

# - JSMC Practical Course - Inferring Phylogeny Based on Sequence Information

**Thursday – Friday, March 21 – 22**

**Room 316, Philosophenweg 12**

Wireless LAN:           eduroam  
Username:                [tagung09@uni-jena.de](mailto:tagung09@uni-jena.de)  
Password:                Gver58ges



# Molecular phylogenetics

Types of molecular features:

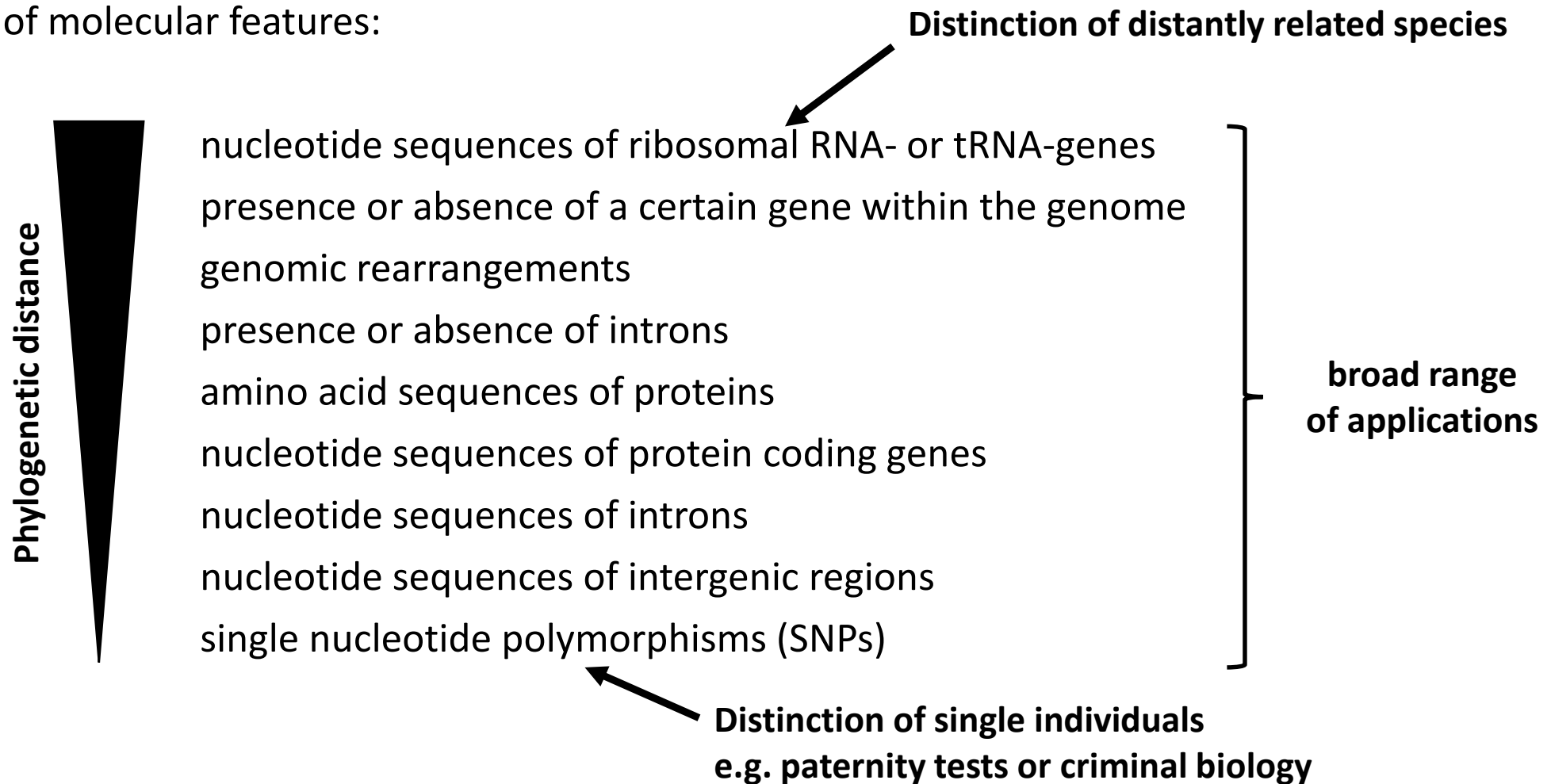
Phylogenetic distance



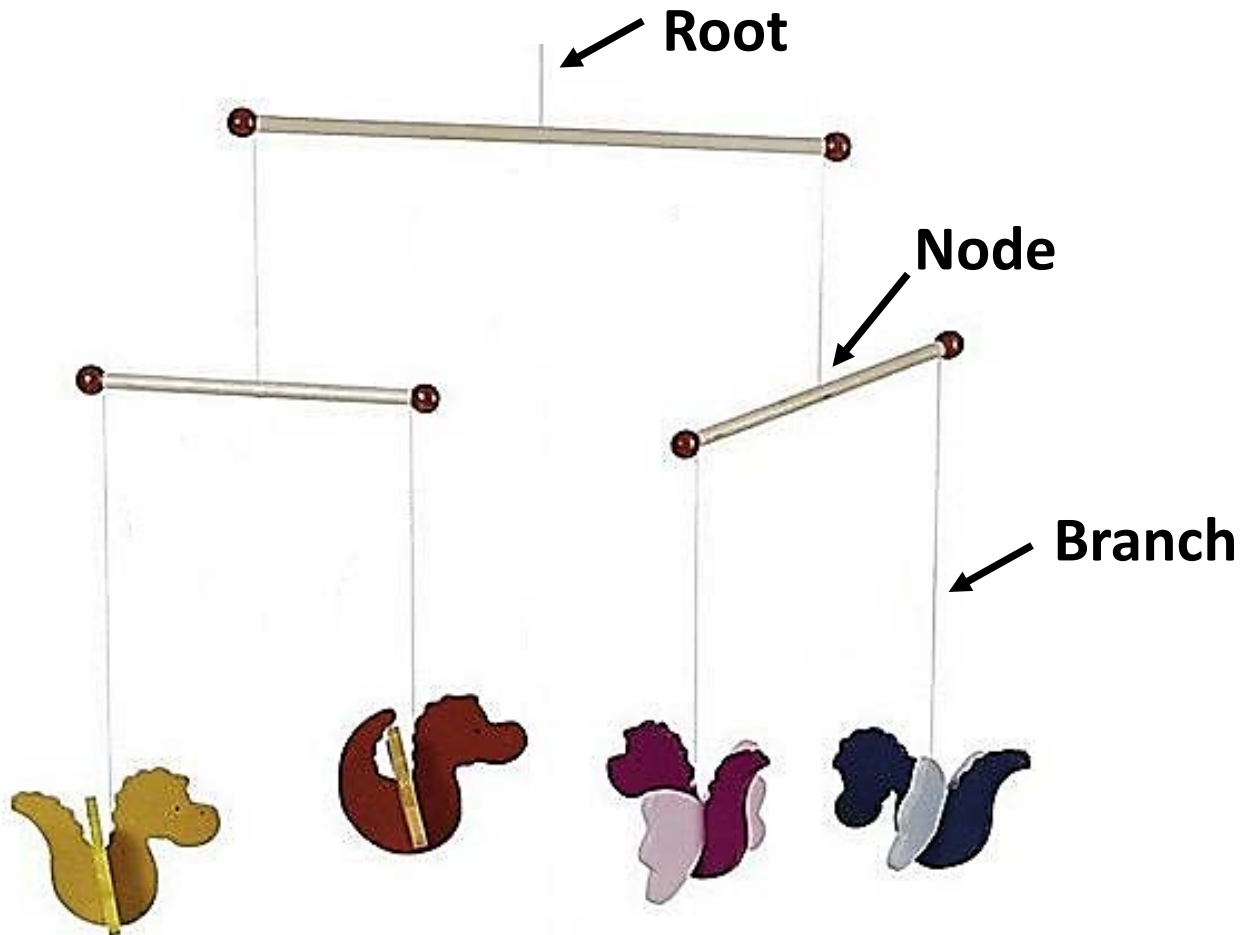
nucleotide sequences of ribosomal RNA- or tRNA-genes  
presence or absence of a certain gene within the genome  
genomic rearrangements  
presence or absence of introns  
amino acid sequences of proteins  
nucleotide sequences of protein coding genes  
nucleotide sequences of introns  
nucleotide sequences of intergenic regions  
single nucleotide polymorphisms (SNPs)

# Molecular phylogenetics

Types of molecular features:



# Interpretation of phylogenetic trees



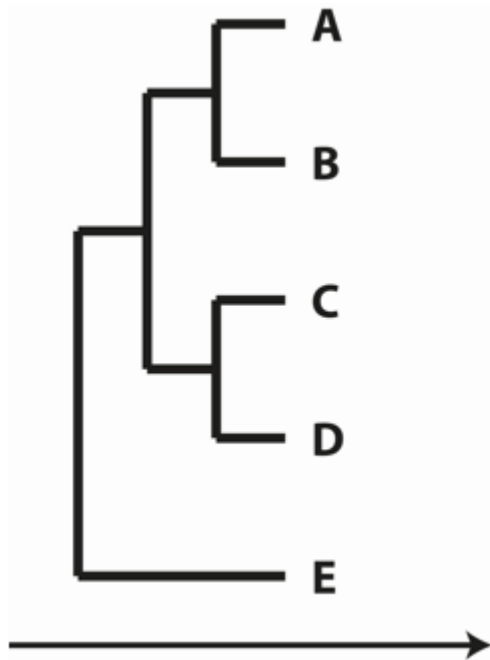
The order of the taxa (terminal branches) is not of importance.

Each sub-tree can be arbitrarily rotated at each node (so that the order of the taxa changes).

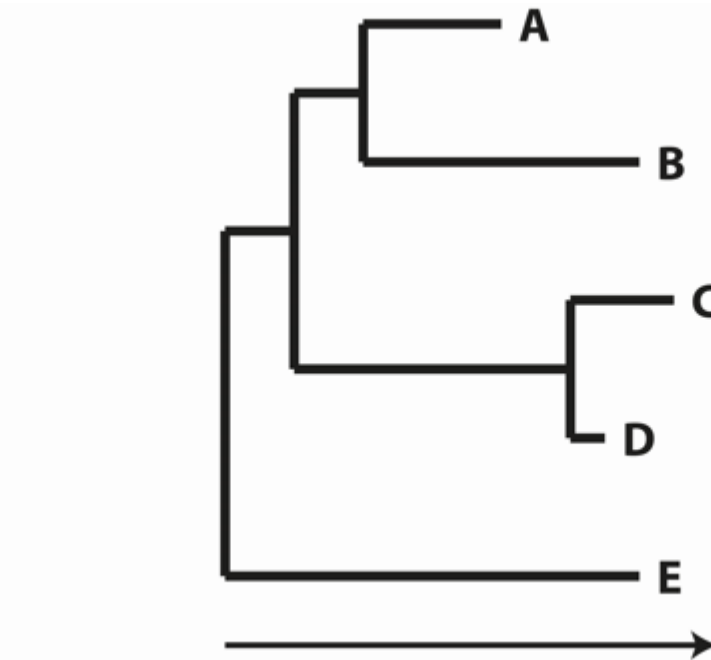
Only the topology of the tree (i.e. the node structure) specifies the phylogenetic relationship!

