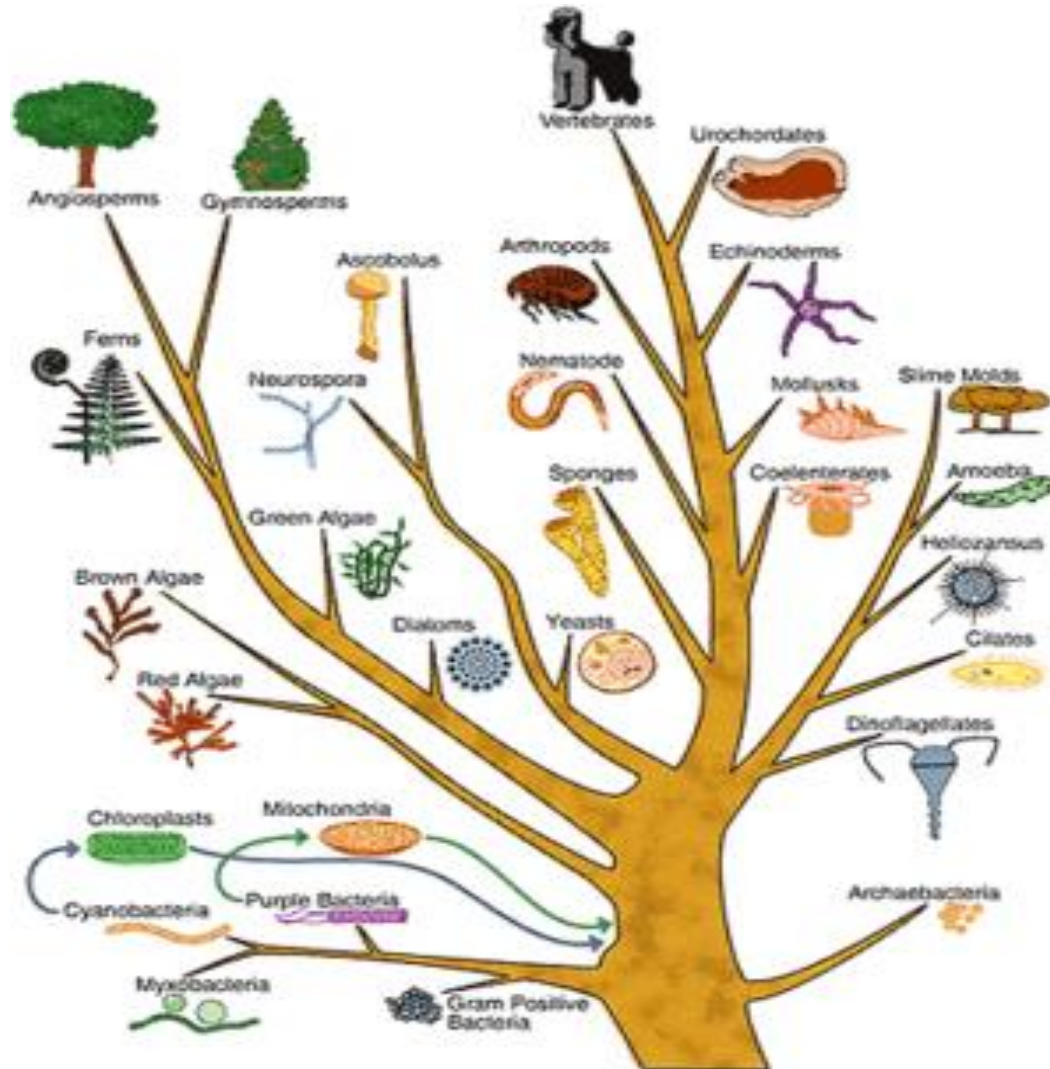


Phylogenetic trees

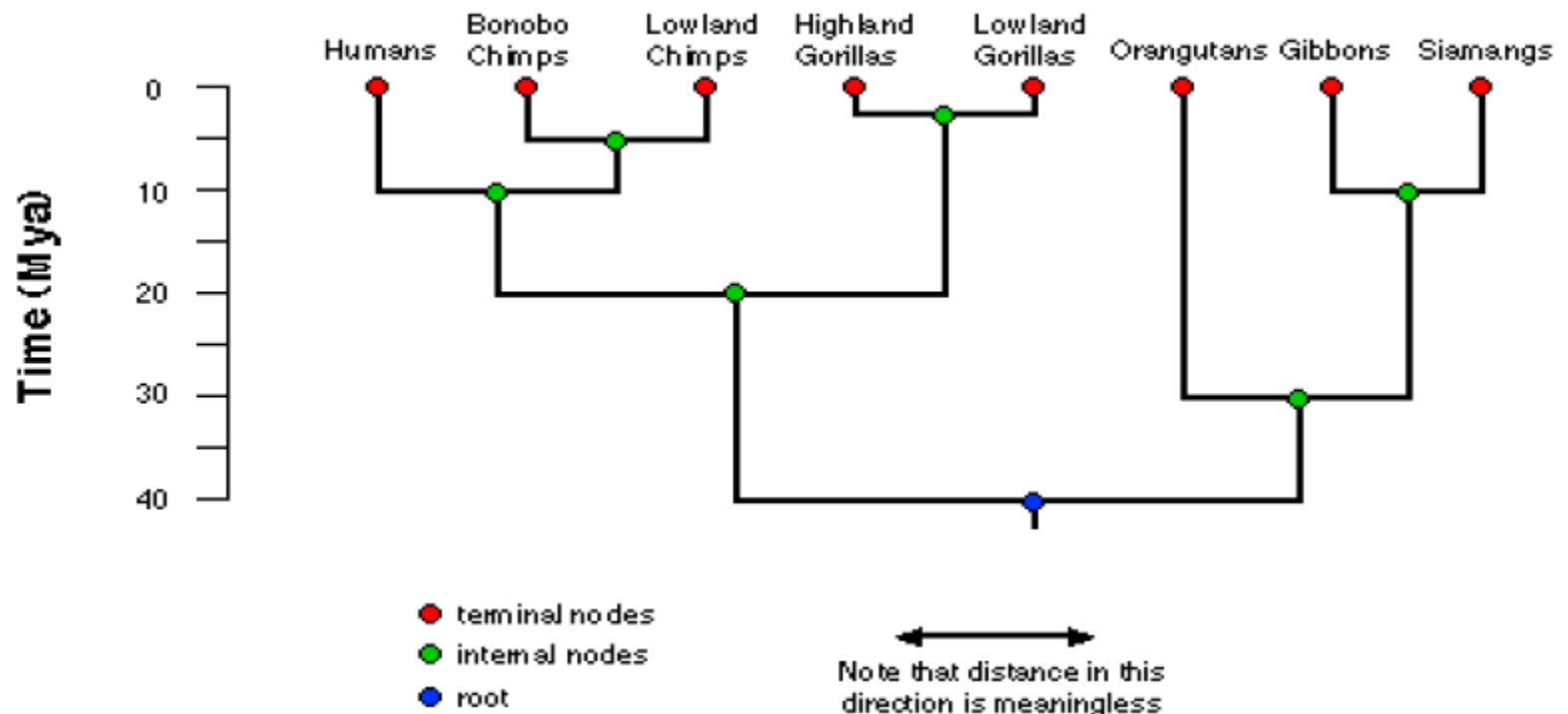


Phylogenetic tree

- The similarity of molecular mechanisms of the organisms that have been studied strongly suggests that all organisms on Earth had a common ancestor.
- Thus any set of species is related (evolutionary divergent), and this relationship is called a phylogeny.
- Usually the relationship can be represented by a phylogenetic tree.
- The task of phylogenetic is to infer this tree from observations upon the existing organisms.
- (Greek: phylon = race and genetic = birth)

