

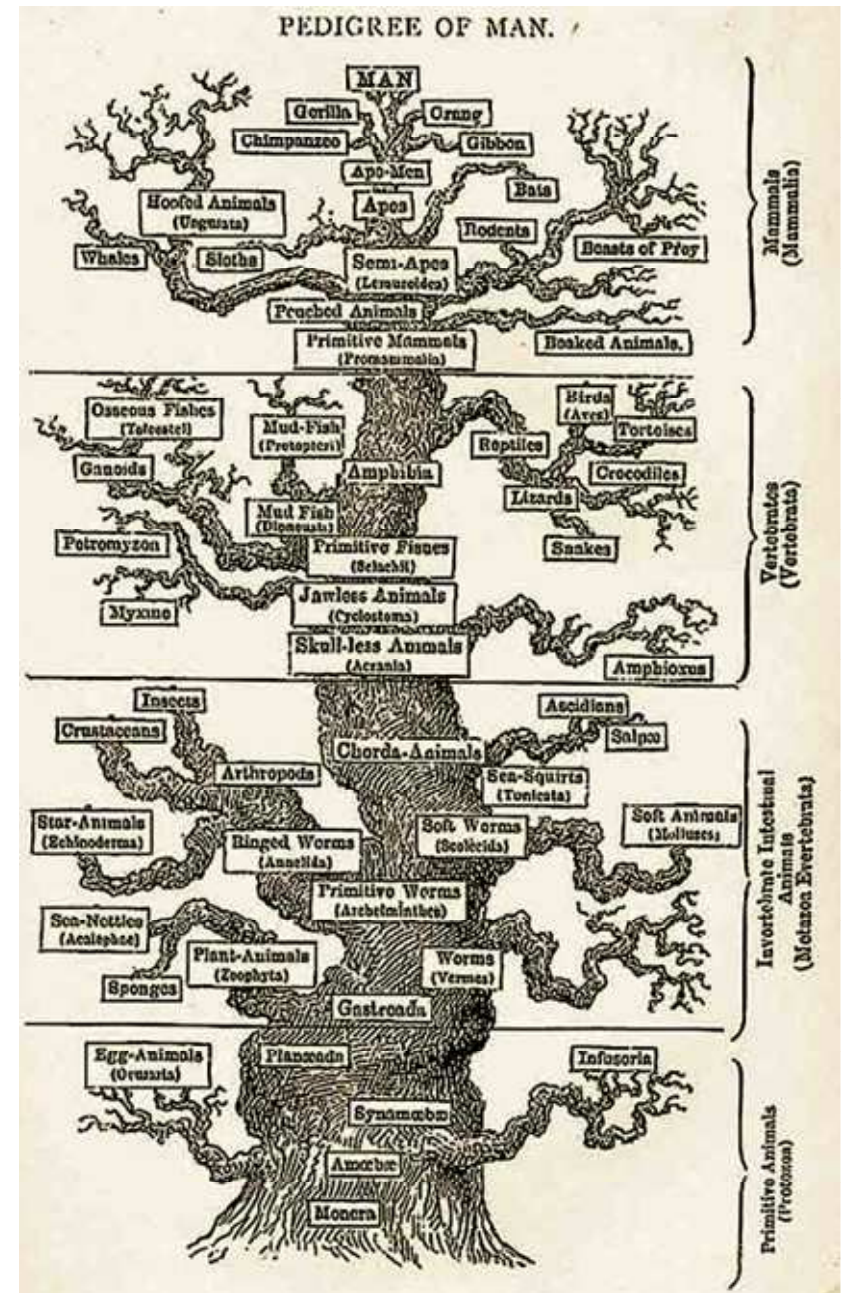


**How to read and make phylogenetic trees**  
**Zuzana Starostová**



**phylogeny** – the evolutionary history of a species or group of related species

- the process of evolution produces a pattern of relationships between species - as lineages evolve and split and modifications are inherited, their evolutionary paths diverge
- this produces a branching pattern of evolutionary relationships
- each lineage has a part of its history that is unique to it alone and parts that are shared with other lineages



# Why is phylogeny important?

- understanding and classifying the diversity of life

We use phylogenetic trees for:

- comparative analysis and character evolution
- biogeography
- dating – age of different taxa
- genetic engineering
- disease epidemiology
- conservation

...



