

## Phylogeny

*Phylogeny:* An evolutionary tree.

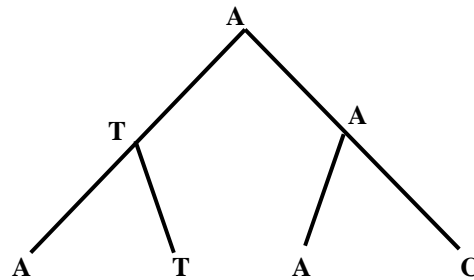
*A hypothesis* concerning the evolutionary history of a group of taxa

*Taxon:* a unit of classification

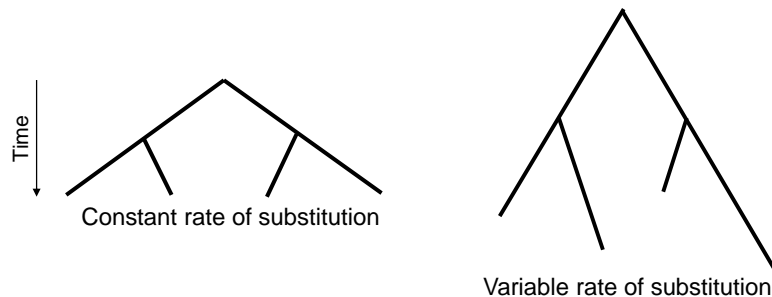
- strain, species, individual, gene
- also called OTUs  
(Operational Taxonomic Units)

## Properties of Trees

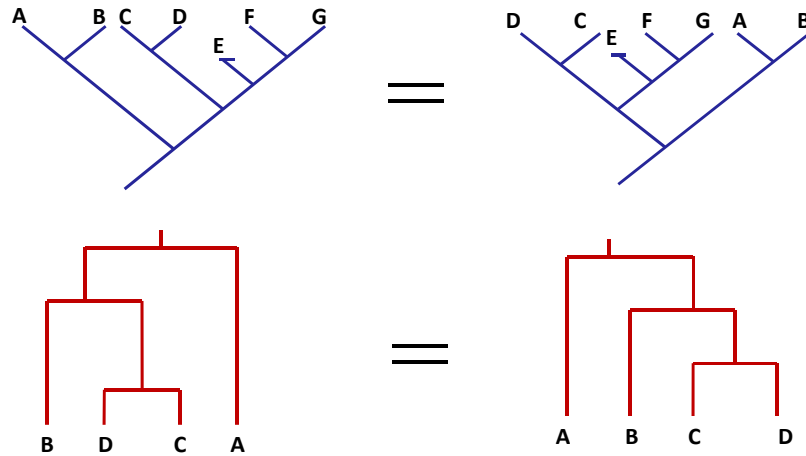
- Leaves– contemporary taxa
- (Internal) nodes – divergence of lineages
- Branch/edge - ancestral taxa
- Topology – relationships between taxa



- Branch lengths represent the amount of change
  - e.g., in substitutions per site
- If the mutation rate is constant in all lineages
  - the branch lengths are proportional to time,
  - the distance from the root to leaf is the same in all lineages
  - the data satisfy the “*molecular clock hypothesis*”.