Terminology

- Phylogenetic tree: Visual representation of evolutionary distances between species



Historical Note

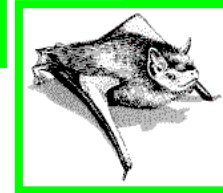
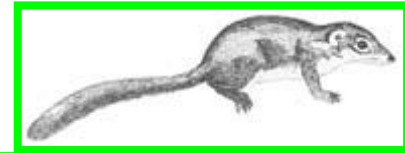
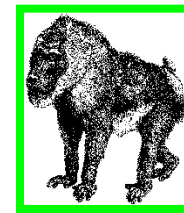
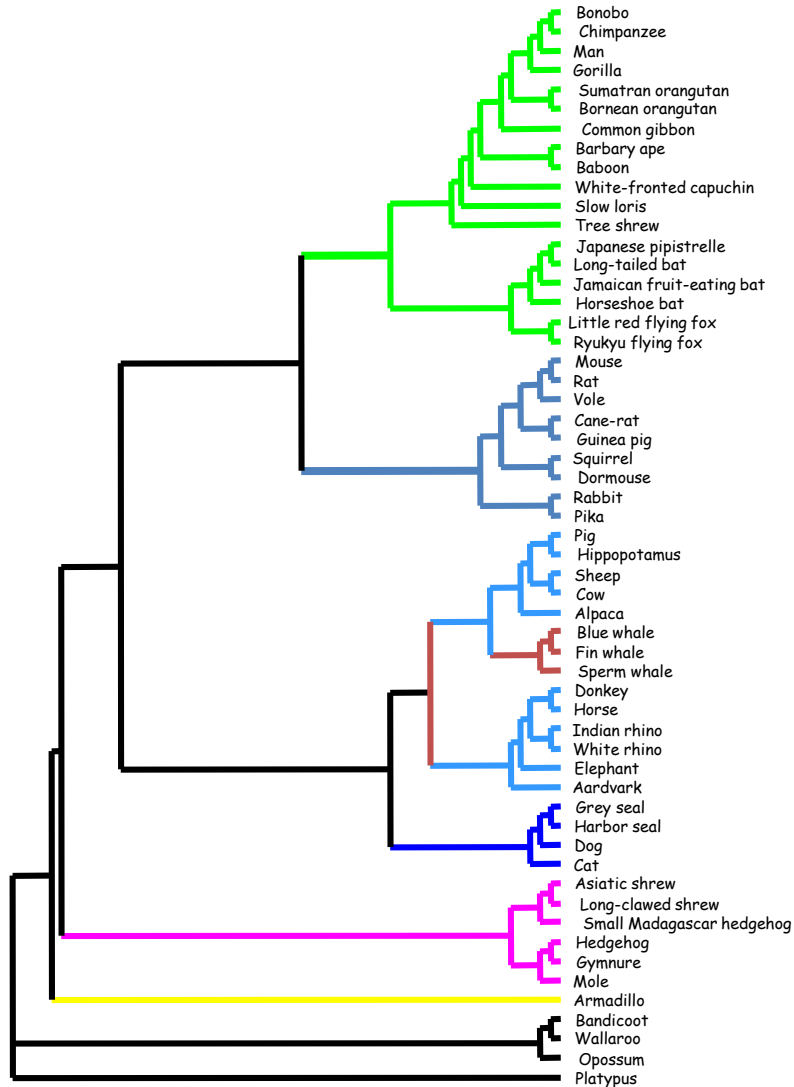
- Approaches
 - Fossil Records , Phylogenetic Trees
- Until mid 1950's phylogenies were constructed by experts based on their opinion (subjective criteria)
- Since then, focus on **objective** criteria for constructing phylogenetic trees
 - Thousands of articles in the last decades

Morphological vs. Molecular

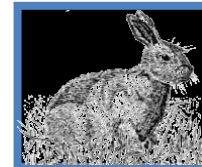
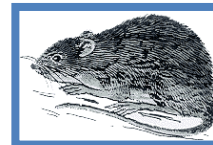
- Classical phylogenetic analysis: **morphological** features: number of legs, lengths of legs, etc.
- Modern biological methods allow to use **molecular** features
 - Gene sequences
 - Protein sequences
- Analysis based on homologous sequences (e.g., globins) in different species

Morphological topology

(Based on Mc Kenna and Bell, 1997)



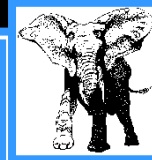
Archonta



Glires



Ungulata



Carnivora



Insectivora



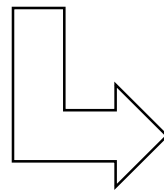
Xenarthra

Morphological limitations

- fossil records have many limitations
- they may be available only for certain species
- Existing fossil data can be fragmentary and their collection is often limited by abundance, habitat, geographic range, and other factors.
- For microorganisms, fossils are essentially nonexistent.