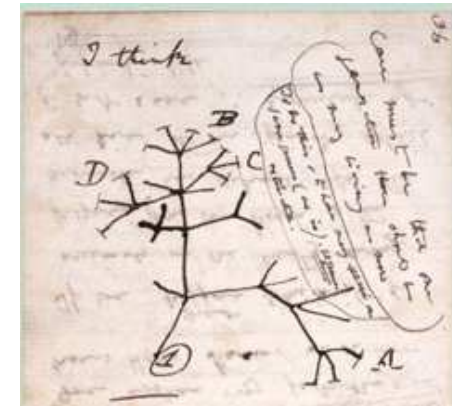
**The Tree of Life Web Project (ToL)** is a collaborative effort of biologists and nature enthusiasts from around the world. On more than 10,000 World Wide Web pages, the project provides information about biodiversity, the characteristics of different groups of organisms, and their evolutionary history.

**<http://tolweb.org/tree/>**

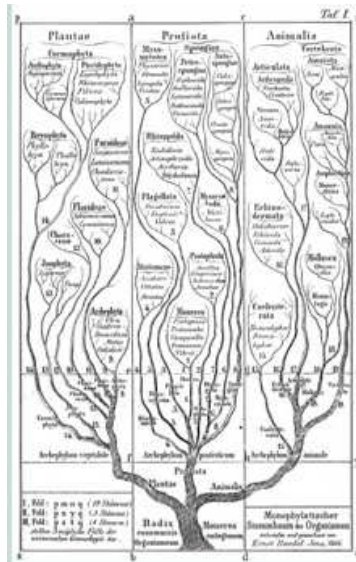
# Tree of life



**Charles Darwin**  
Relationships among species compared to  
“the great Tree of Life”

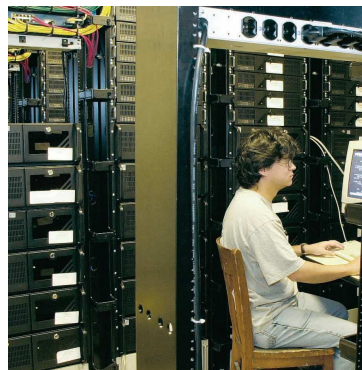
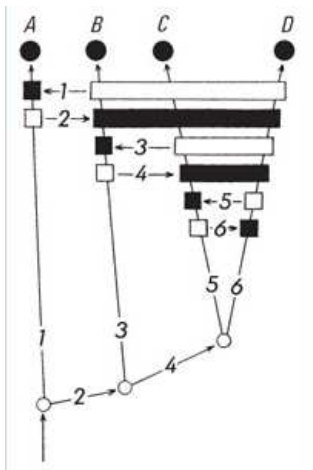