# Types of phylogenetic trees

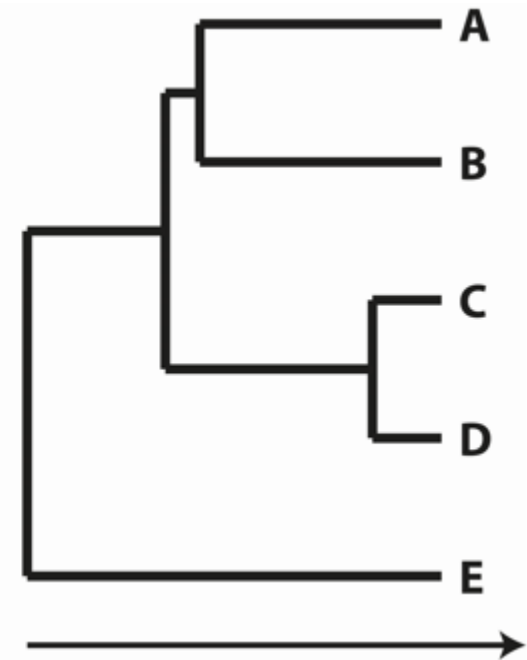
**Cladogram**



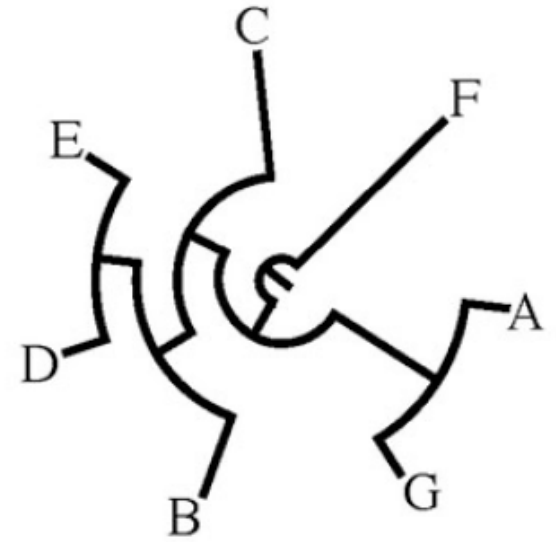
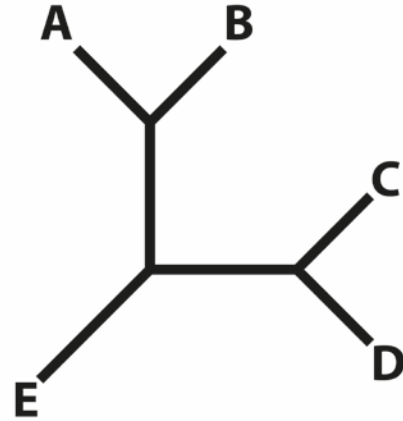
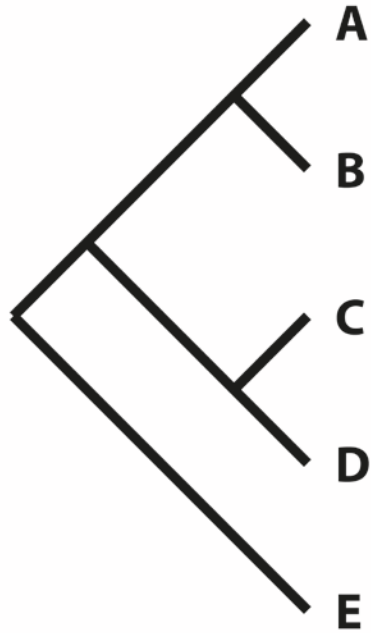
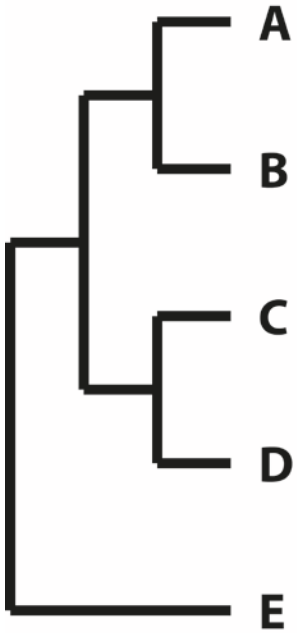
**Dendrogram**



**Phylogram**

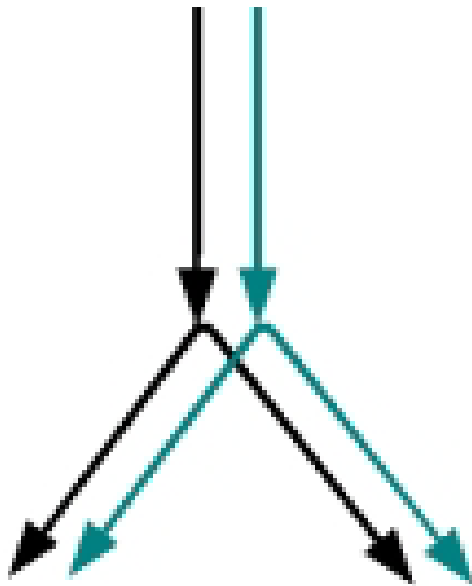


# Display formats

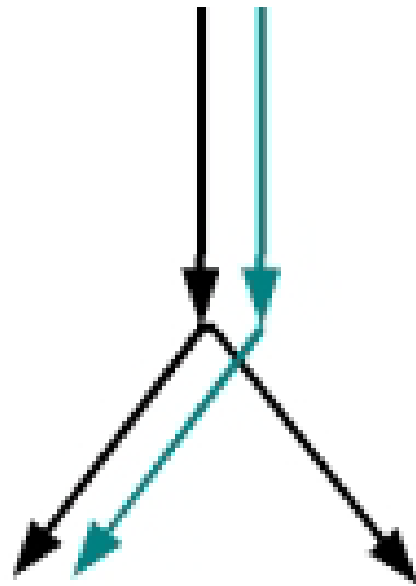


# Gene trees and species trees

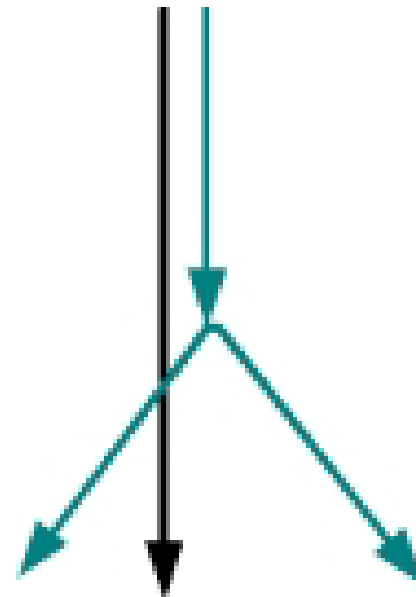
**Speciation**



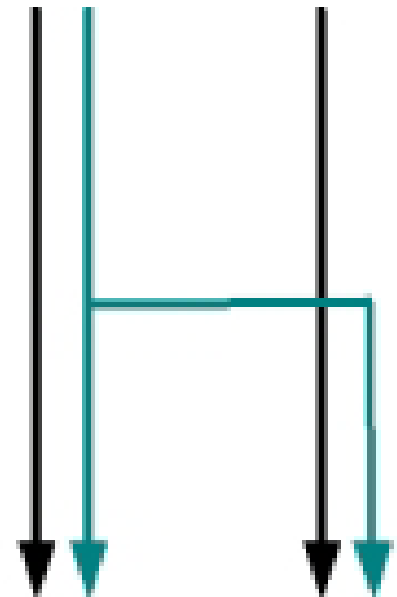
**Gene loss**



**Gene duplication**



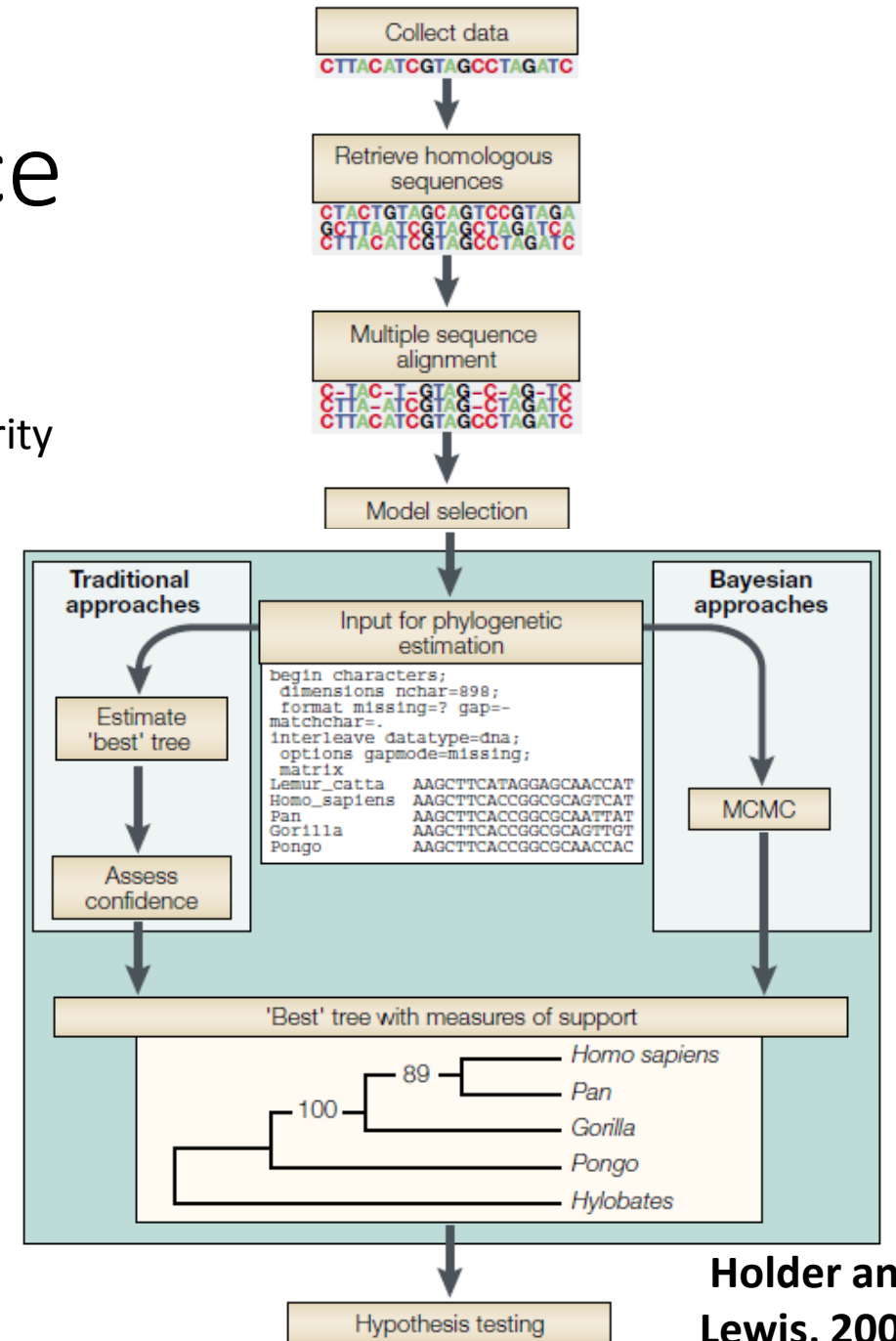
**Horizontal gene transfer**





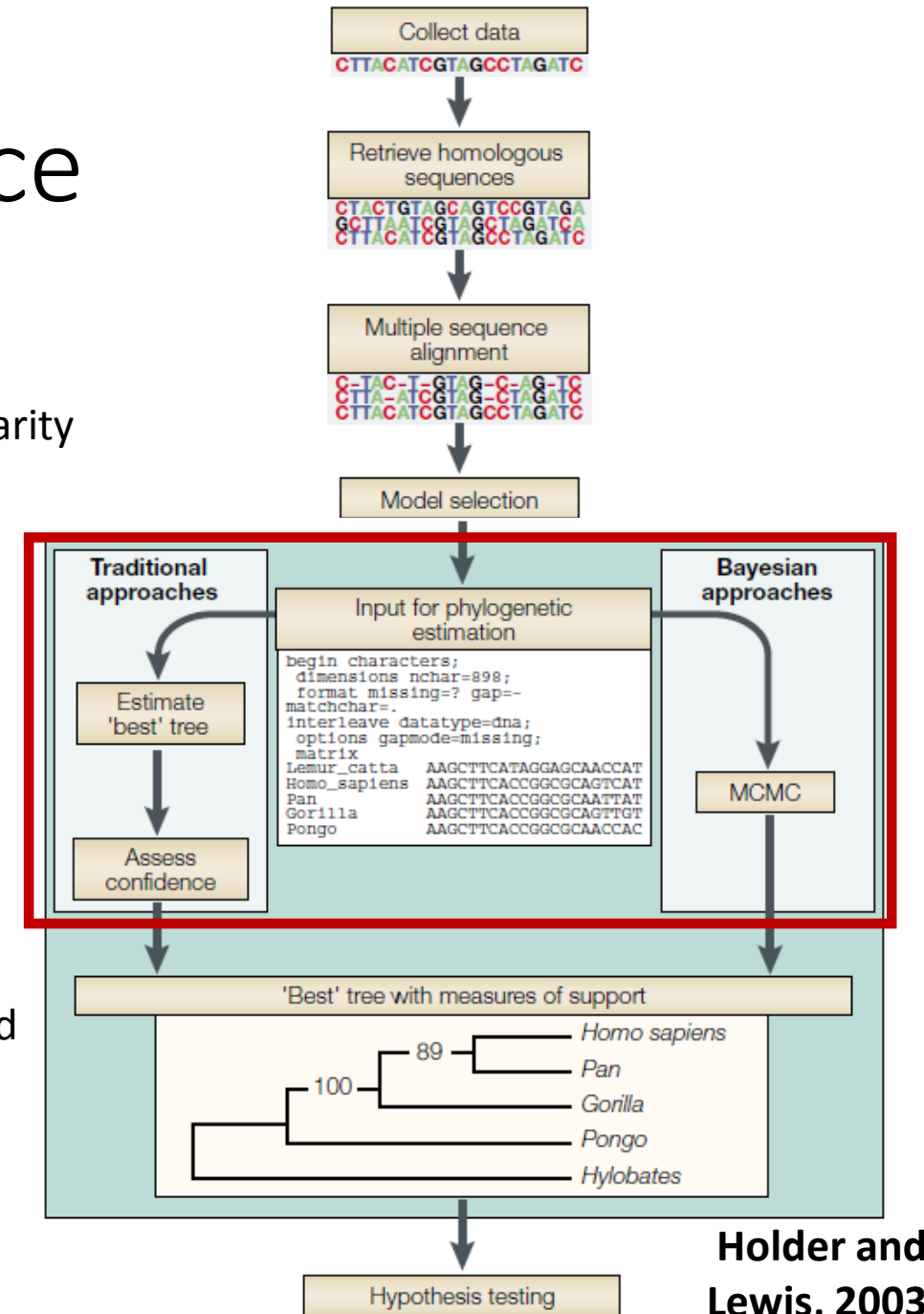
# Process of phylogeny inference

- Data collection
  - Homologous sequences are searched based on sequence similarity  
-> **BLAST**
- Multiple sequence alignment
  - Homologous sites are detected and aligned along each other  
-> **MAFFT**
- Selection of an appropriate model to infer phylogeny
  - Based on the level of sequence identity/similarity among the alignment members their phylogenetic relation is reconstructed  
-> **Neighbour joining**  
-> **Bayesian phylogeny inference**