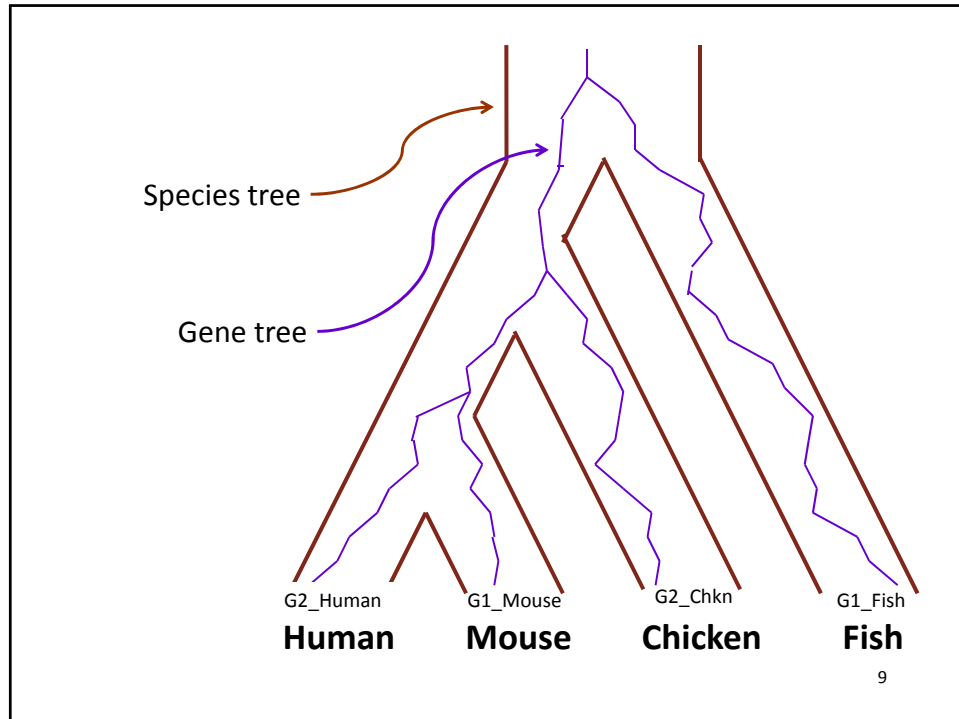
Trees represent the order of branching only



## Gene Trees vs Species Trees

Gene sequences can be used to

- infer the history of speciation
- infer the history of gene families



## Gene Trees vs Species Trees

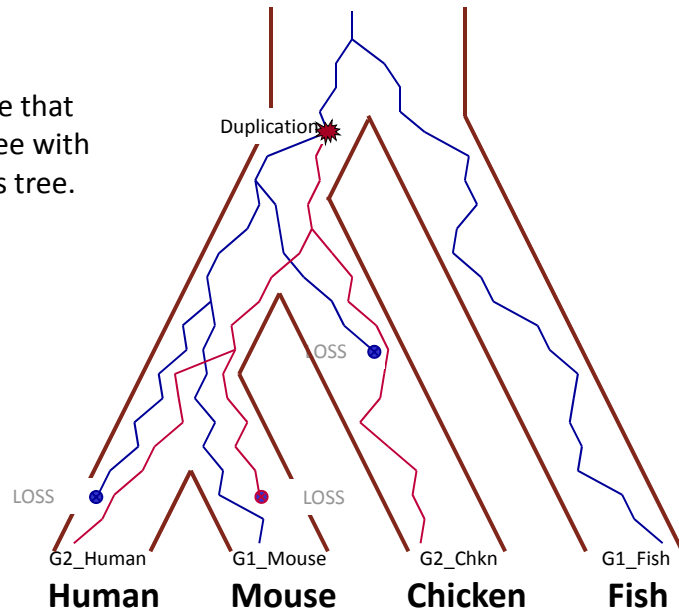
Gene sequences can be used to

- infer the history of speciation
- infer the history of gene families

*Caveat emptor:* the history of the gene, may not be the same as the history of the organism

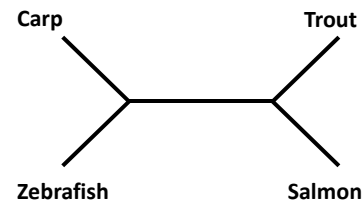
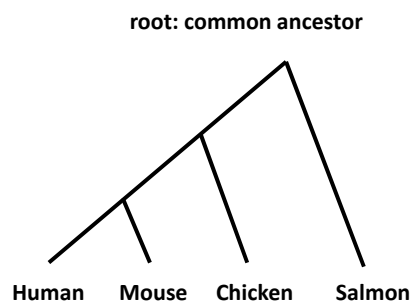
- gene duplications
- horizontal gene transfer

A gene tree that  
doesn't agree with  
the species tree.