Darwin's diary, 1837

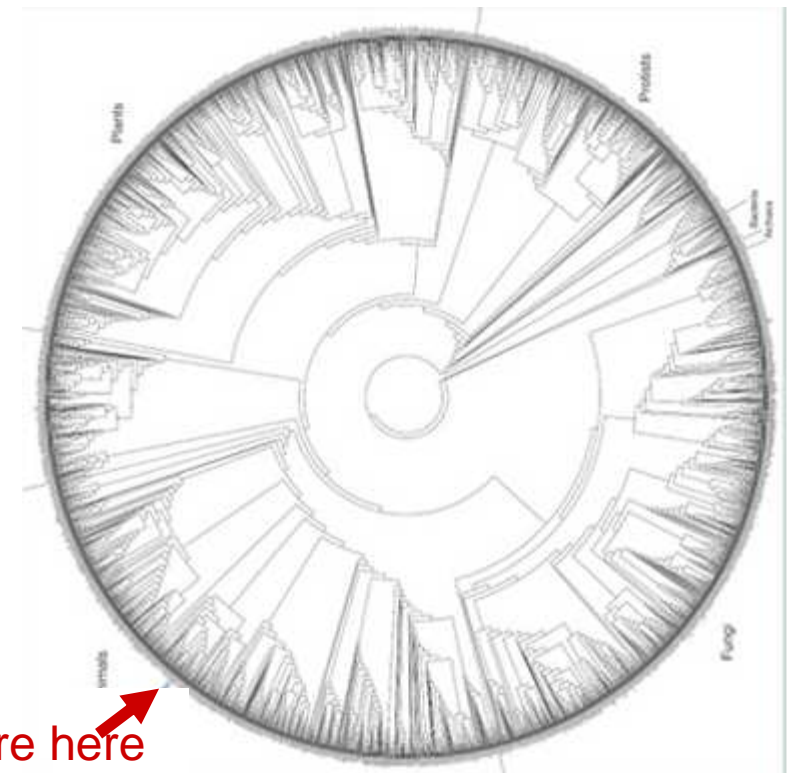


**Ernst Haeckel – 1866**  
diagram of relationships among species  
based on general similarity

**Willi Hennig – 1960s**  
foundation of modern  
phylogenetics

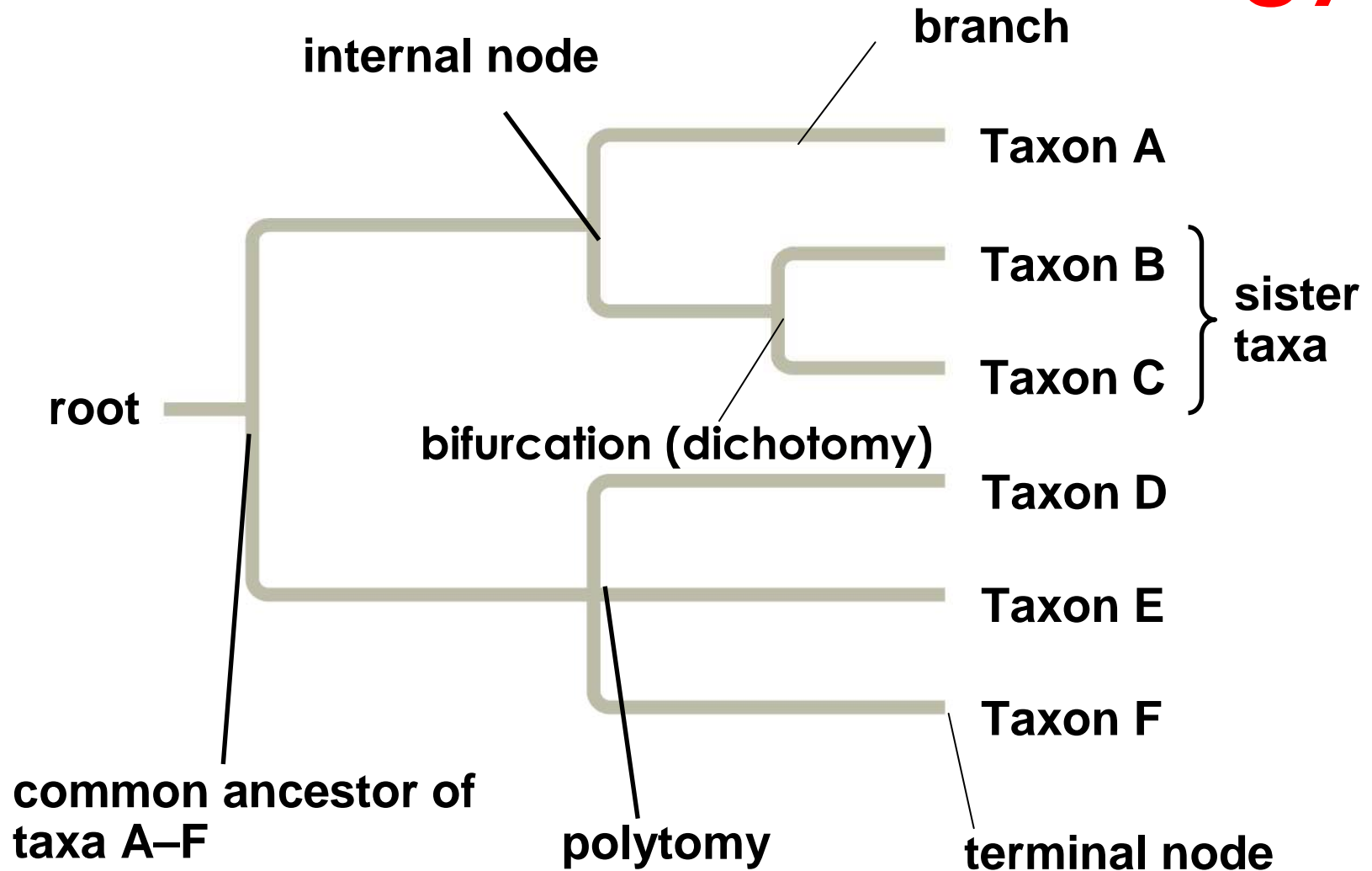


you are here



<http://www.peabody.yale.edu/exhibits/treeoflife/index.html>

# how to read trees - terminology



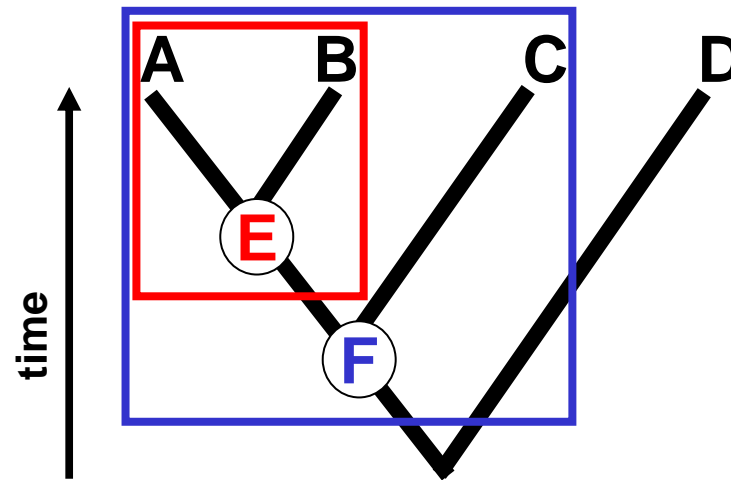
- root of the tree represents the ancestral lineage, and the tips of the branches represent the descendants of that ancestor