# Process of phylogeny inference

- Data collection
  - Homologous sequences are searched based on sequence similarity  
-> **BLAST**
- Multiple sequence alignment
  - Homologous sites are detected and aligned along each other  
-> **MAFFT**
- Selection of an appropriate model to infer phylogeny
  - Based on the level of sequence identity/similarity among the alignment members their phylogenetic relation is reconstructed  
-> **Neighbour joining**  
-> **Bayesian phylogeny inference**



# Process of phylogeny inference

## **Tree construction and searching methods**

- Stepwise addition
- Star decomposition
- Heuristic search
- Exact search

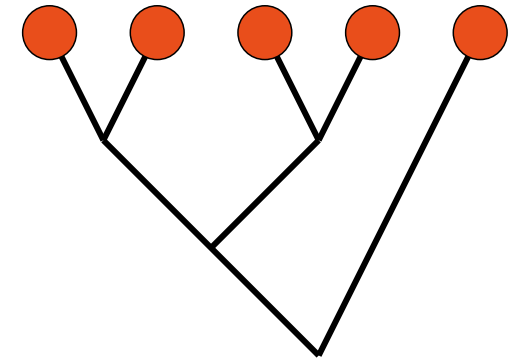
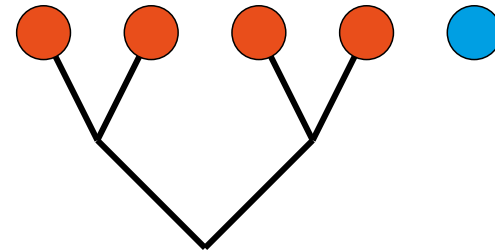
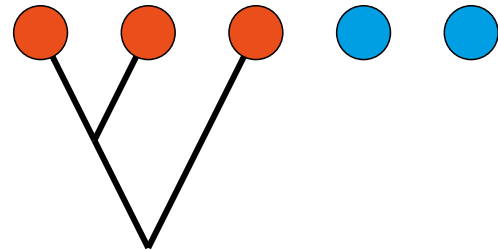
## **Tree evaluation methods (optimality criteria)**

- Minimum evolution
- Parsimony
- Maximum likelihood

# Tree construction and searching methods

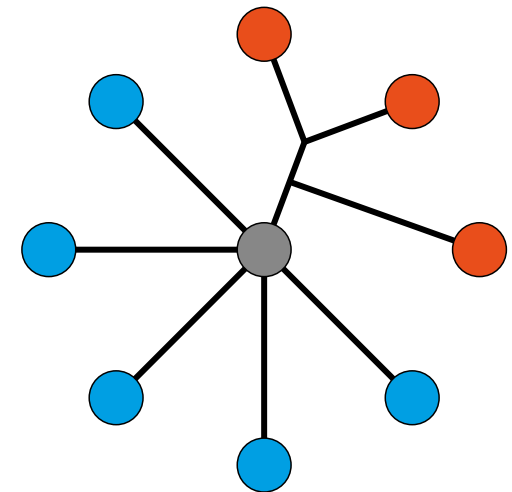
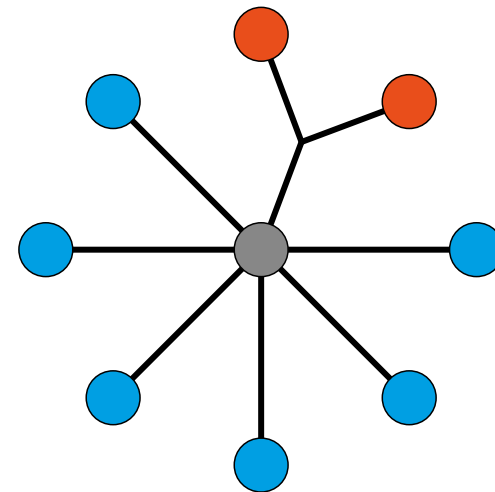
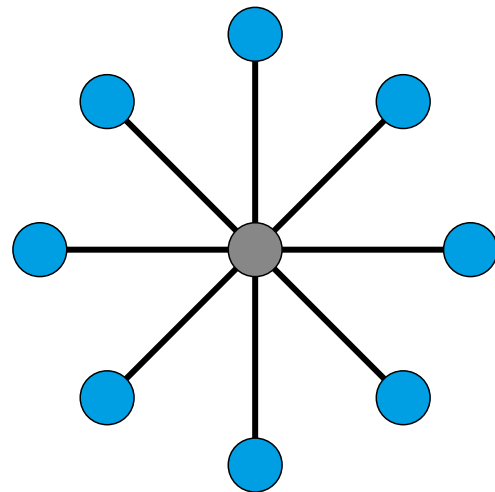
## - Stepwise addition

Attaches lineage by lineage according to their relative similarity



## - Star decomposition

Joins lineage by lineage according to their relative similarity



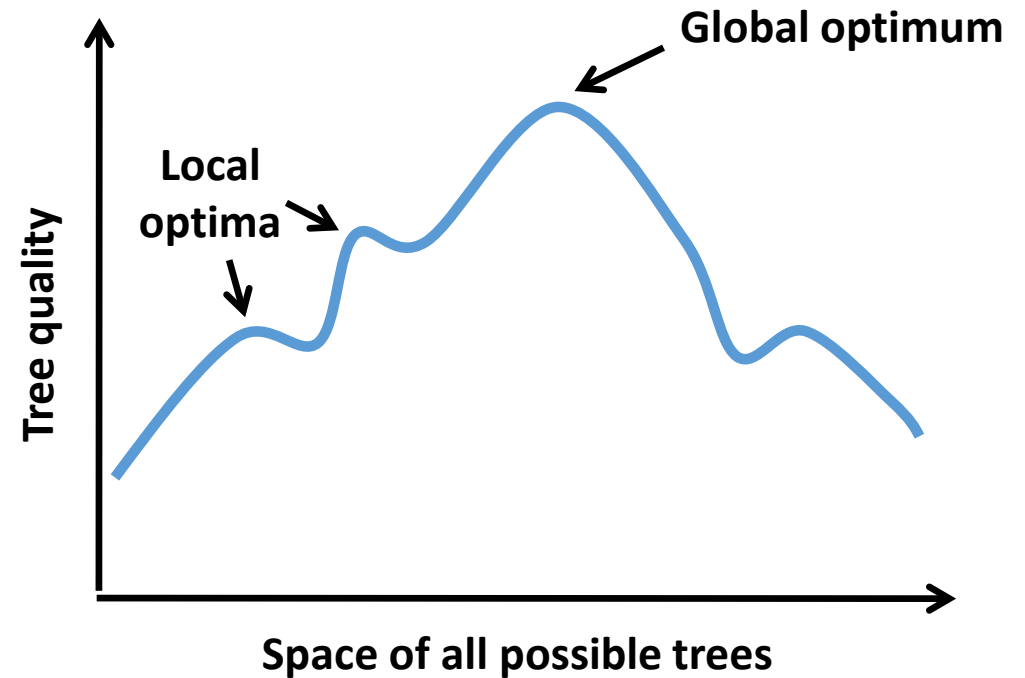
# Tree construction and searching methods

- **Heuristic search**

Performs branch swapping to generate alternative trees in attempt to find a better tree

- **Exact search**

Searches the complete 'space' of possible trees



# Tree evaluation methods (optimality criteria)

## Minimum Evolution

- Uses a distance matrix to evaluate tree quality
- For every tree the branch lengths are estimated that best explain the observed distances

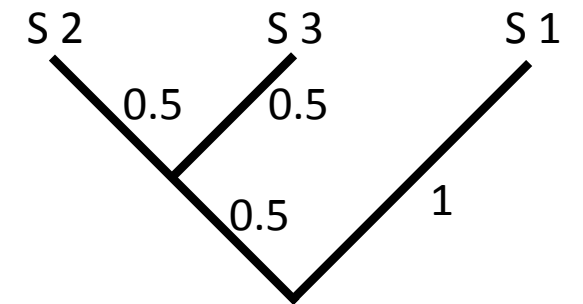
-> fast

-> can correct for unseen changes

-> weaknesses for long branches (i.e. high evolutionary distances)

	Position 1	Position 2	Position 3
Sequence 1	A	A	A
Sequence 2	A	T	G
Sequence 3	A	T	C

	S 1	S 2	S 3
S 1	0	2	2
S 2	2	0	1
S 3	2	1	0



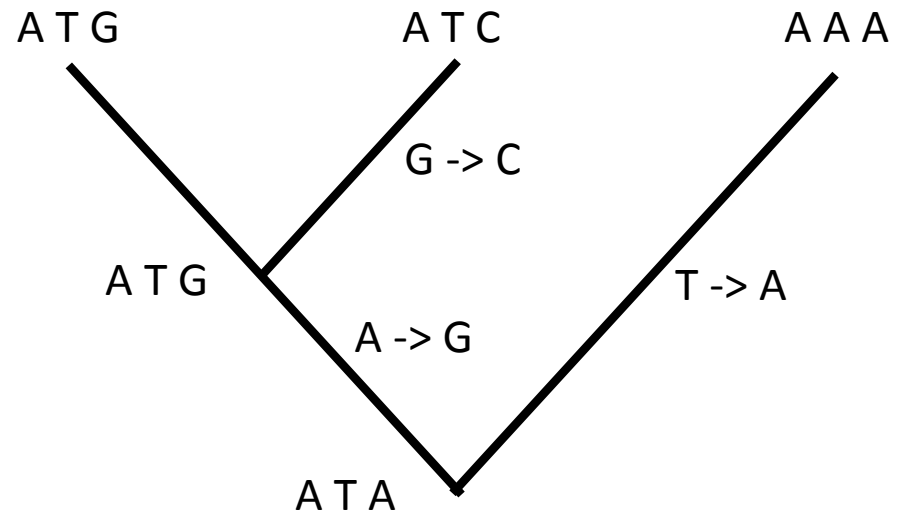
# Tree evaluation methods (optimality criteria)

## Parsimony

- Maps sequence history onto tree
- Evaluates tree quality by finding the minimum number of mutations that could explain the data

- > fast enough for hundreds of sequences
- > does not correct for multiple mutational pathways of the same tree
- > performs poorly if branch length differ

	Position 1	Position 2	Position 3
Sequence 1	A	A	A
Sequence 2	A	T	G
Sequence 3	A	T	C



# Tree evaluation methods (optimality criteria)

## Maximum likelihood

- Maps sequence history onto tree
- Finds the tree that is most likely to explain the data
  - > captures all possible mutational pathways
  - > corrects for multiple mutational events at the same site
  - > slow