From sequences to a phylogenetic tree

Rat	QEPGGLVVPPTDA
Rabbit	QEPGGMVVPPTDA
Gorilla	QEPGGLVVPPTDA
Cat	REPGGLVVPPTTEG



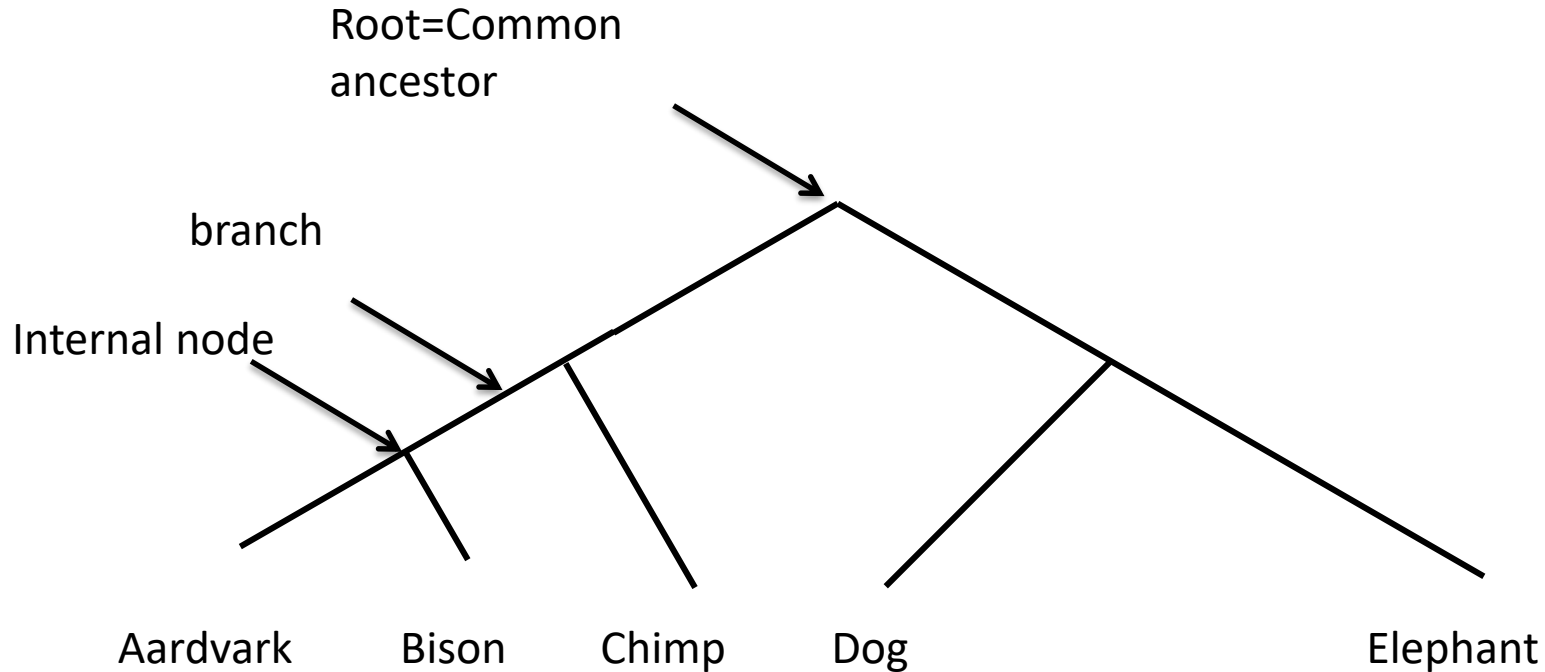
There are many possible types of sequences to use (e.g. Mitochondrial vs Nuclear proteins).



Basic Assumptions

- ◆ A universal ancestor exists for all life forms.
- ◆ Molecular difference in homologous genes (or protein sequences) are positively correlated with evolution time.
- ◆ each position in a sequence evolved independently.
- ◆ Phylogenetic relation can be expressed by a dendrogram (a “tree”) .

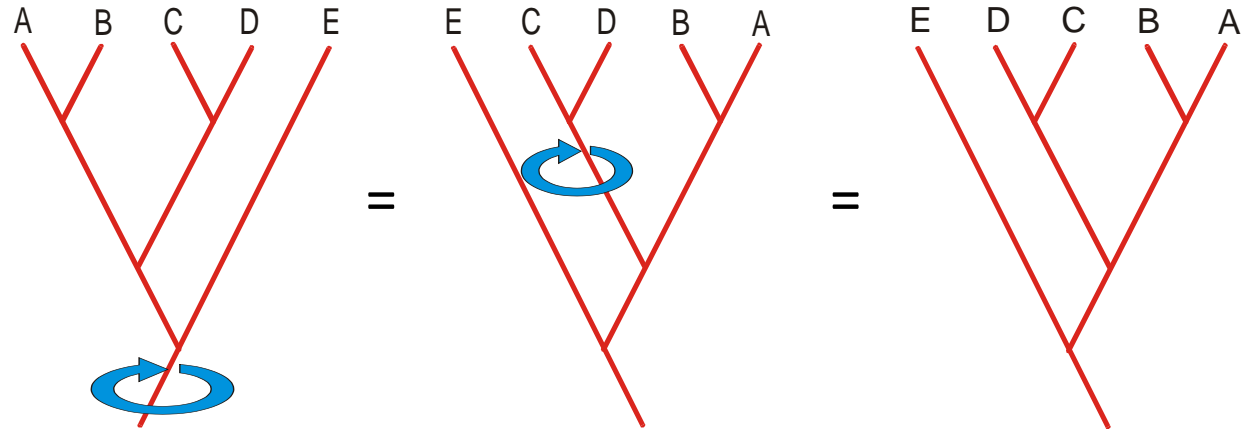
Phylogenetic trees



- **Leafs** - current day species(Taxa)
- **Nodes** - hypothetical most recent common ancestors
- **Edges length** - “time” from one speciation to the next

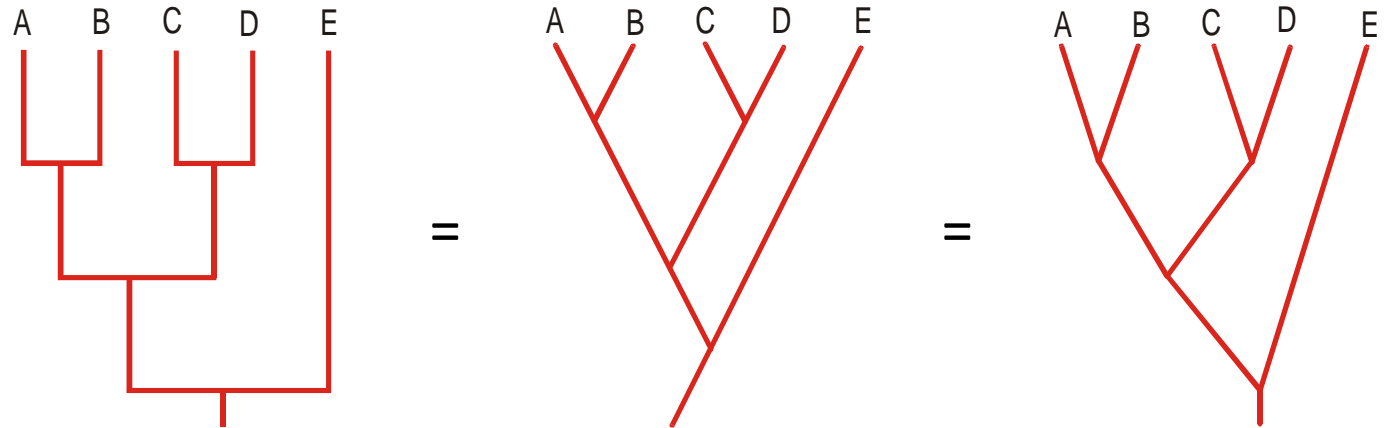
Phylogenetic trees

There are many ways of drawing a tree



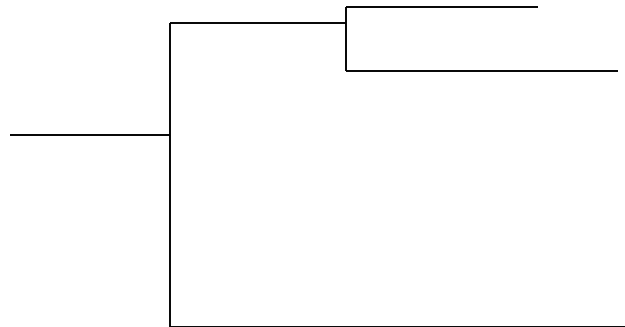
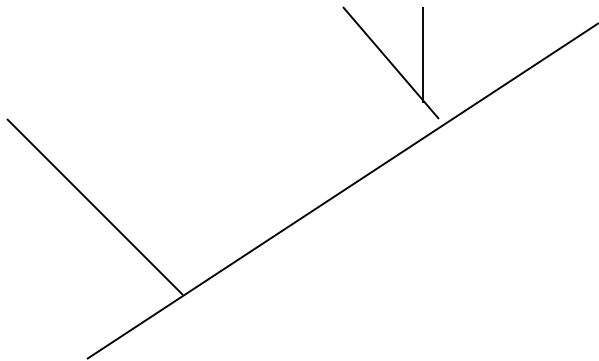
Phylogenetic trees