- as one moves from the root to the tips = moving forward in time



# phylogenetic tree

branching diagram showing relationships between taxa based on their shared common ancestors

taxa (e.g.:species):



**A and B are most closely related because they share a common ancestor (here “E”) that C and D do not share**

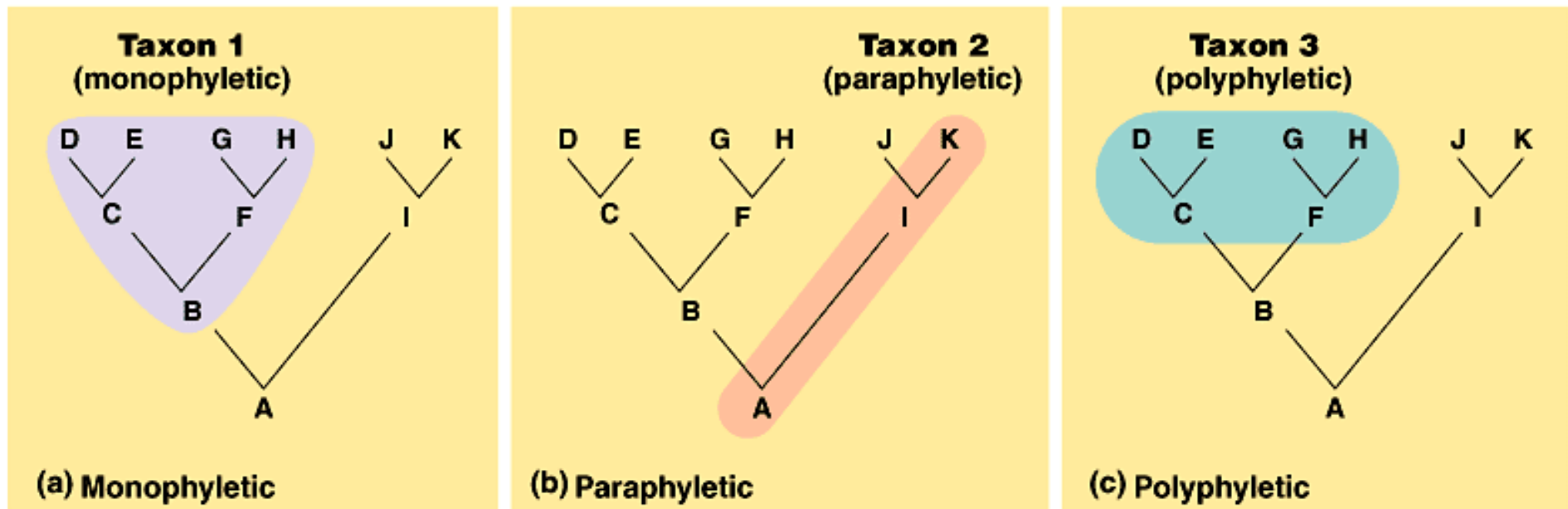
**A+B+C are more closely related to each other than to D because they share a common ancestor (“F”) that D does not share**