# Process of phylogeny inference

## Tree construction and searching methods

- Stepwise addition
- Star decomposition
- Heuristic search
- Exact search

## Tree evaluation methods (optimality criteria)

- Minimum evolution
- Parsimony
- Maximum likelihood

**Produce only  
one tree**

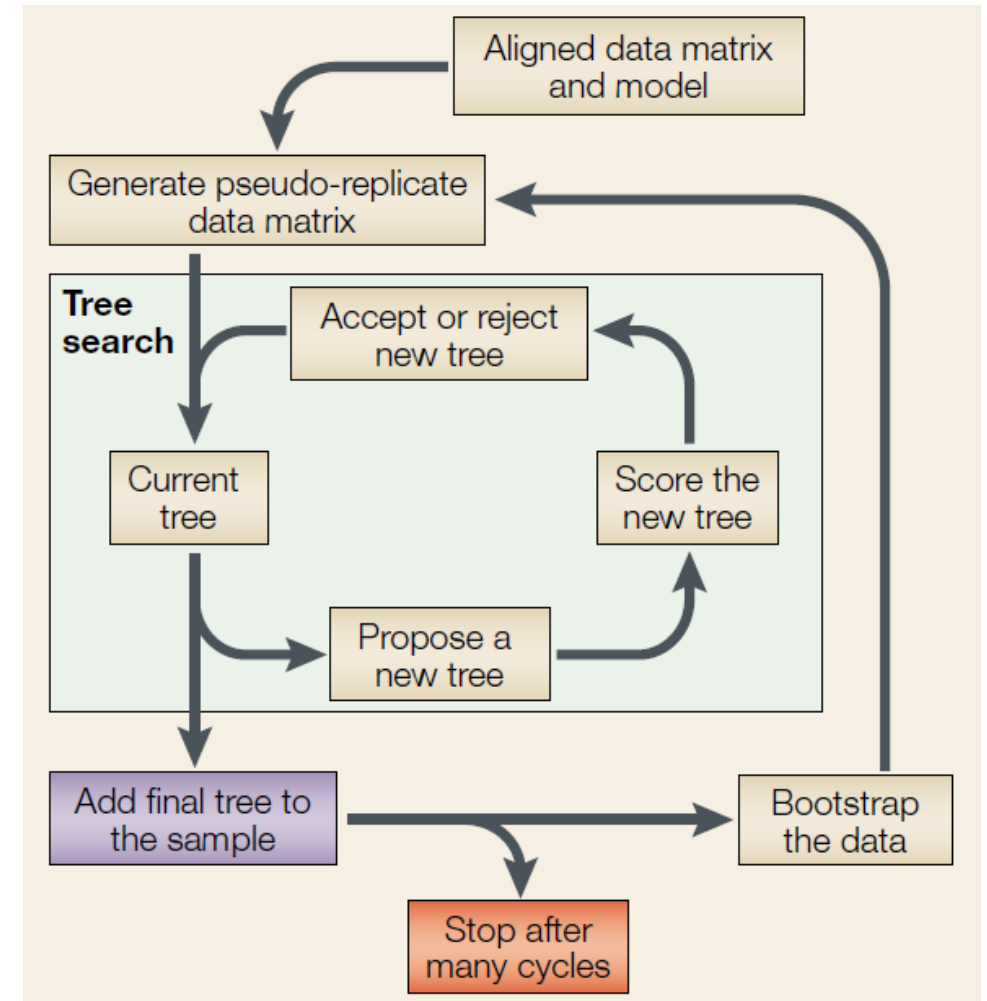
**No information about the reliability of  
single branches**

**-> Bootstrapping**

# Bootstrapping

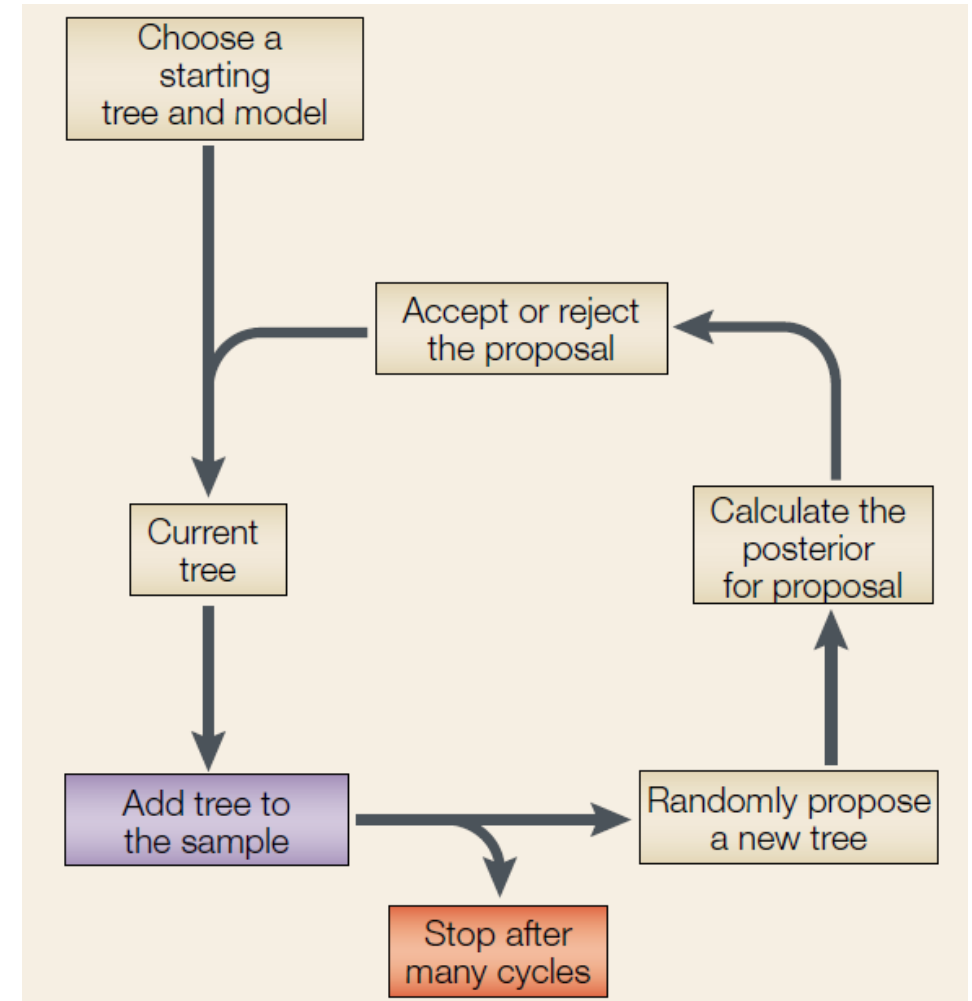
- Creates pseudo-replicates of original data
- Performs the same tree search for all pseudo-replicates and stores the trees
- The reliability of a certain grouping is determined based on the number of trees that show this grouping

-> **very time consuming**



# Bayesian phylogenetics

- Performs tree search and measure of support simultaneously
- Uses Markov chain Monte Carlo (MCMC) simulations to produce alternative trees
- Not a strict 'hill-climber' (does not only accept better trees)
- Higher probability to reach global optimum



# Bayesian phylogenetics

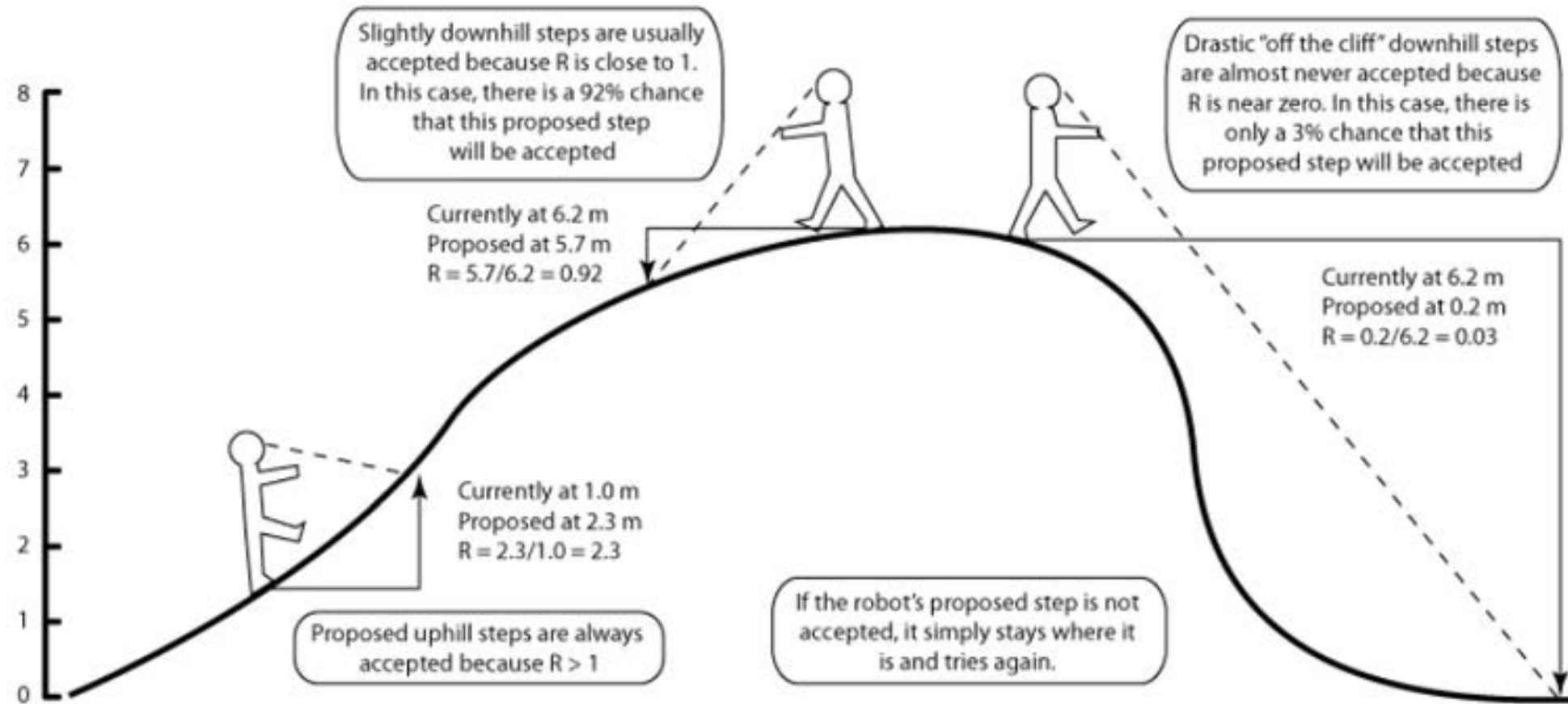
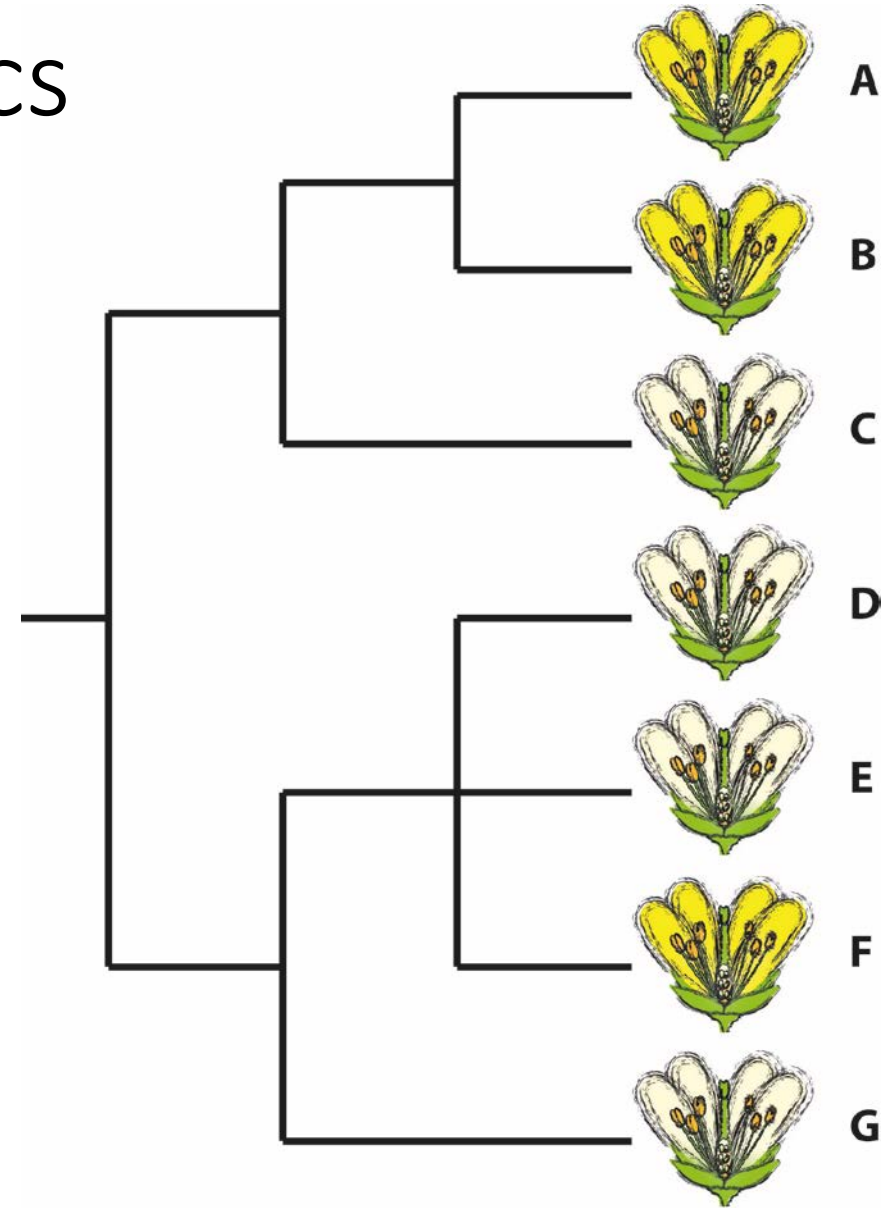


