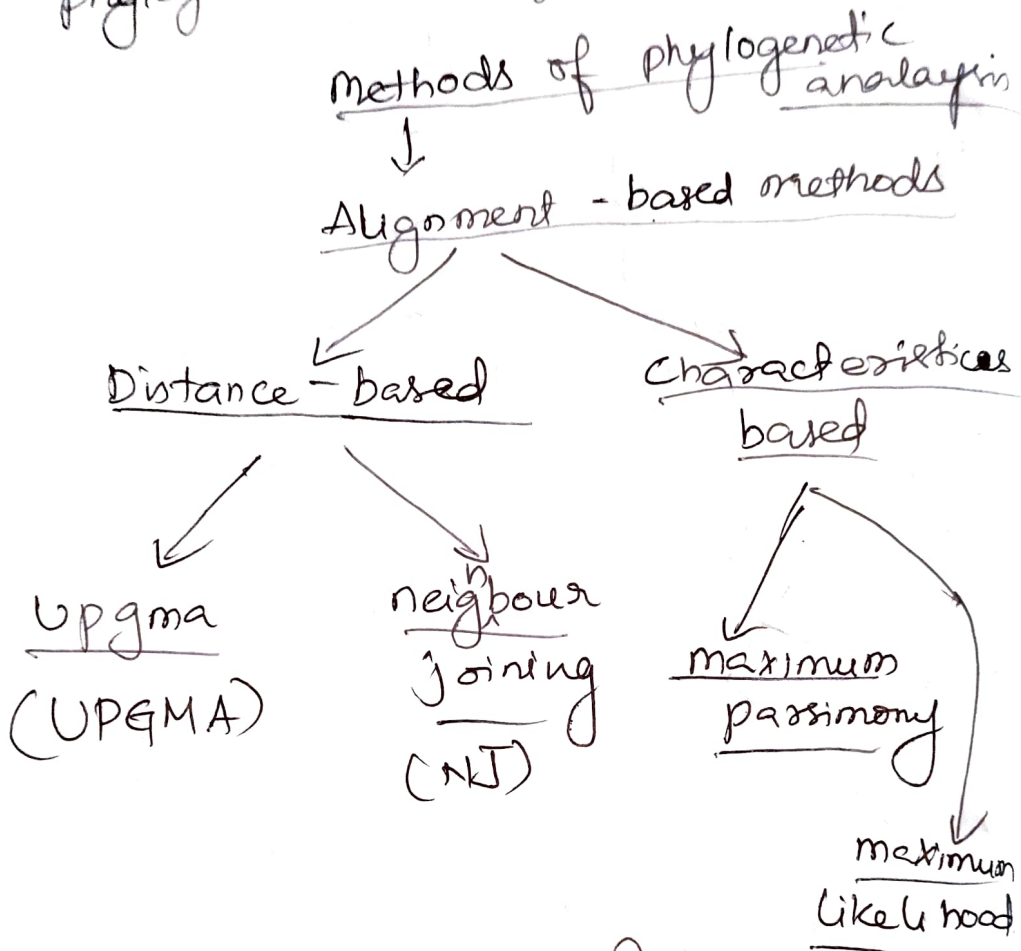
→ Search for a new species

→ To identify the most recent common ancestor.

→ microbes can be traced along with their evolutionary histories.

Methods of phylogenetic Analysis

→ Alignment - Based methods of phylogenetic analysis.



→ Distance - based methods of phylogenetic

→ Use the dissimilarity (the distance) b/w two sequences to construct trees.

→ They are much less computationally intensive than the character based methods are mostly accurate as they take mutations into count

→ for tree generation, generally ~~to~~ hierarchical clustering

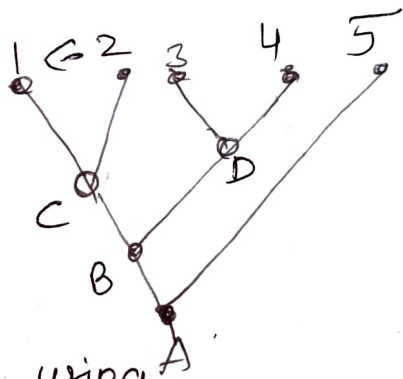
is used in which dendrograms (clusters) are created.

a) UPGMA :

(Unweighted Pair Group Method with Arithmetic Mean)

- A simple clustering method that assumes a constant rate of evolution (molecular clock hypothesis).
- It needs a distance matrix of the analysed taxa that can be calculated from a multiple alignment.

eg:-



Steps involved to construct a phylogenetic tree using UPGMA :

- Align & Name them
- Compare sequence them using pair-wise sequence alignment
- Count the mismatches and record them in the table
- Complete the table by comparing all the sequences.
- look for the lowest value to find the first group.

→ By finding the arithmetic mean step by step, and construct the tree.

Example:-

A - ATCGATCG
 B - GTA GACGA
 C - ACCGTACG
 D - TCA GTCAC
 E - CCTACAG

→ Pair-wise sequence alignment by mis match

① A - ATCGATCG mismatch - 5
 B - GTA GACGA sequences

② A - ATCGATCG
 C - ACCGTACG
 → 3

	A	B	C	D	E
A		5	3	6	3
B			7	5	6
C				4	5
D					4
E					

Find lowest value - ③

A
 └─ C.

1st group.

	A/C	B	D	E
A/C		6	5	5
B			5	6
D				4
E				

lowest value

$$\textcircled{1} \frac{(A-B) + (C-B)}{2} = \frac{5+7}{2} = \frac{12}{2} = 6$$

$$\textcircled{2} \frac{(A-D) + (C-D)}{2} = \frac{6+4}{2} = \frac{10}{2} = 5$$

Average 1st group

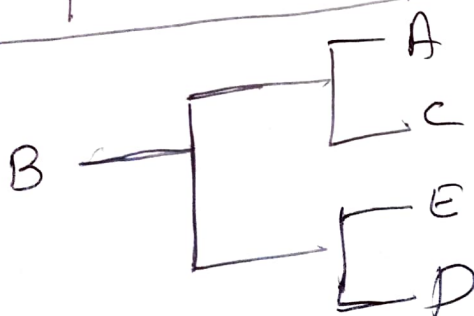
second group

	A/C	D/E	B
A/C		5	6
D/E			5.5
B			

lowest value

$$\frac{A}{B} = \frac{[A/C-E] + [A/C-D]}{2}$$

$$= \frac{5+5}{2} = \frac{10}{2} = 5$$



Phylogenetic Tree

$$D/E \& B = \frac{D-B + E-B}{2} = \frac{5+6}{2} = 5.5$$

b) Neighbour joining (NJ)

- bottom-up clustering method that also needs a distance matrix
- NJ is a heuristic approach that does not guarantee to find the perfect result, but under normal conditions has a very high probability to do so. It has a very good computational efficiency making it well suited for large datasets.
 - constant rate can't be assumed
 - mutation rate - Base pair
 - Branch length is proportional to rate of evolution

2) characteristics based methods

- character based methods compare:
- All sequences simultaneously considering one character/site at a time
- These are maximum parsimony and maximum likelihood
- These methods use probability

and consider Variation in a set of sequences.

a) Maximum Parsimony (MP)

→ This method tries to create a phylogeny that requires the least evolutionary change.

→ It may suffer from long branch attraction, a problem that leads to incorrect trees in rapidly evolving lineages.

b) Maximum - Likelihood (ML)

→ ML uses a statistical approach to infer a phylogenetic tree.

→ ML is well suited for the analysis of distantly related sequences, but is computationally expensive and thus not that well suited for larger input data.