11

Rooted vs unrooted trees:



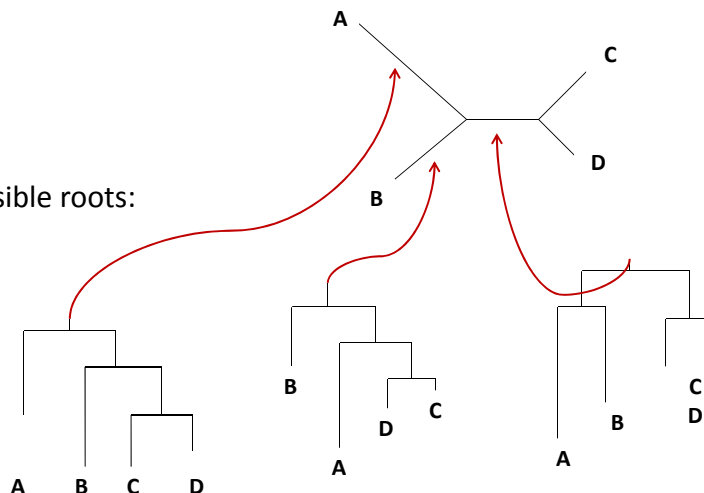
*Unrooted trees give no information about the order of  
speciation events*

## Unrooted vs. Rooted Trees

- An **unrooted** tree gives information about the relationships between taxa.
- A **rooted** gene tree gives information about the order of events.
- Most tree reconstruction programs output unrooted trees.

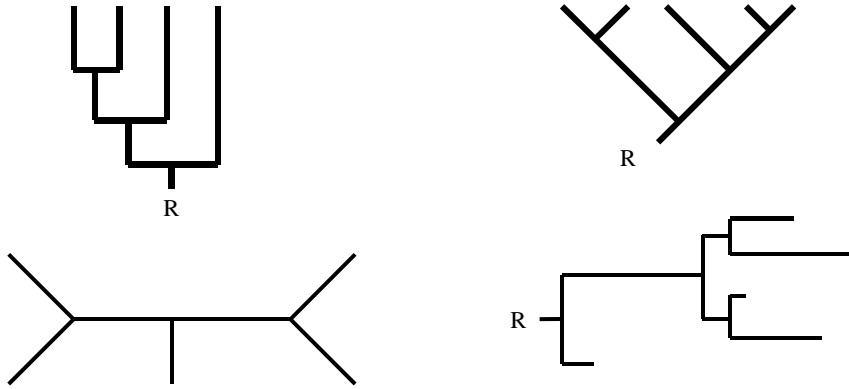
Unrooted tree: branch lengths proportional to evolutionary distance

Three possible roots:



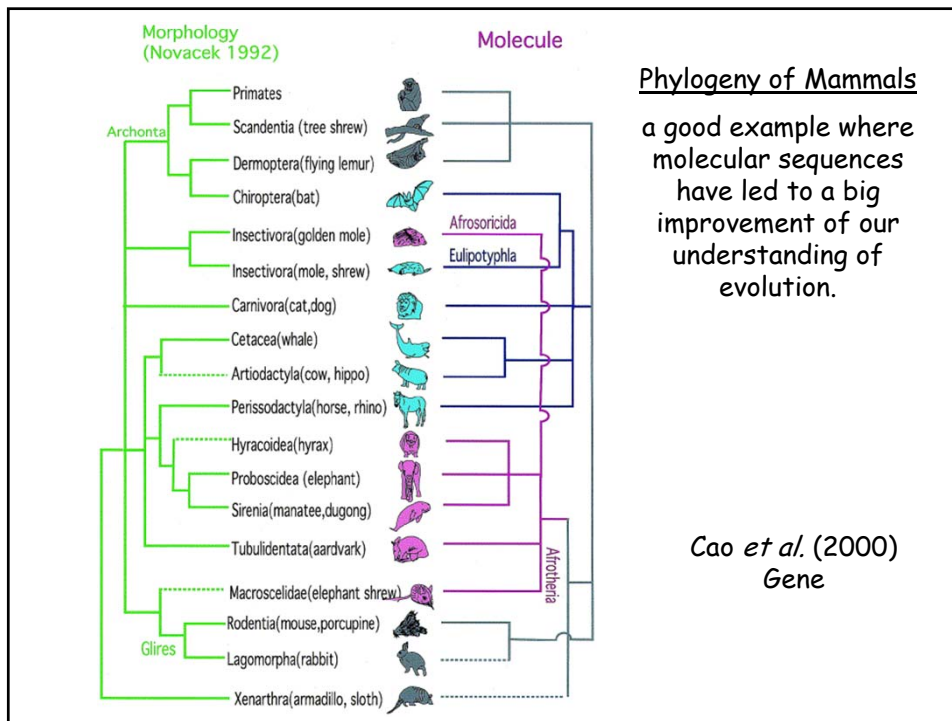
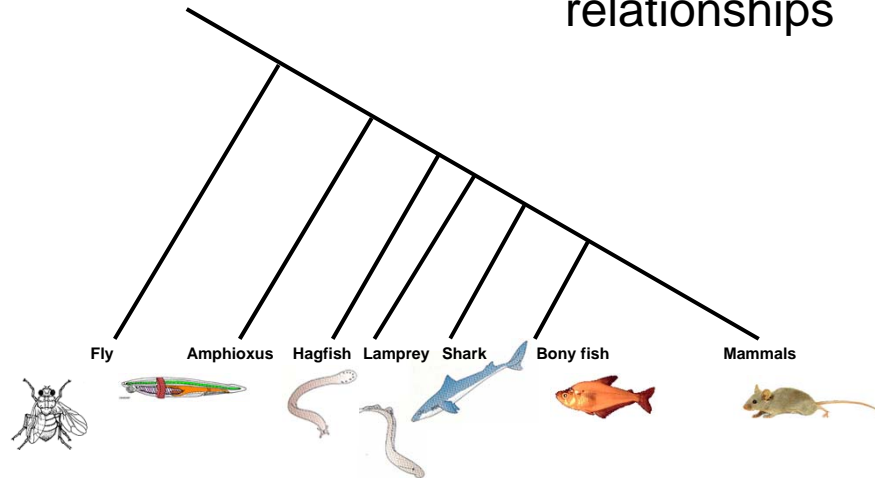
Most phylogenetic methods do not tell you where the root is.