Illustration of MCMC method process (Lewis, 2011)

# Basic terms of phylogenetics

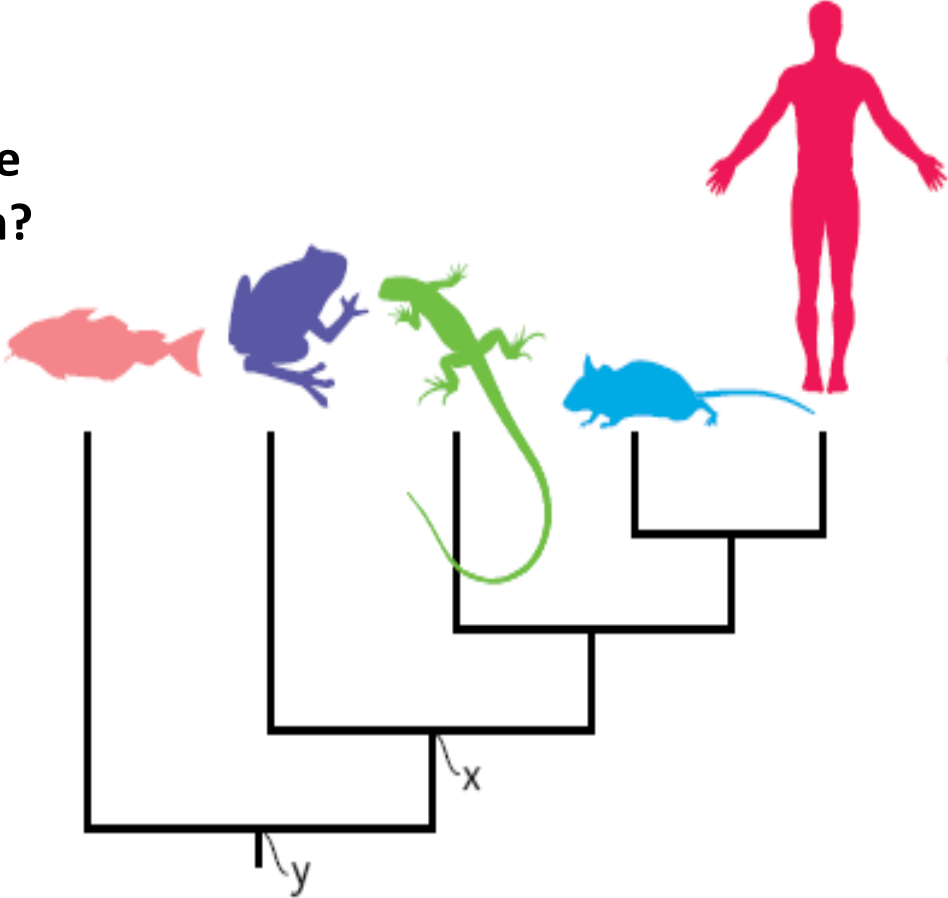
- Monophyletic group
- Paraphyletic group
- Polyphyletic group
- Apomorphic feature
- Plesiomorphic feature
- Autapomorphy
- Synapomorphy
- Sympleiomorphy
- Polytomy

- Analogy
- Homology
- Homoplasy
- Convergence
- MRCA
- Extant species
- Extinct species
- Dichotomy



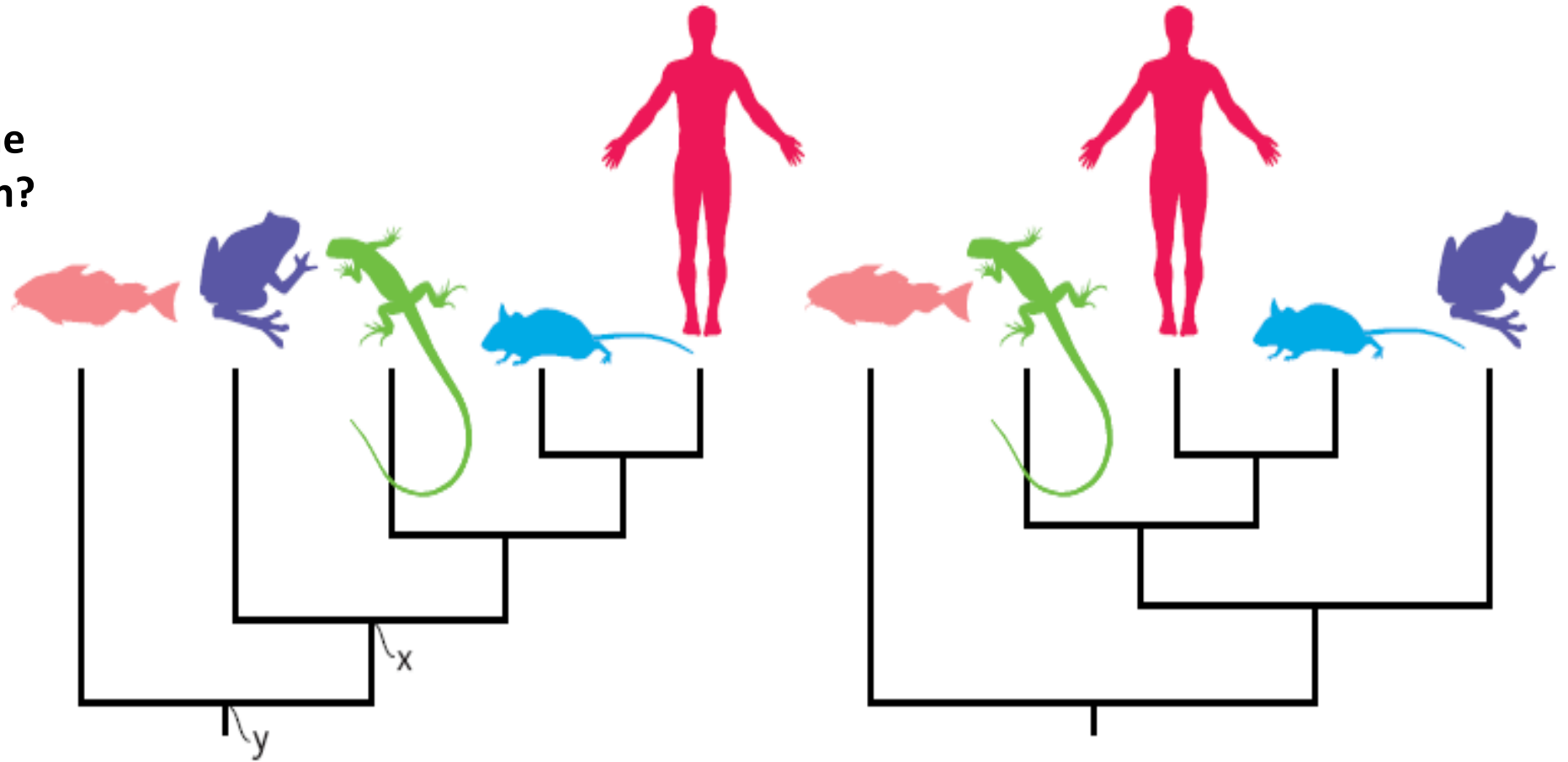
# The Tree Thinking Challenge

Is the frog more closely related to the human or to the fish?

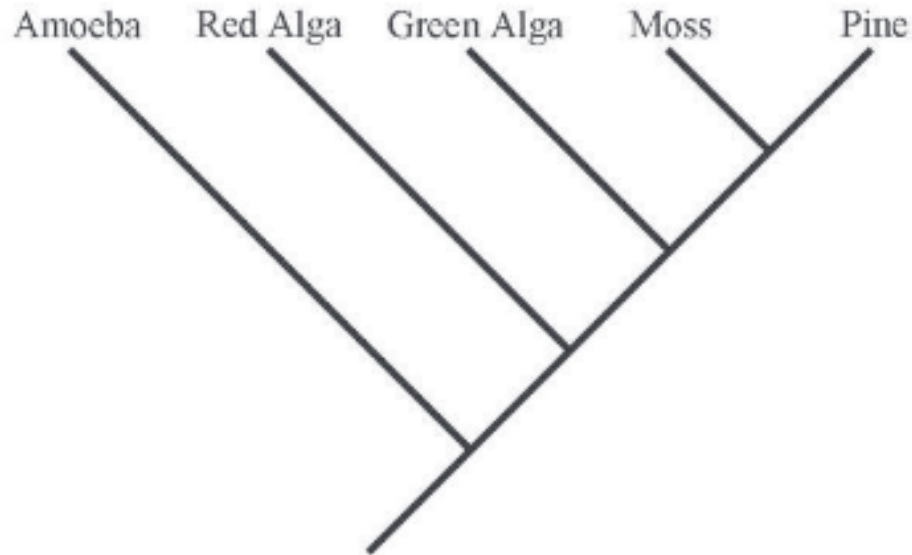


# The Tree Thinking Challenge

Is the frog more closely related to the human or to the fish?



# The Tree Thinking Challenge

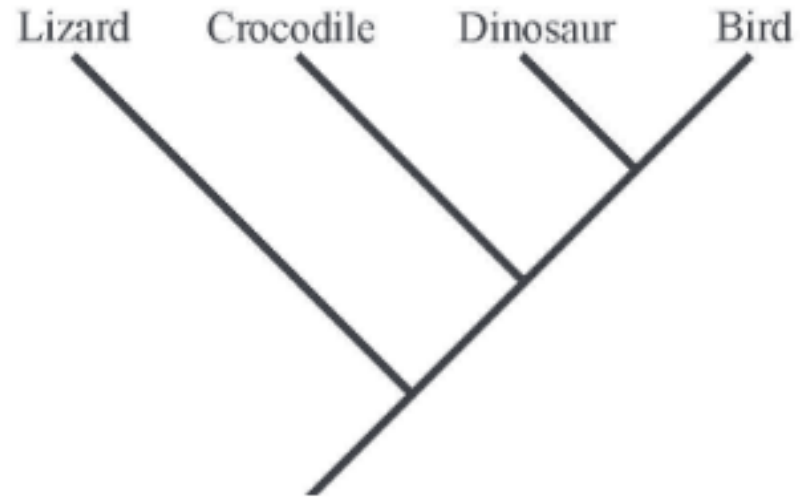


1) By reference to the tree above, which of the following is an accurate statement of relationships?

- a) A green alga is more closely related to a red alga than to a moss
- b) A green alga is more closely related to a moss than to a red alga
- c) A green alga is equally related to a red alga and a moss
- d) A green alga is related to a red alga, but is not related to a moss



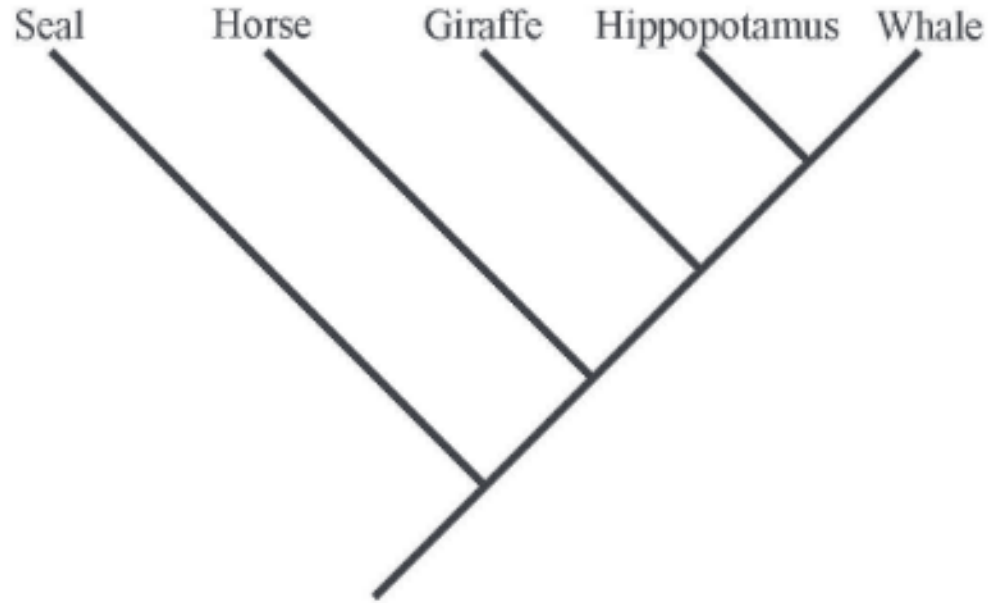
# The Tree Thinking Challenge



2) By reference to the tree above, which of the following is an accurate statement of relationships?