There are many ways of drawing a tree



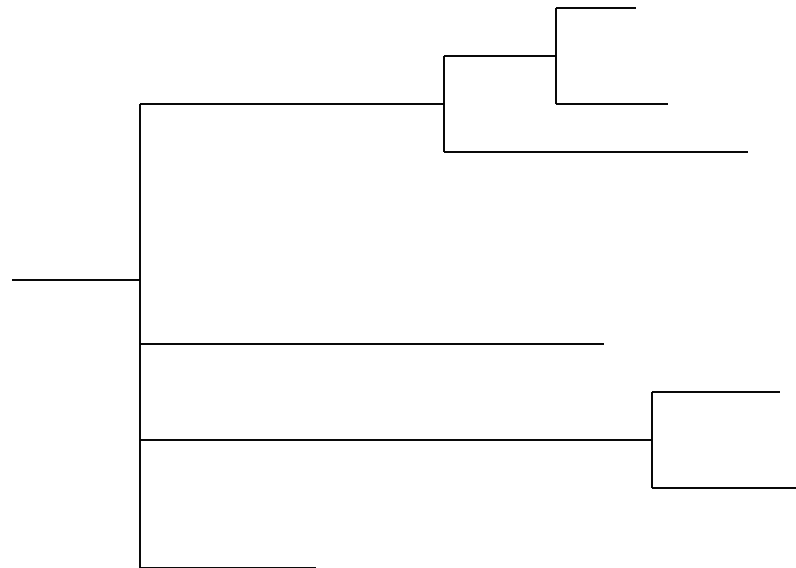
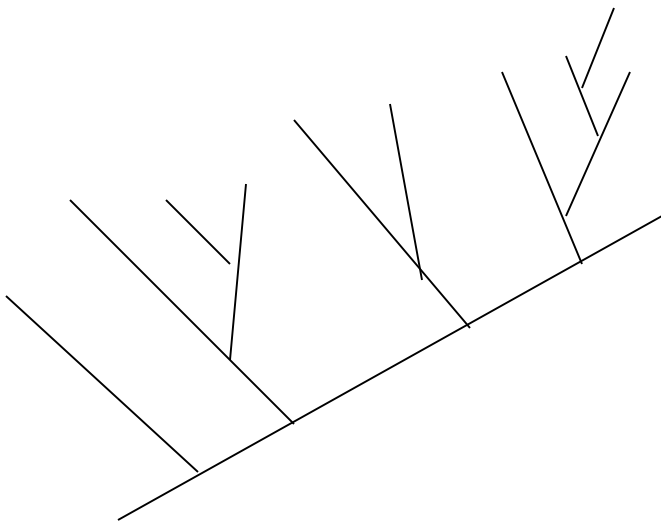
The bifurcating tree

- A tree that bifurcates has a maximum of 2 descendants arising from each of the interior nodes.



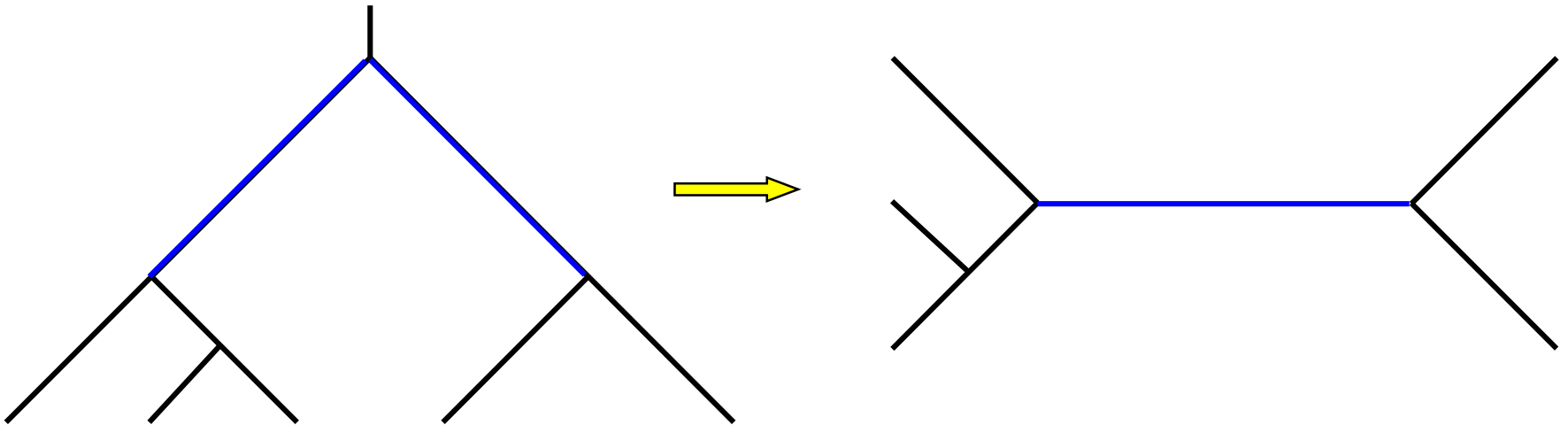
The multi-furcating tree

- A tree that multi-furcates has multiple descendants arising from each of the interior nodes.



Types of trees

- A natural model to consider is that of **rooted** trees
- **Unrooted** tree represents the same phylogeny without the root node

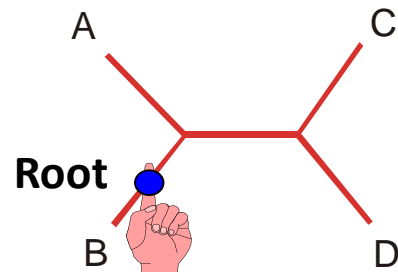
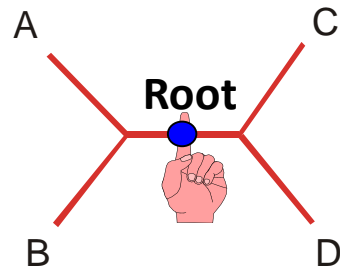


Depending on the model, data from current day species does not distinguish between different placements of the root.