## Tree representations

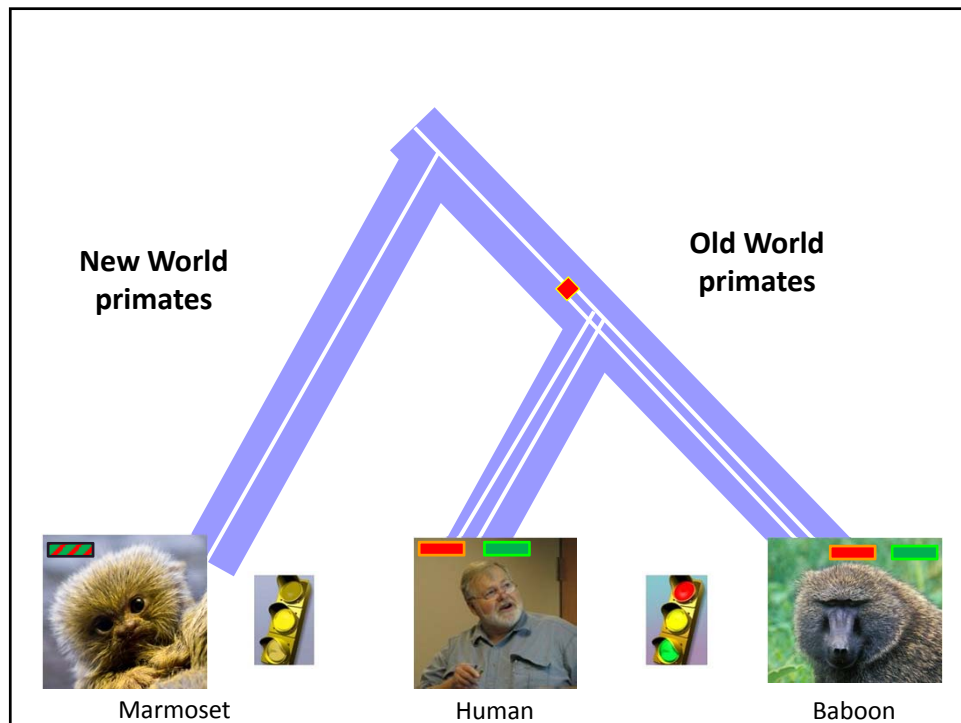
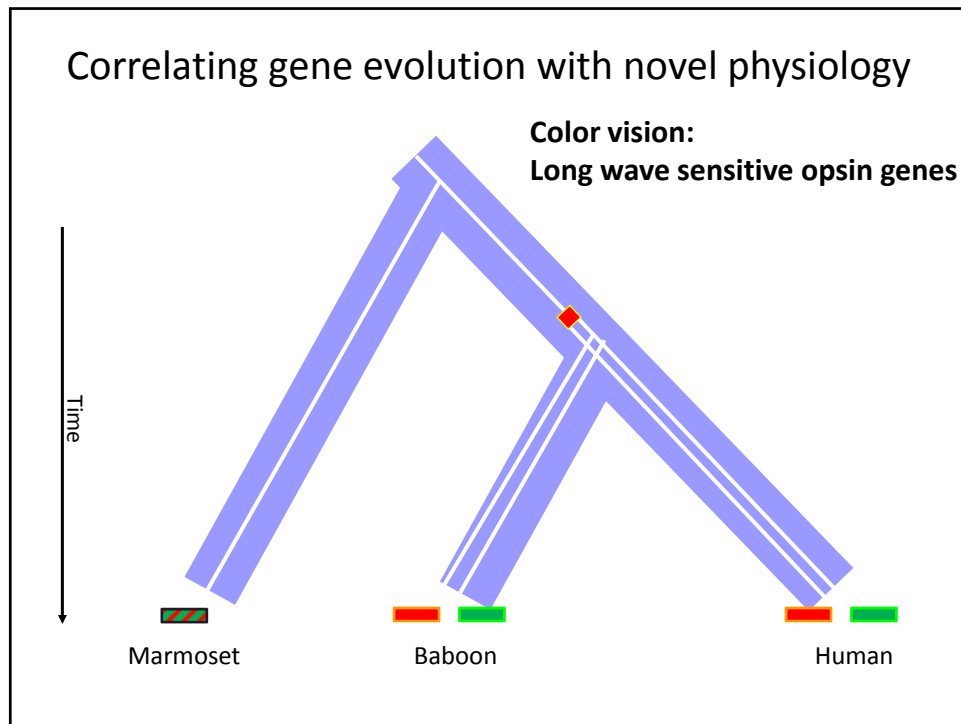