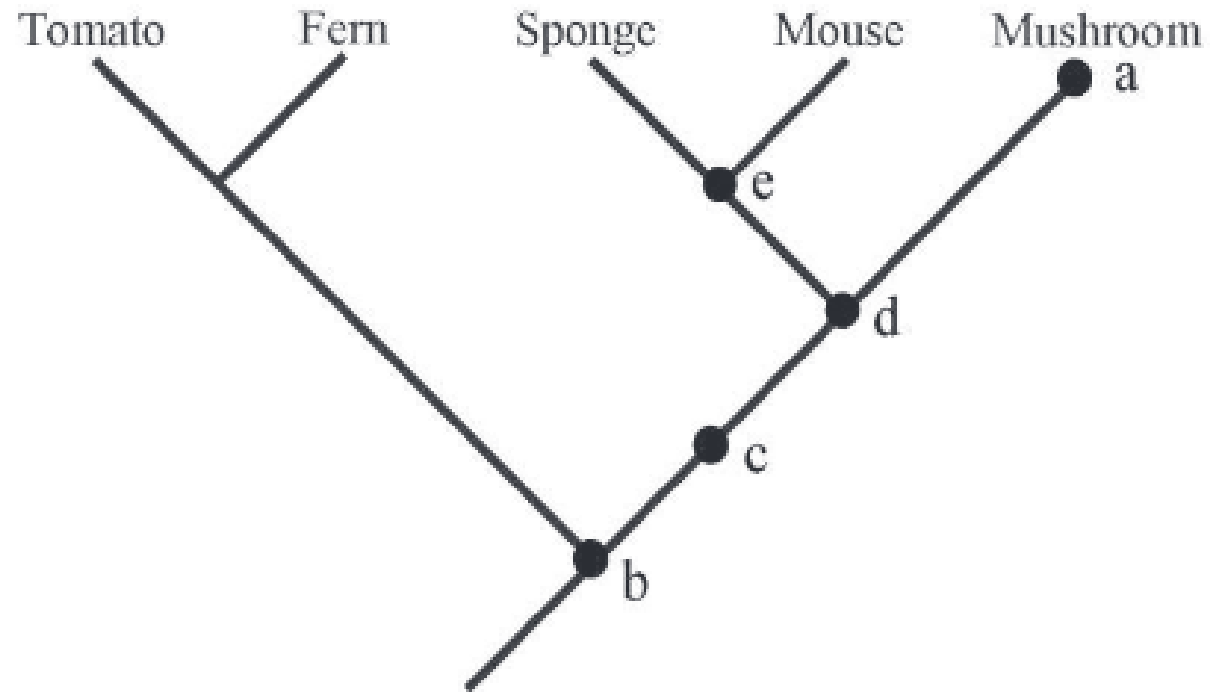
- a) A crocodile is more closely related to a lizard than to a bird
- b) A crocodile is more closely related to a bird than to a lizard
- c) A crocodile is equally related to a lizard and a bird
- d) A crocodile is related to a lizard, but is not related to a bird

# The Tree Thinking Challenge



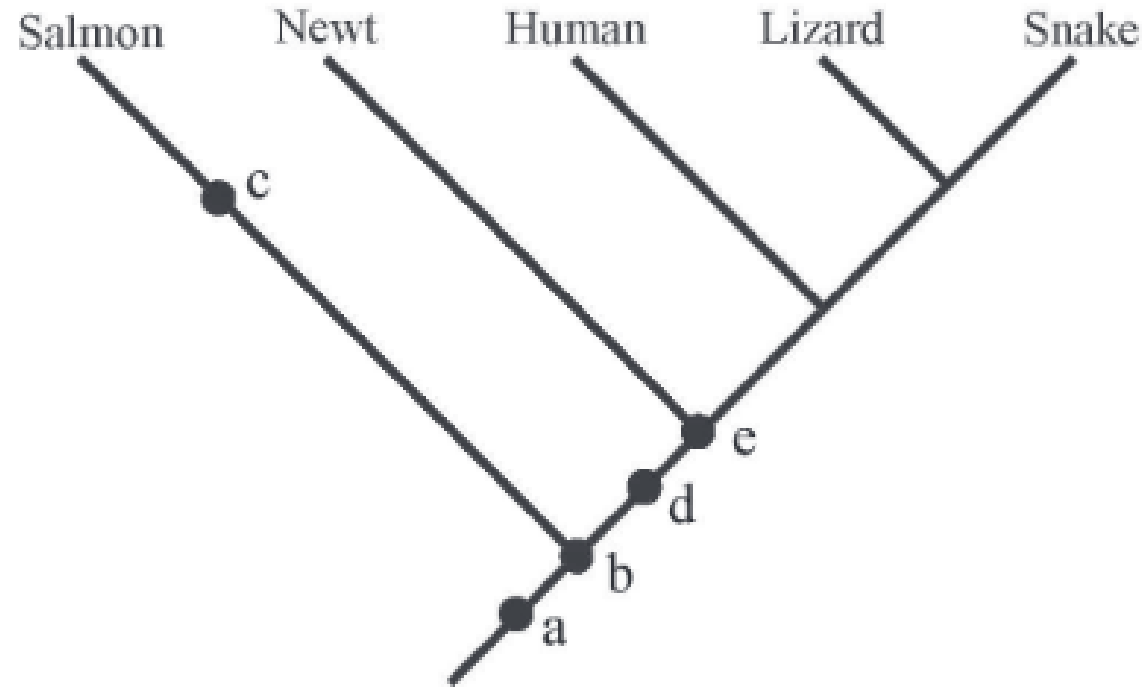
- 3) By reference to the tree above, which of the following is an accurate statement of relationships?
- a) A seal is more closely related to a horse than to a whale
  - b) A seal is more closely related to a whale than to a horse
  - c) A seal is equally related to a horse and a whale
  - d) A seal is related to a whale, but is not related to a horse

# The Tree Thinking Challenge



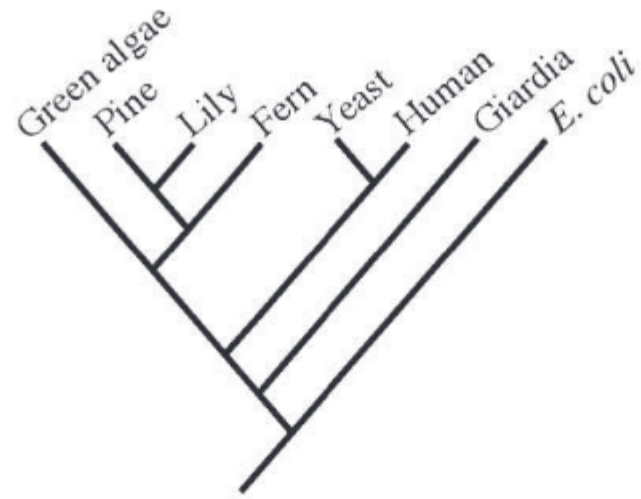
4) Which of the five marks in the tree above corresponds to the most recent common ancestor of a mushroom and a sponge?

# The Tree Thinking Challenge

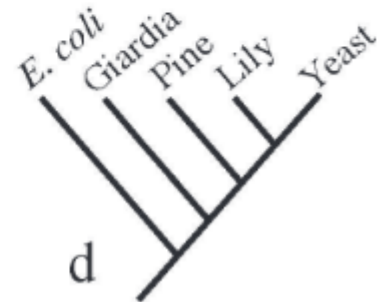
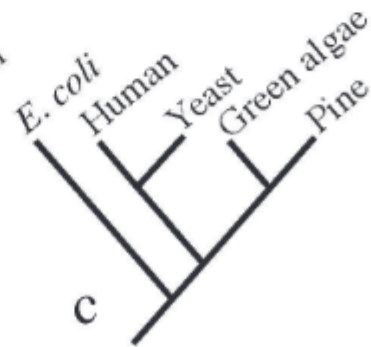
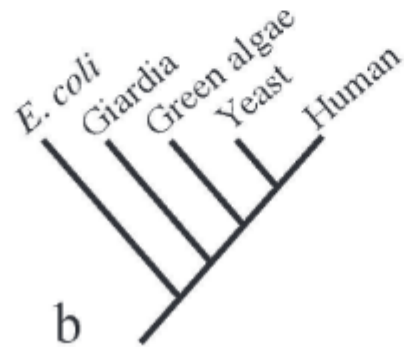
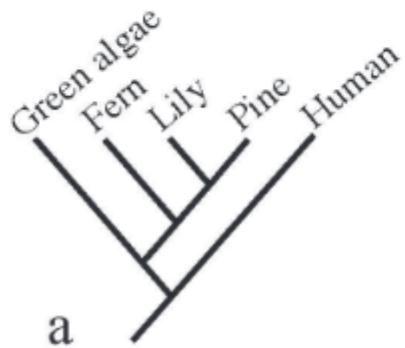


5) If you were to add a trout to the phylogeny shown above, where would its lineage attach to the rest of the tree?

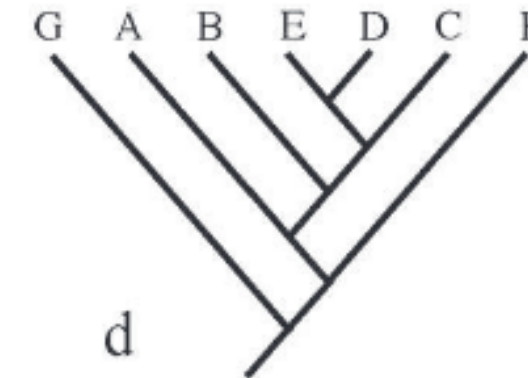
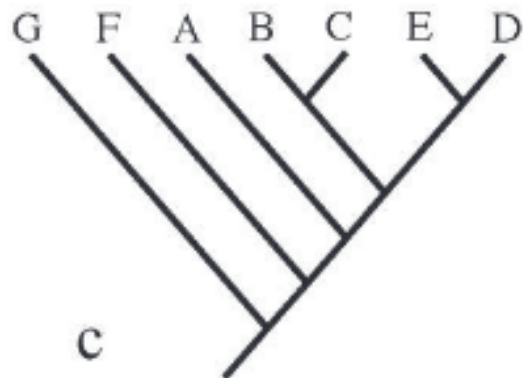
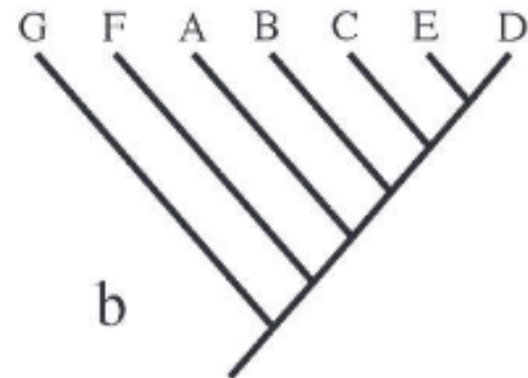
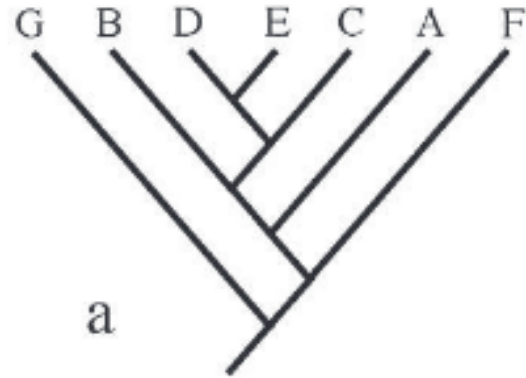
# The Tree Thinking Challenge



6) Which of trees below is false given the larger phylogeny above?