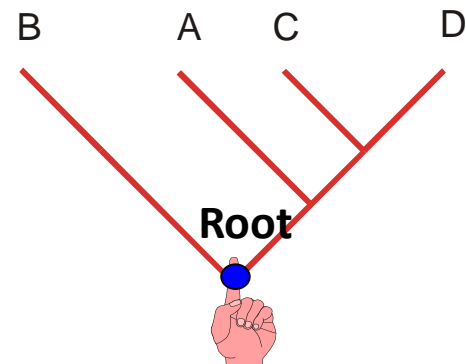
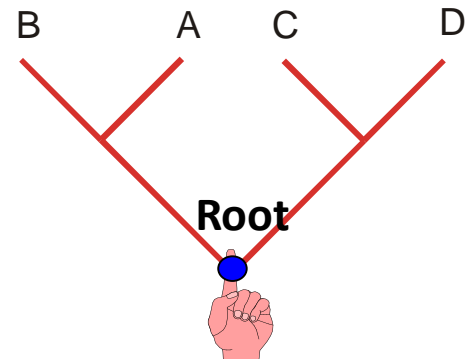
Phylogenetic trees

Trees can be unrooted or rooted

Unrooted tree



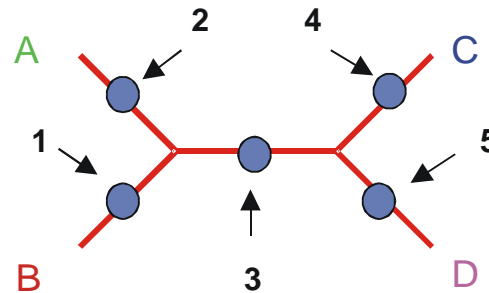
Rooted tree



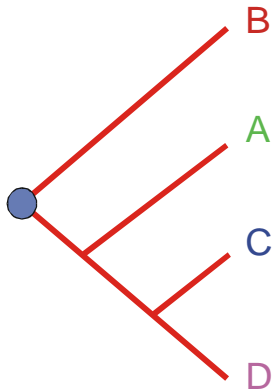
Phylogenetic trees

Trees can be unrooted or rooted

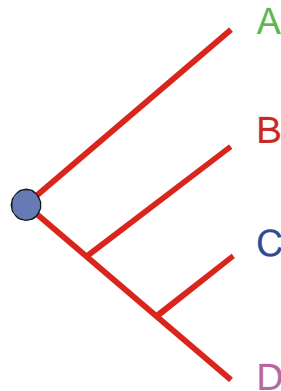
Unrooted tree



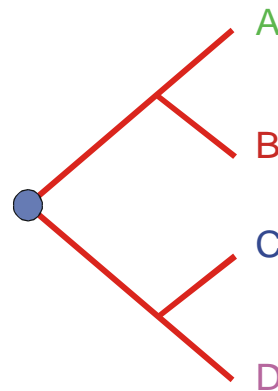
Rooted tree 1



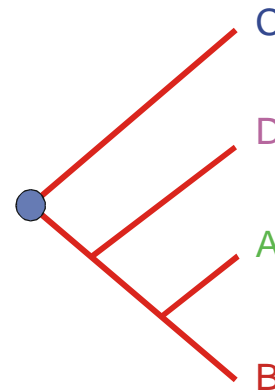
Rooted tree 2



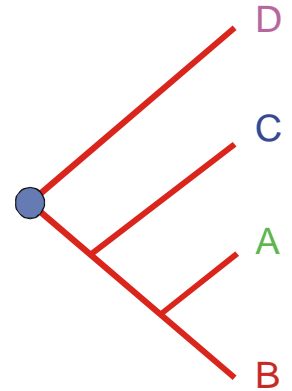
Rooted tree 3



Rooted tree 4



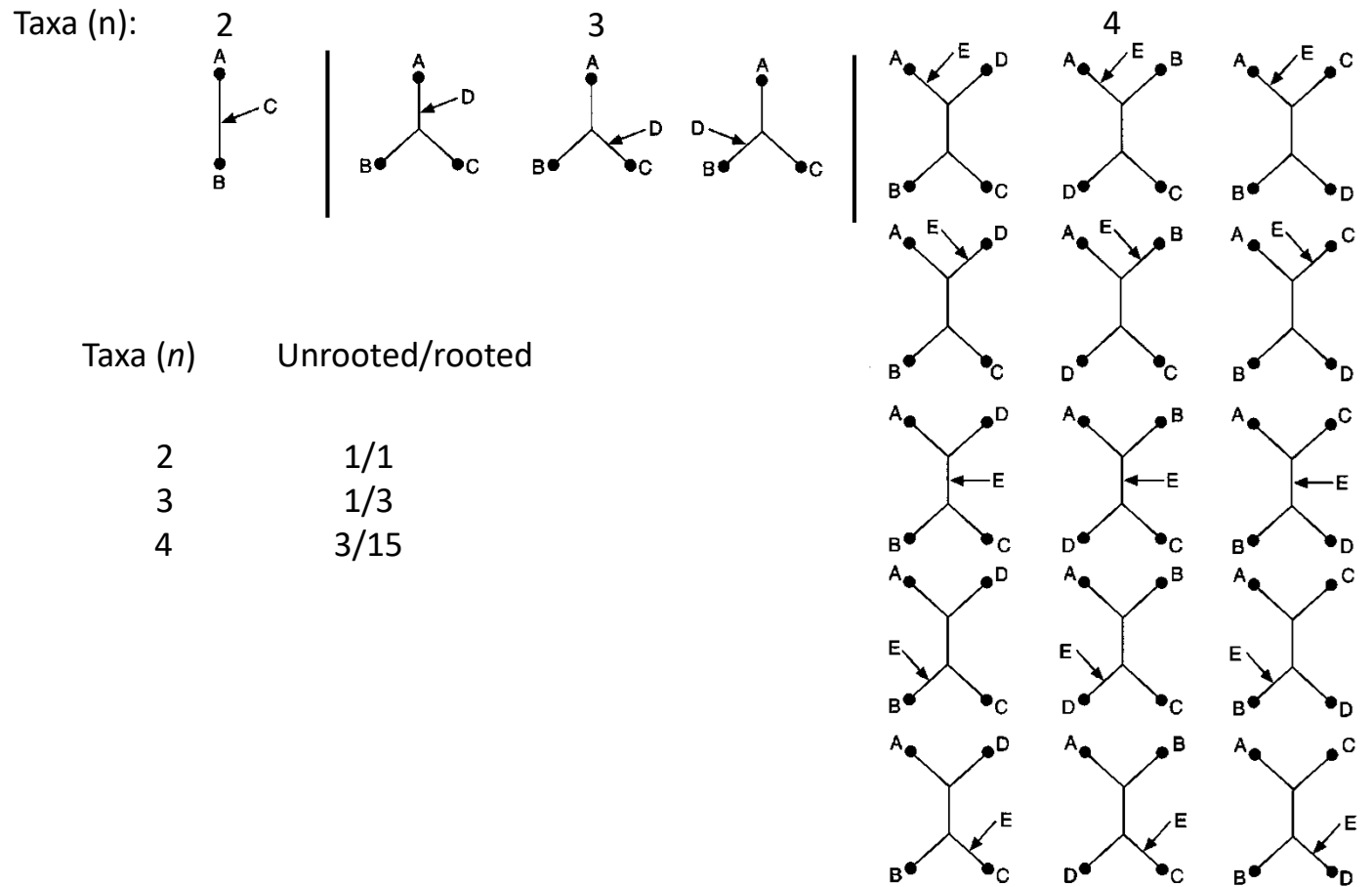
Rooted tree 5



These trees show five different evolutionary relationships among the taxa!