# Phylogeny and classification

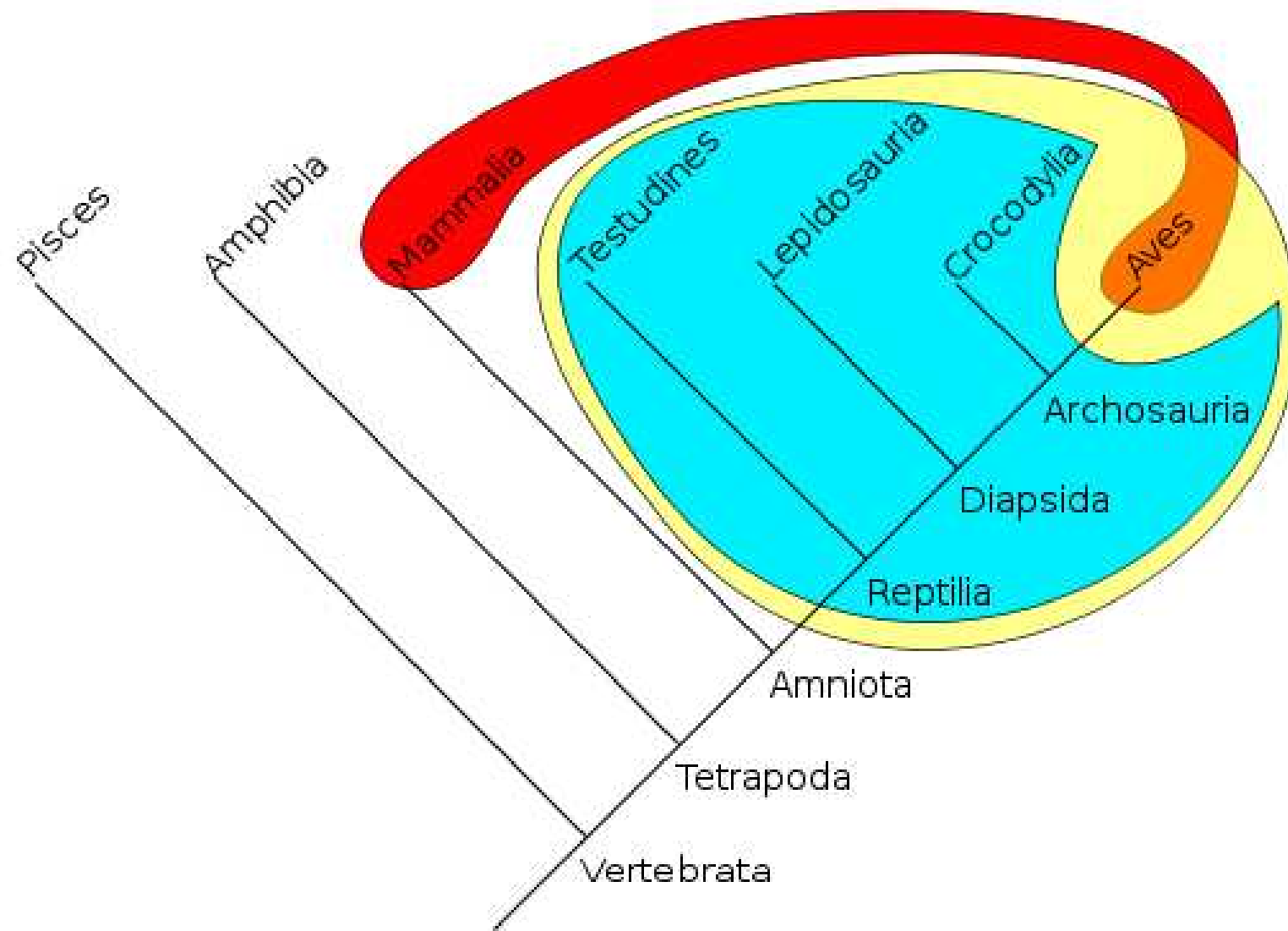
phylogenetic (cladistic) classification reflects evolutionary history

the only valid group for classification is **monophyletic**

**group = clade** = group that includes a common ancestor and all the descendants (living and extinct) of that ancestor.



- Monophyly
- Paraphyly
- Polyphyly





# rooted vs. unrooted phylogenetic trees

rooted tree – root leads to the common ancestor of all studied taxa (e.g.: species)