Why phylogeny reconstruction?  
Some applications

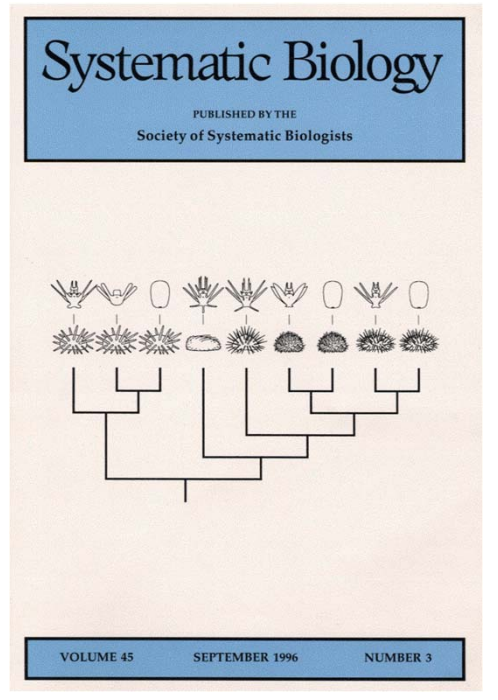
## Species relationships



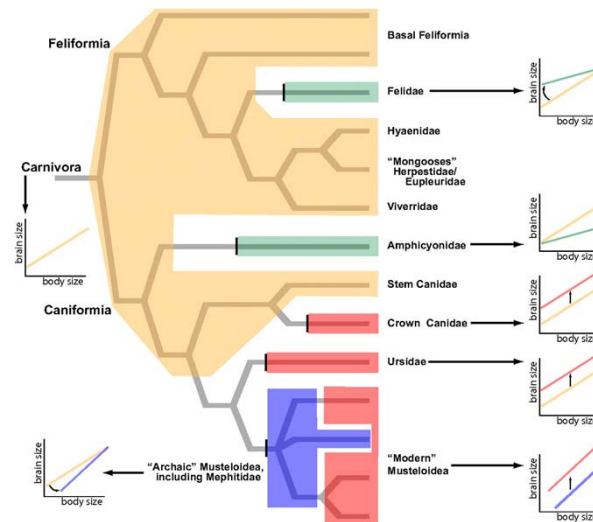




## Character evolution



The evolution of carnivoran encephalization mapped onto the branching pattern of the phylogeny presented in Fig.



Finarelli J A , Flynn J J PNAS 2009;106:9345-9349

©2009 by National Academy of Sciences

PNAS

## Biogeography

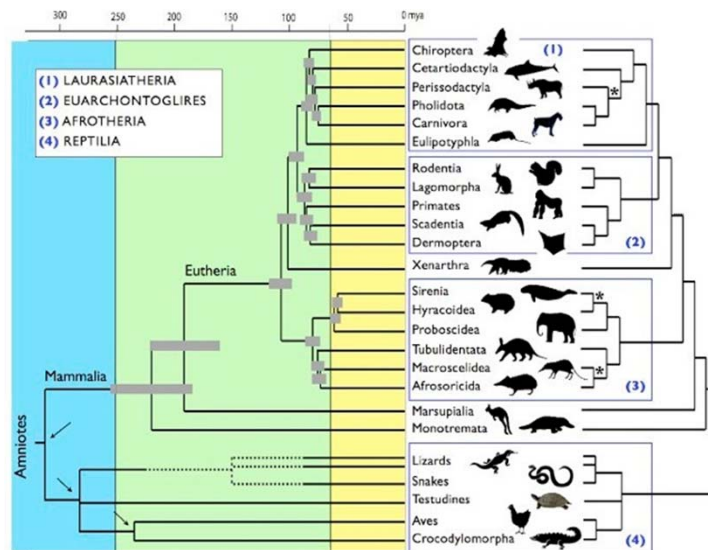
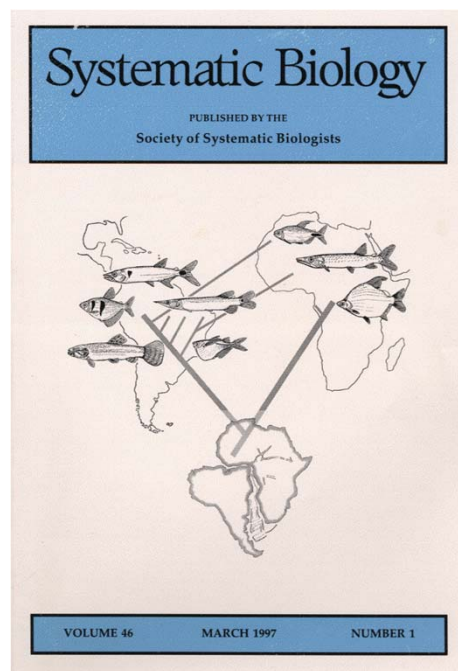
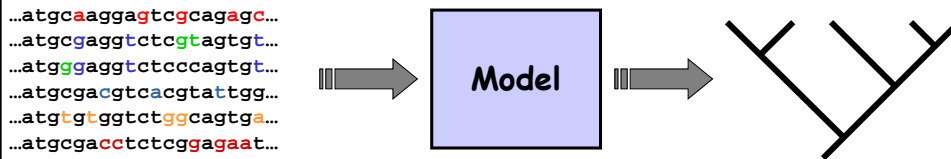