Phylogenetic trees

Possible evolutionary trees

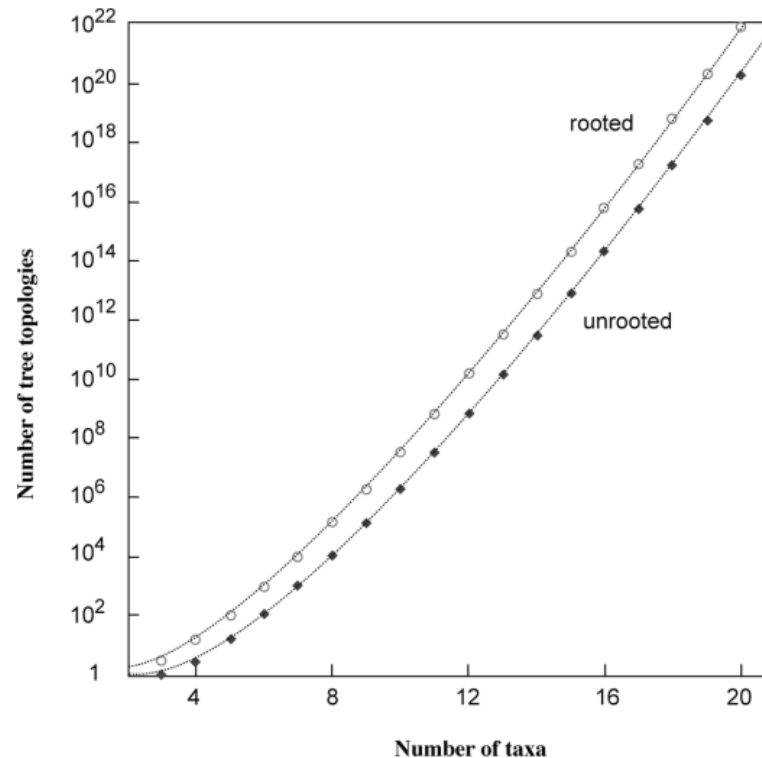


Phylogenetic trees

Possible evolutionary trees

Taxa (n)	rooted $(2n-3)!/(2^{n-2}(n-2)!)$	unrooted $(2n-5)!/(2^{n-3}(n-3)!)$
2	1	1
3	3	1
4	15	3
5	105	15
6	954	105
7	10,395	954
8	135,135	10,395
9	2,027,025	135,135
10	34,459,425	2,027,025

Number of Rooted VS Unrooted Trees



$$NR = (2n - 3)! / 2^{(n-2)} * (n - 2)!$$

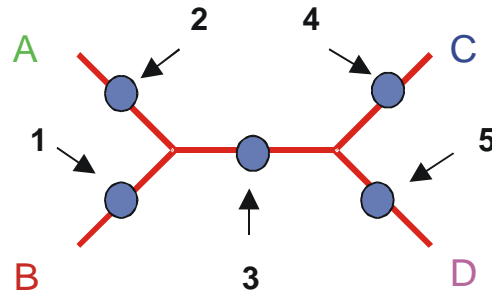
$$NU = (2n - 5)! / 2^{(n-3)} * (n - 3)!$$

But only one of these represents the true turn of events!

Most phylogenetic trees generated with molecular data are thus referred to as *inferred trees*.

Phylogenetic trees

How to root?

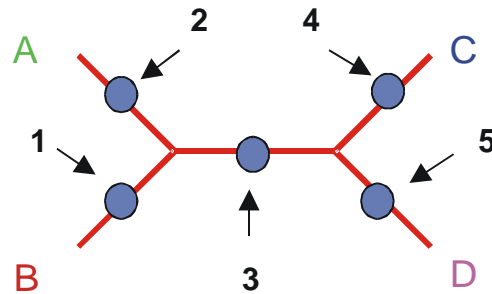


- ◆ Use information from ancestors

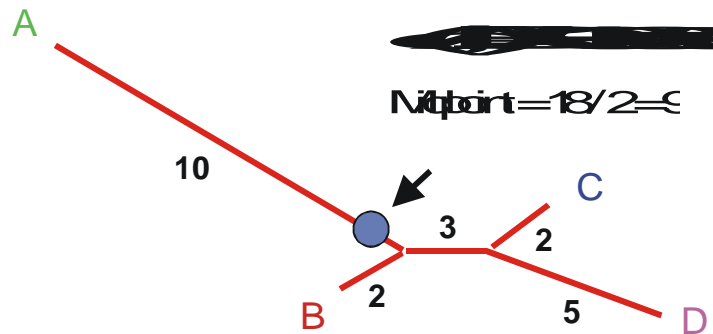
In most cases not available

Phylogenetic trees

How to root?



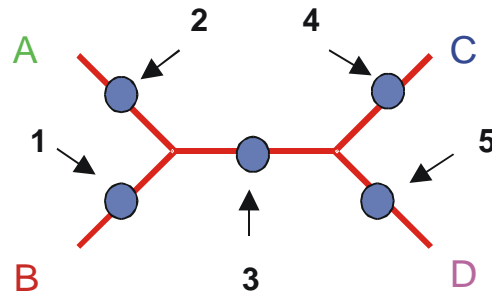
- ◆ Use statistical tools will root trees automatically (e.g. mid-point rooting)



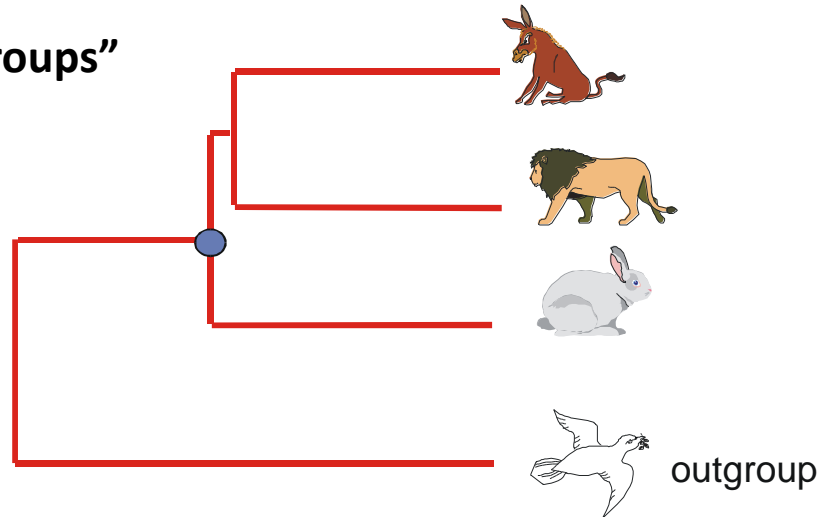
This must involve assumptions ... BEWARE!

Phylogenetic trees

How to root?