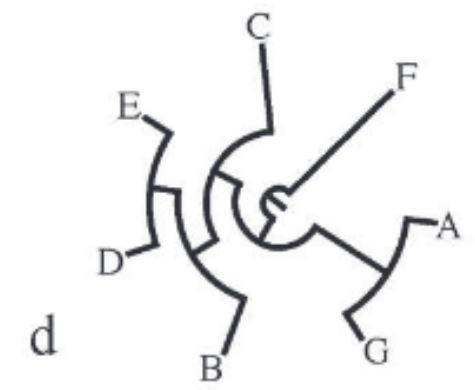
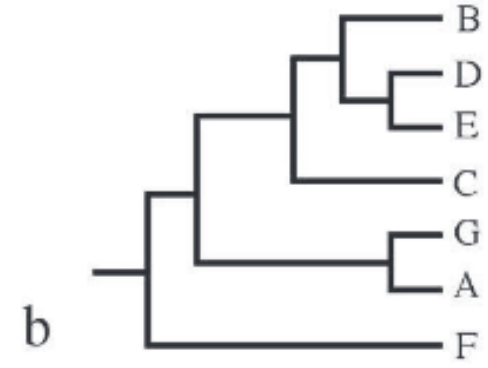
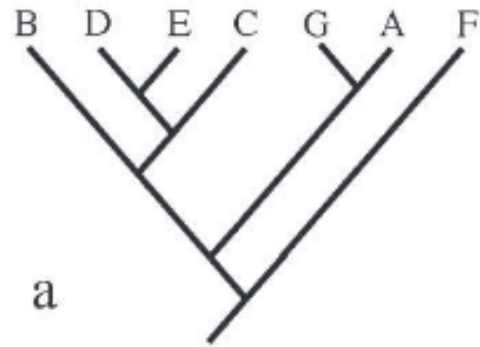


# The Tree Thinking Challenge



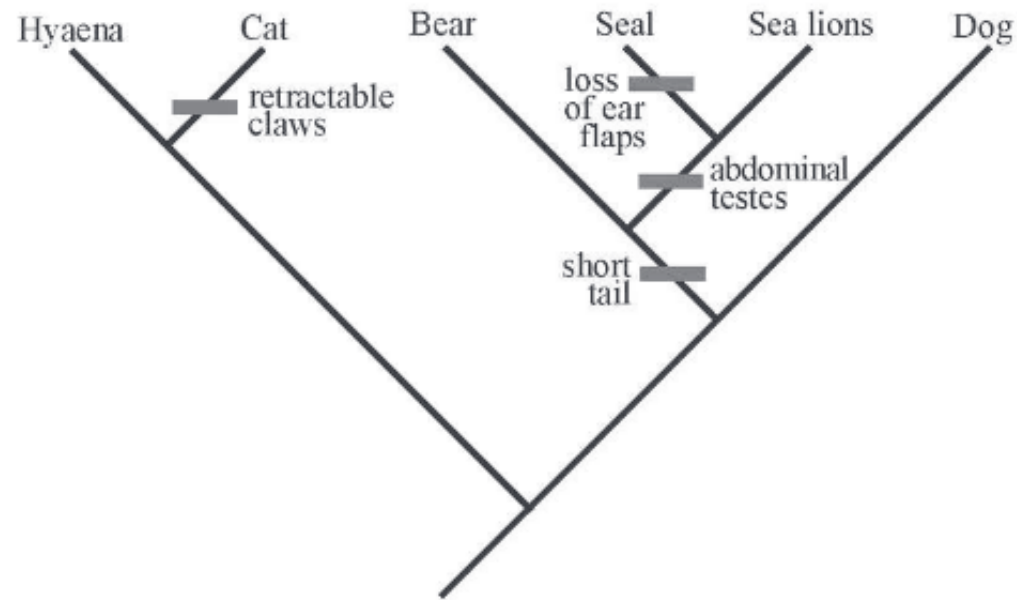
7) Which of the four trees above depicts a different pattern of relationships than the others?

# The Tree Thinking Challenge



8) Which of the four trees above depicts a different pattern of relationships than the others?

# The Tree Thinking Challenge

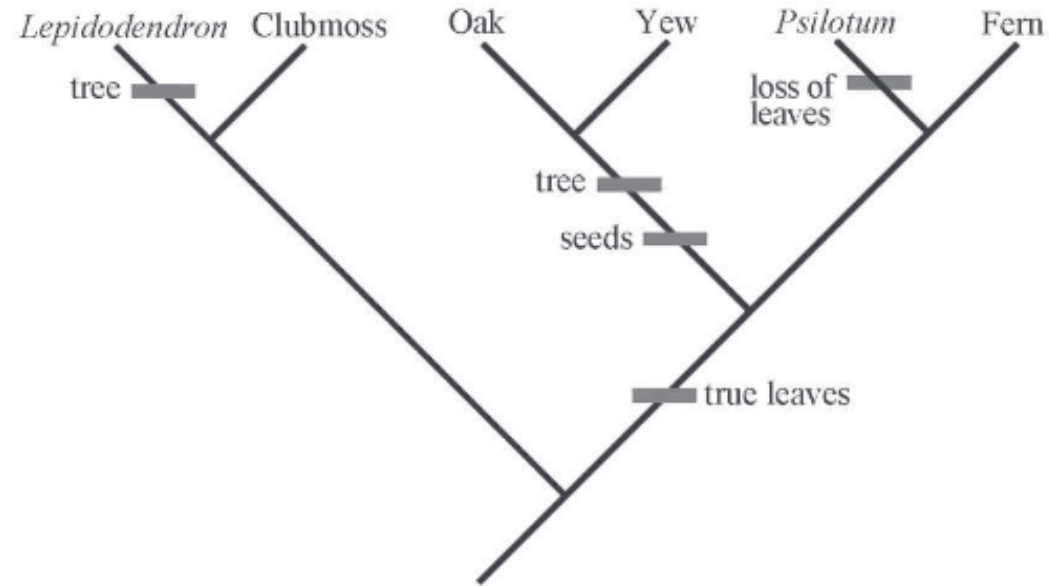


9) In the above tree, assume that the ancestor had a long tail, ear flaps, external testes, and fixed claws. Based on the tree and assuming that all evolutionary changes in these traits are shown, what traits does a sea lion have?

- a) long tail, ear flaps, external testes, and fixed claws
- b) short tail, no ear flaps, external testes, and fixed claws
- c) short tail, no ear flaps, abdominal testes, and fixed claws
- d) short tail, ear flaps, abdominal testes, and fixed claws
- e) long tail, ear flaps, abdominal testes, and retractable claws



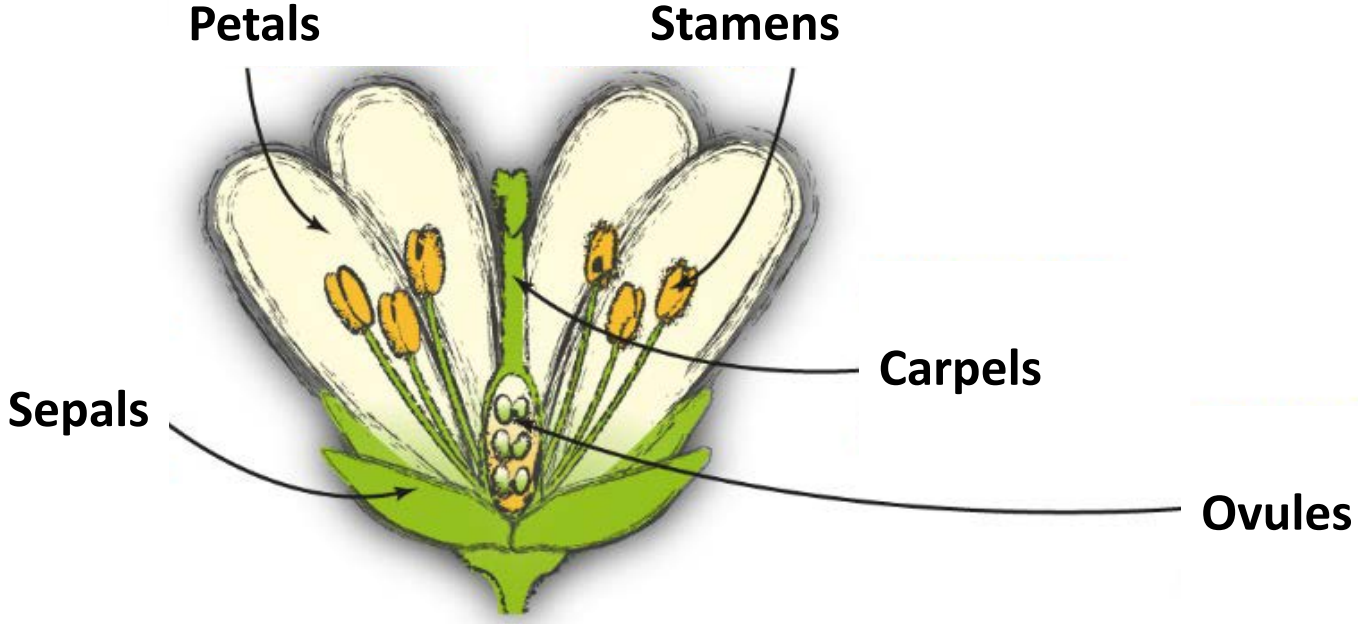
# The Tree Thinking Challenge



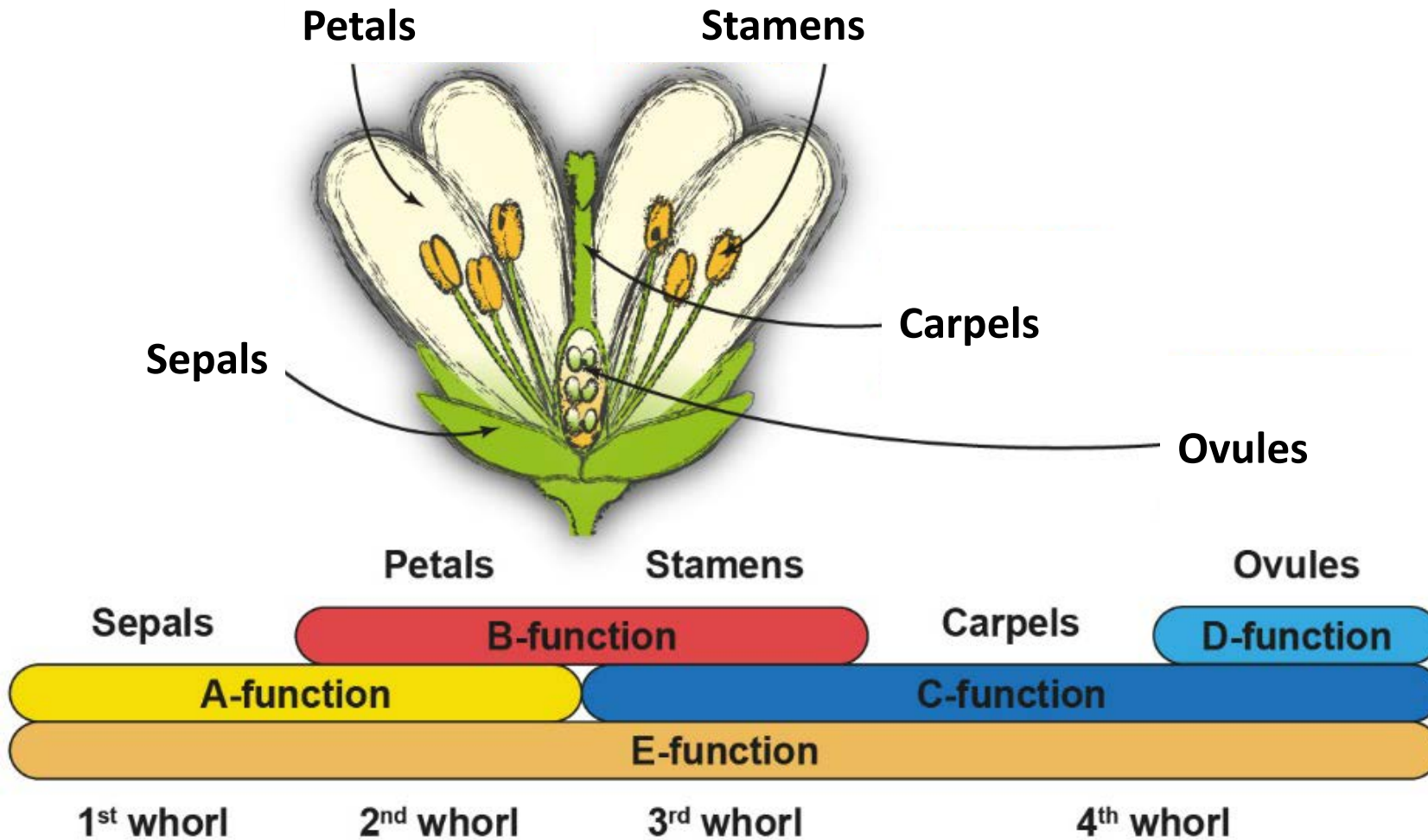
10) In the above tree, assume that the ancestor was a herb (not a tree) without leaves or seeds. Based on the tree and assuming that all evolutionary changes in these traits are shown, which of the tips has a tree habit and lacks true leaves?