Figure from Tzika & Milinkovitch 2008

## Phylogeny reconstruction



Given

- Multiple sequence alignment
- Model of sequence evolution

find the (binary) tree that is the best explains the data with respect to the model.

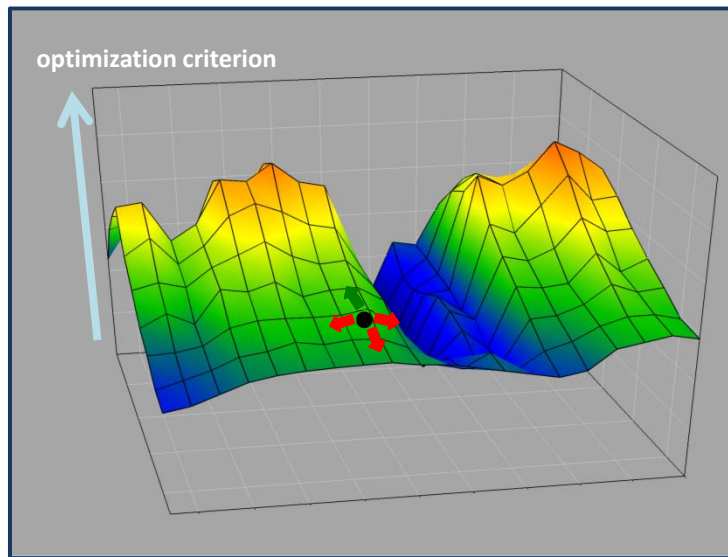
## Finding the optimal tree

Given  $k$  taxa,

- Consider all trees with  $k$  leaves
- Score each tree with respect to chosen optimization criterion.
- Select the optimal tree(s)

Phylogeny reconstruction is *NP*-complete:

Except in special cases when the data obeys specific constraints, the only way to find the best tree is to consider all trees.



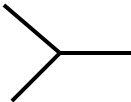
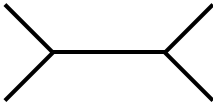
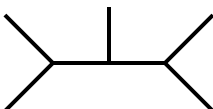
## Finding the optimal tree

Given  $k$  taxa,

- Consider all trees with  $k$  leaves
  - Score each tree with respect to chosen evolutionary model.
  - Select highest scoring tree(s)

How many trees are there?

How many unrooted trees with  $k$  leaves?

		$k$	$E(k)$	$T(k)$
• Three taxa		3	3	1
• Four Taxa		4	5	3
• Five taxa		5	7	15
	.....			

Number of unrooted trees for  $k$  taxa

$$E(k) = E(k-1) + 2 = 2k - 3$$

$$T(k) = E(k-1)T(k-1) = \prod_{i=3}^{k-1} (2i-3)$$

$$T(k) = \frac{(2k-5)!}{2^{k-3}(k-3)!}$$

## The number of trees gets big fast

Number of leaves

Number of unrooted binary trees

3	1
4	3
5	15
6	105
10	2,027,025
20	$2.2 \times 10^{20}$
50	$2.8 \times 10^{74}$
500	$1 \times 10^{1074}$