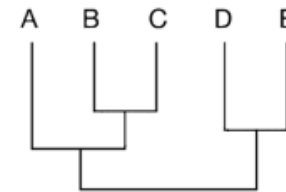
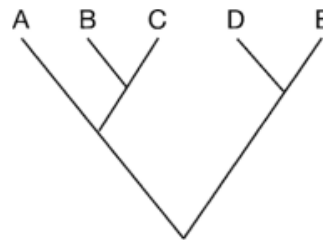
◆ Using “outgroups”



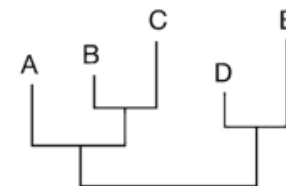
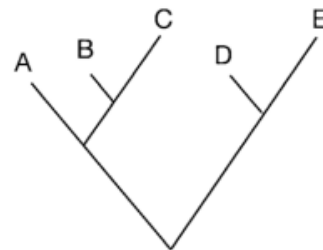
- the outgroup should be a taxon known to be less closely related to the rest of the taxa (ingroups)
- it should ideally be as closely related as possible to the rest of the taxa while still satisfying the above condition

Terminology

- Unrooted tree
- Rooted tree
 - Cladograms: Branch length have no meaning
 - Phylograms: Branch length represent evolutionary change

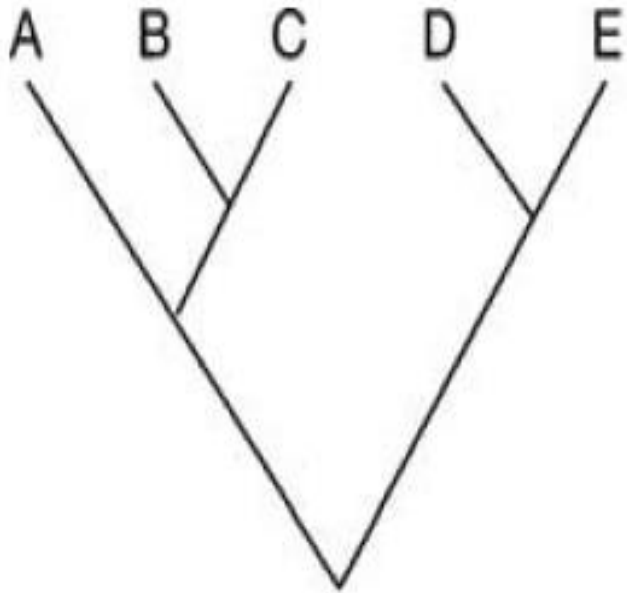


Cladogram

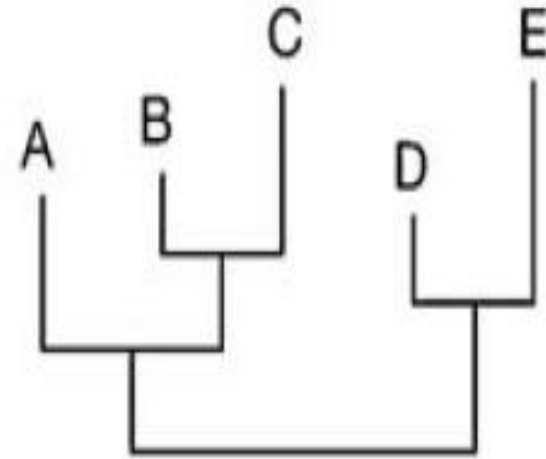


Phylogram

Formal Representation



`((B,C),A),(D,E))`



`((B:1,C:2),A:2),(D:1.2,E:2.5))`

Newick format

How to construct a phylogenetic tree?

- Step 1:

Choice of Molecular Markers

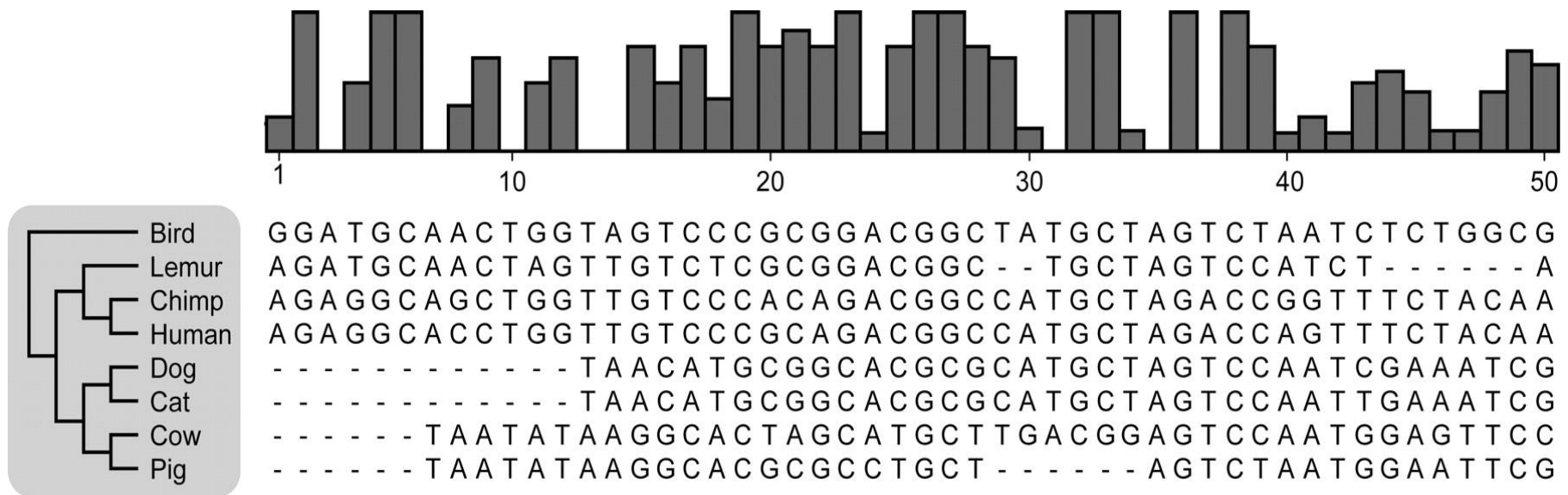
amino acid sequence

DNA sequence

RNA sequence



How to construct a phylogenetic tree?



□ Step 2:

Check the multiple alignment if it reflects the evolutionary process.

How to construct a phylogenetic tree?

cont

- Step3:

Choose what method we are going to use and calculate the distance or use the result depending on the method

- Step 4:

Verify the result statistically.

Where do we get distances?

- commonly obtained from sequence alignments

$$f_{ij} = \frac{\text{\#mismatches}}{\text{\#matches} + \text{\#mismatches}}$$

in alignment of sequence i with sequence j

$$\text{dist}(i, j) = f_{ij}$$

- to consider evolutionary time between sequences:

$$\text{dist}_{\text{Jukes-Cantor}}(i, j) = -\frac{3}{4} \ln \left(1 - \frac{4}{3} f_{ij} \right)$$

Distance Matrix methods

- ❑ Calculate all the distance between leaves (taxa)
- ❑ Based on the distance, construct a tree
- ❑ Not very accurate
- ❑ Fastest method
 - ❑ UPGMA
 - ❑ Neighbor-joining

Neighbor-joining method

- ❑ Developed in 1987 by Saitou and Nei
- ❑ Works in a similar fashion to UPGMA
- ❑ Still fast – works great for large dataset

How to construct a tree with Neighbor-joining method?