- a) *Lepidodendron*
- b) Clubmoss
- c) Oak
- d) *Psilotum*
- e) Fern

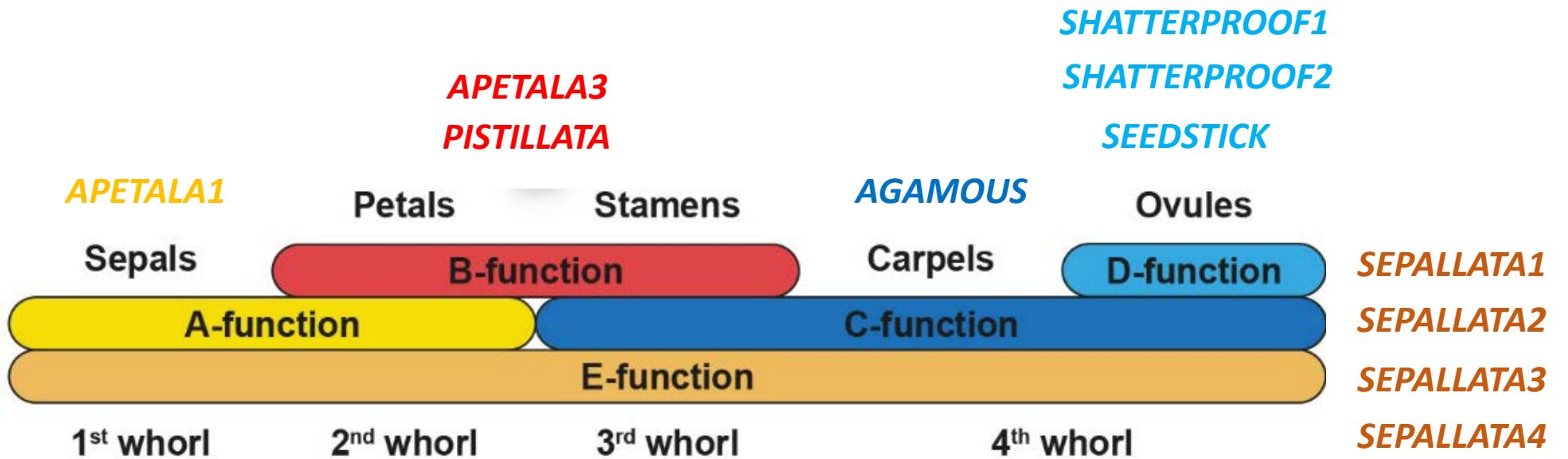
# The flower development of angiosperms



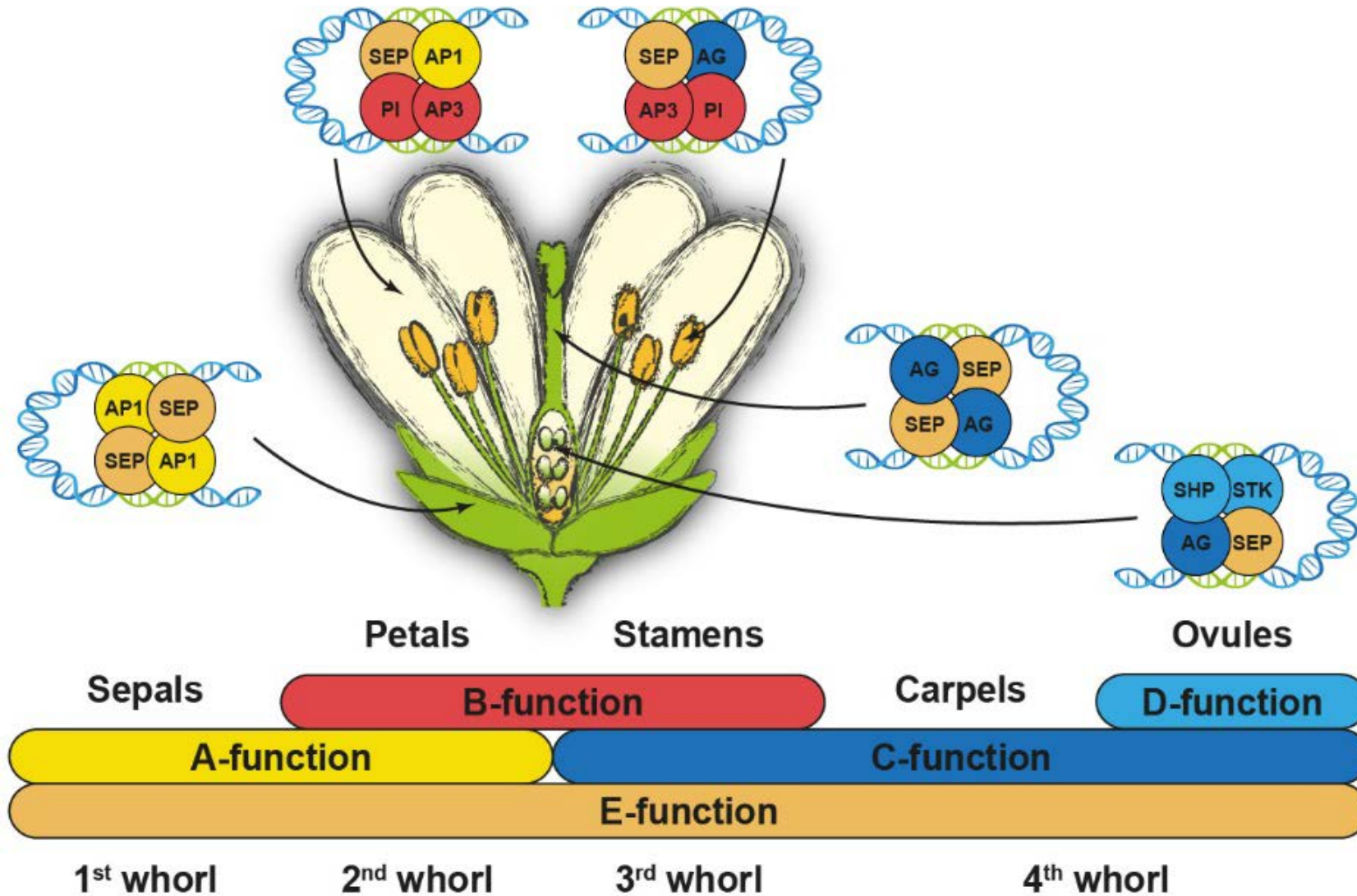
# The ABCDE model



# The ABCDE model

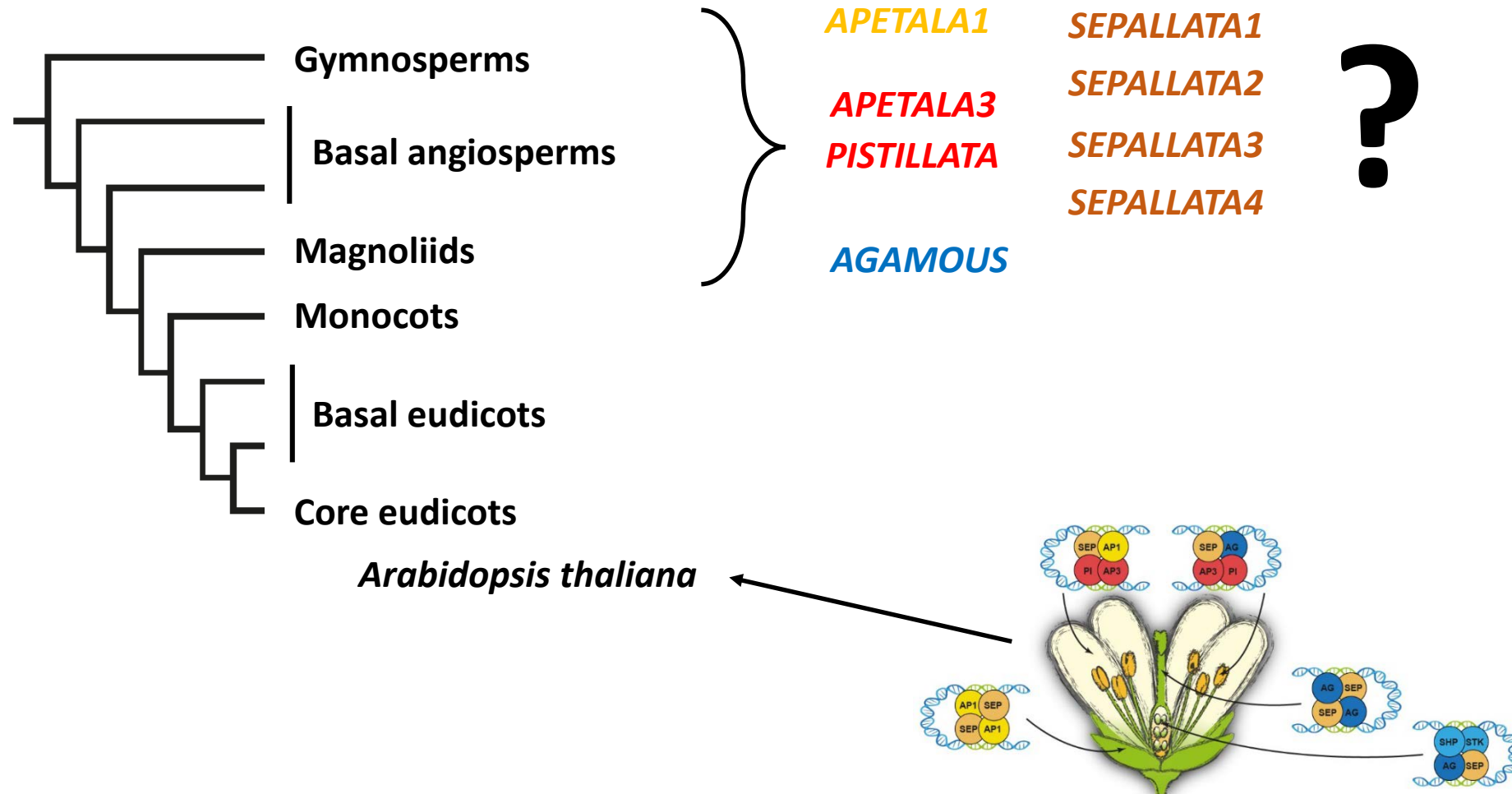


# The floral quartet model

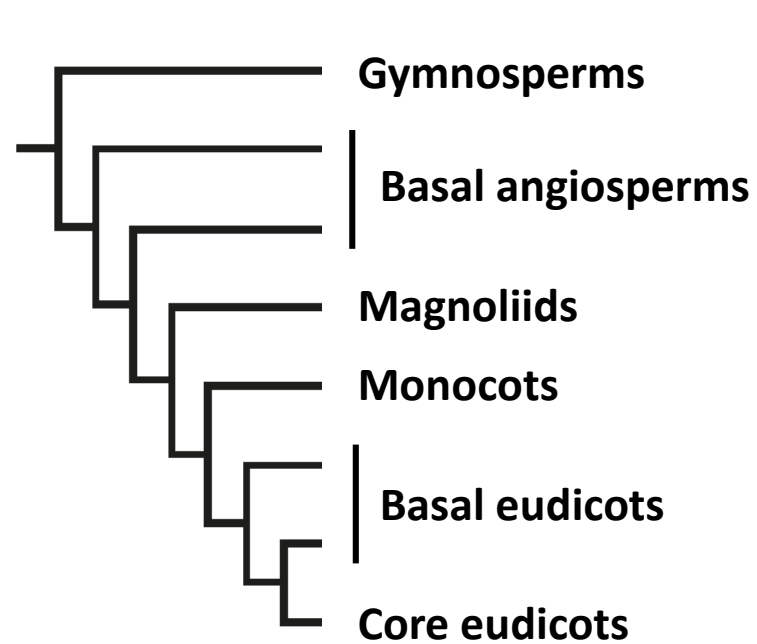




# Phylogeny of seed plants



# Phylogeny of seed plants



*Arabidopsis thaliana*

**APETALA1**

**APETALA3**

**PISTILLATA**

**AGAMOUS**

**SEPALLATA1**

**SEPALLATA2**

**SEPALLATA3**

**SEPALLATA4**

- Search for orthologs of floral homeotic genes in distantly related angiosperm and gymnosperm species

- Examine the phylogenetic relationship of the gene families