## How big is that?

Number of leaves

Number of unrooted binary trees

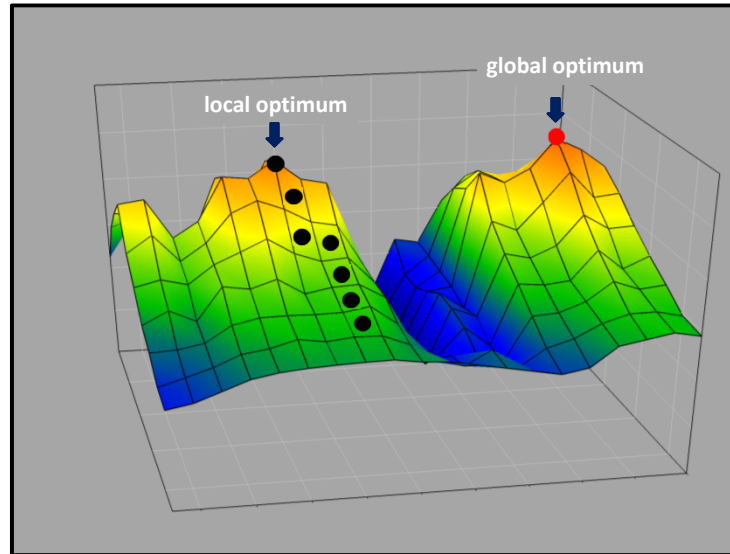
...	...
20	$2.2 \times 10^{20}$
50	$2.8 \times 10^{74}$
500	$1 \times 10^{1074}$

Age of the universe (seconds):  $4.42 \times 10^{17}$

Diameter of the universe:  $2.70 \times 10^{10}$

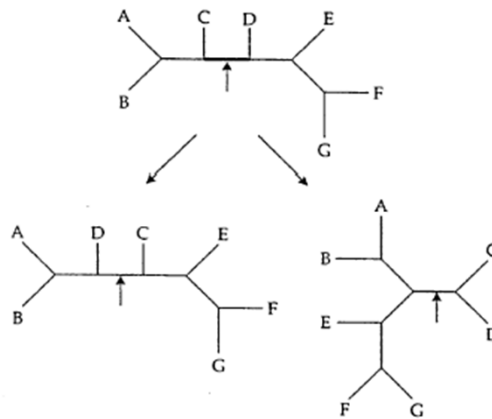
Number of stars in the universe:  $10^{22}$

## Phylogeny reconstruction uses heuristic search



## Traversing tree space

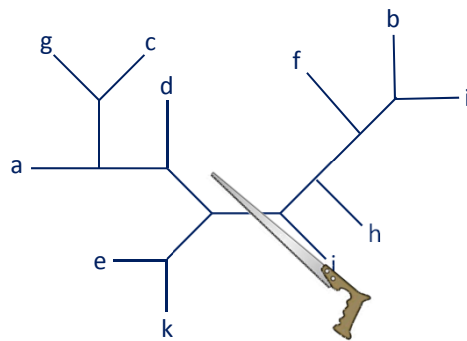
Nearest-neighbor interchange (NNI)





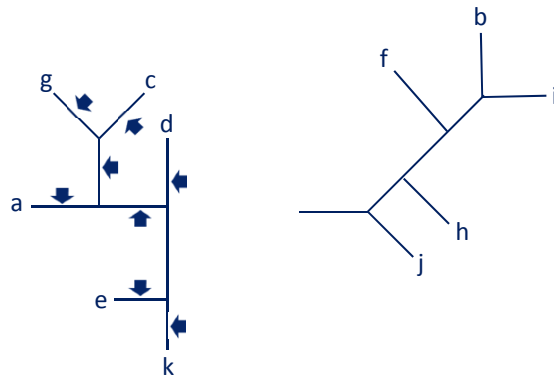
## Traversing tree space

Subtree pruning and regrafting (SPR)



## Traversing tree space

Subtree pruning and regrafting (SPR)



# Traversing tree space

Subtree pruning and regrafting (SPR)