□ Step 1:

- ▣ Calculate sum all distance from x and divide by (leaves – 2)

- $S_x = (\text{sum all } D_x) / (\text{leaves} - 2)$

□ Step 2:

- ▣ Calculate pair with smallest M

- $M_{ij} = \text{Distance } ij - S_i - S_j$

□ Step 3:

- ▣ Create a node U that joins pair with lowest M_{ij}

- $S_U = (D_{ij} / 2) + (S_i - S_j) / 2$

How to construct a tree with Neighbor-joining method?

- Step 4:
 - ▣ Join i and j according to S and make all other taxa in form of a star
- Step 5:
 - ▣ Recalculate new distance matrix of all other taxa to U with:
 - $D_{xU} = D_{ix} + D_{jx} - D_{ij}$

Example of Neighbor-joining

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

□ Step 1: S calculation : $S_x = (\text{sum all } D_x) / (\text{leaves} - 2)$

□ $S(A) = (5 + 4 + 7 + 6 + 8) / 4 = 7.5$

□ $S(B) = (5 + 7 + 10 + 9 + 11) / 4 = 10.5$

□ $S(C) = (4 + 7 + 7 + 6 + 8) / 4 = 8$

□ $S(D) = (7 + 10 + 7 + 5 + 9) / 4 = 9.5$

□ $S(E) = (6 + 9 + 6 + 5 + 8) / 4 = 8.5$

□ $S(F) = (8 + 11 + 8 + 9 + 8) / 4 = 11$

Example of Neighbor-joining cont 1

- Step 2: Calculate pair with smallest M

$$M_{ij} = \text{Distance } ij - S_i - S_j$$

- Smallest are

- ▣ $M(AB) = d(AB) - S(A) - S(B) = 5 - 7.5 - 10.5 = -13$

- ▣ $M(DE) = 5 - 9.5 - 8.5 = -13$

	A	B	C	D	E
B	-13				
C	-11.5	-11.5			
D	-10	-10	-10.5		
E	-10	-10	-10.5	-13	
F	-10.5	-10.5	-11	-11.5	-11.5

Example of Neighbor-joining cont 2

- Step 3: Create a node U

$$S_{1U} = (D_{ij} / 2) + (S_i - S_j) / 2$$

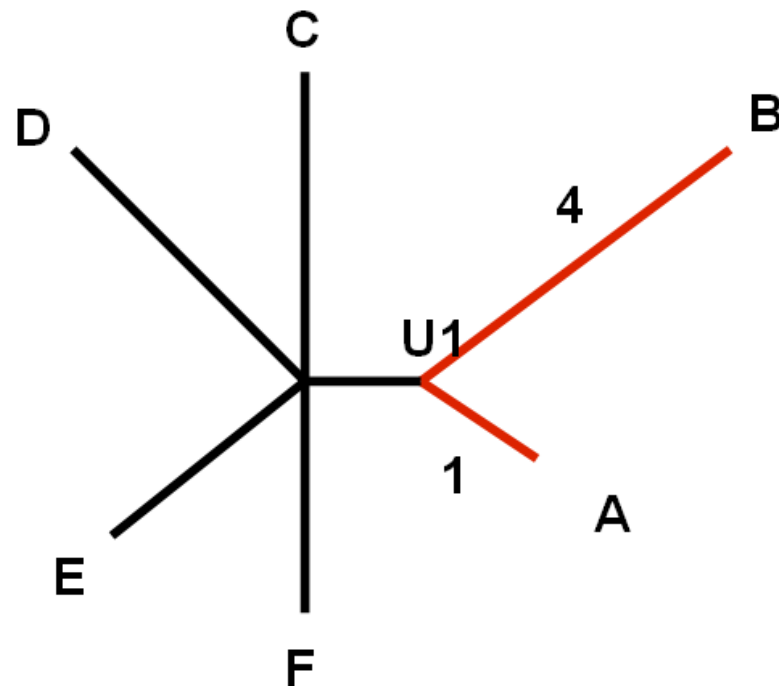
- U1 joins A and B:

- ▣ $S(AU1) = d(AB) / 2 + (S(A) - S(B)) / 2$
 $= 5 / 2 + (7.5 - 10.5) / 2 = 1$

- ▣ $S(BU1) = d(AB) / 2 + (S(B) - S(A)) / 2$
 $= 5 / 2 + (10.5 - 7.5) / 2 = 4$

Example of Neighbor-joining cont 3

- Step 4: Join A and B according to S , and make all other taxa in form of a star. Branches in black are unknown length and Branches in red are known length



Example of Neighbor-joining cont 4

- Step5: Calculate new distance matrix

$$D_{xu} = (D_{ix} + D_{jx} - D_{ij}) / 2$$

- ▣ $d(CU) = (d(AC) + d(BC) - d(AB)) / 2$
 $= (4 + 7 - 5) / 2 = 3$

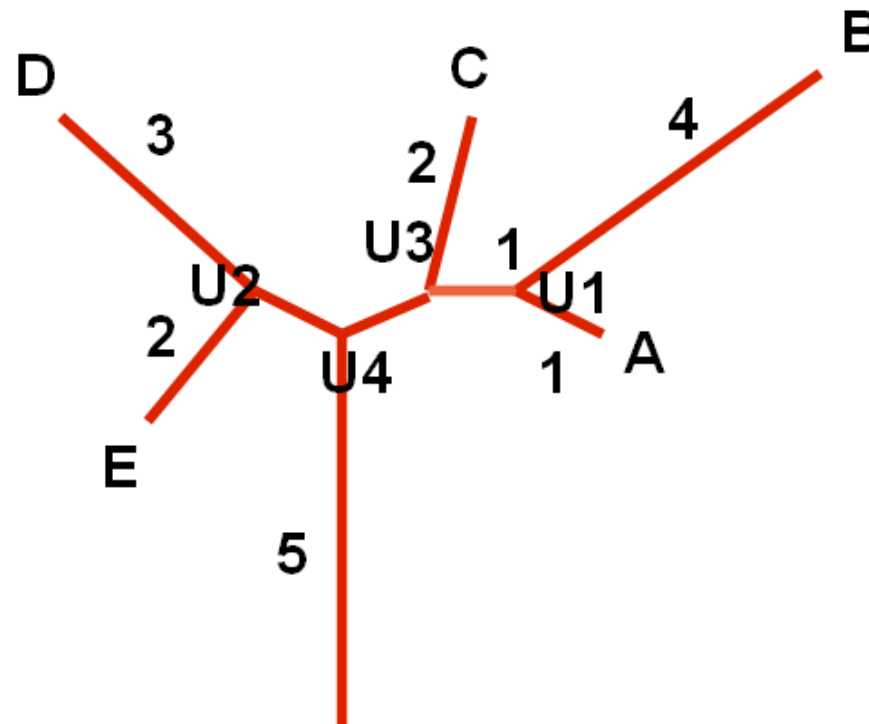
- ▣ $d(DU) = d(AD) + d(BD) - d(AB) / 2 = 6$
Same as EU and FU

- Then we get the new distance matrix

	U1	C	D	E
C	3			
D	6	7		
E	5	6	5	
F	7	8	9	8

Example of Neighbor-joining cont 5

- Repeat 1 to 5 until all branches are done
- In this example, we will get this at the end



Downside of Neighbor-joining

- Generates only one possible tree
- Generates only unrooted tree