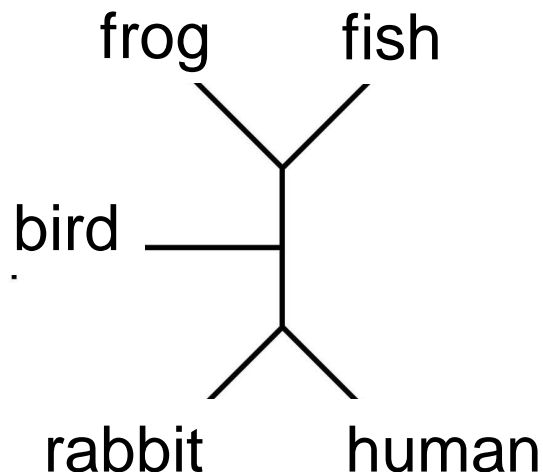
rooting the tree = indicating the direction of the evolutionary process

helps determine what is plesiomorphic and apomorphic

## How to root a tree?

### - introducing outgroup

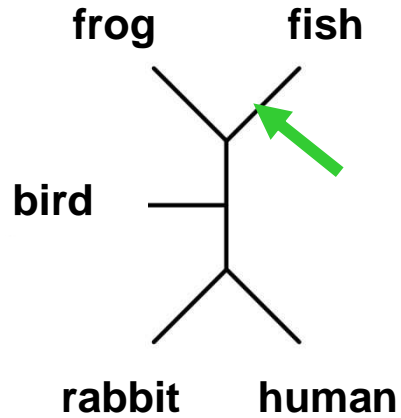
**outgroup** – a species or a set of species that are definitely distant from all the species of interest (ingroup), but still closely related



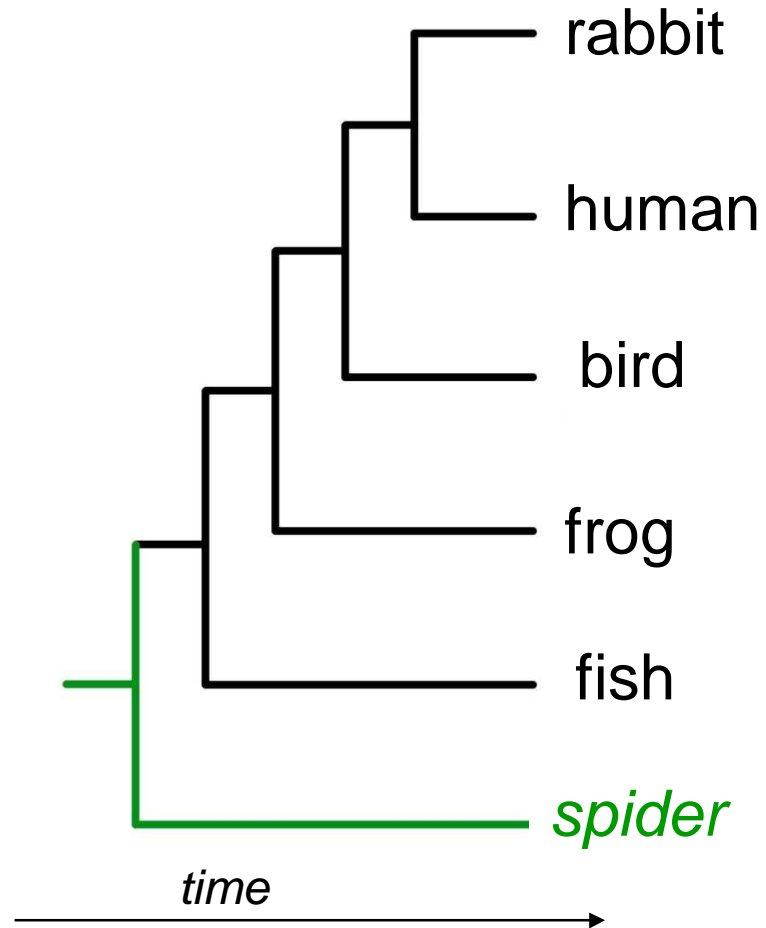
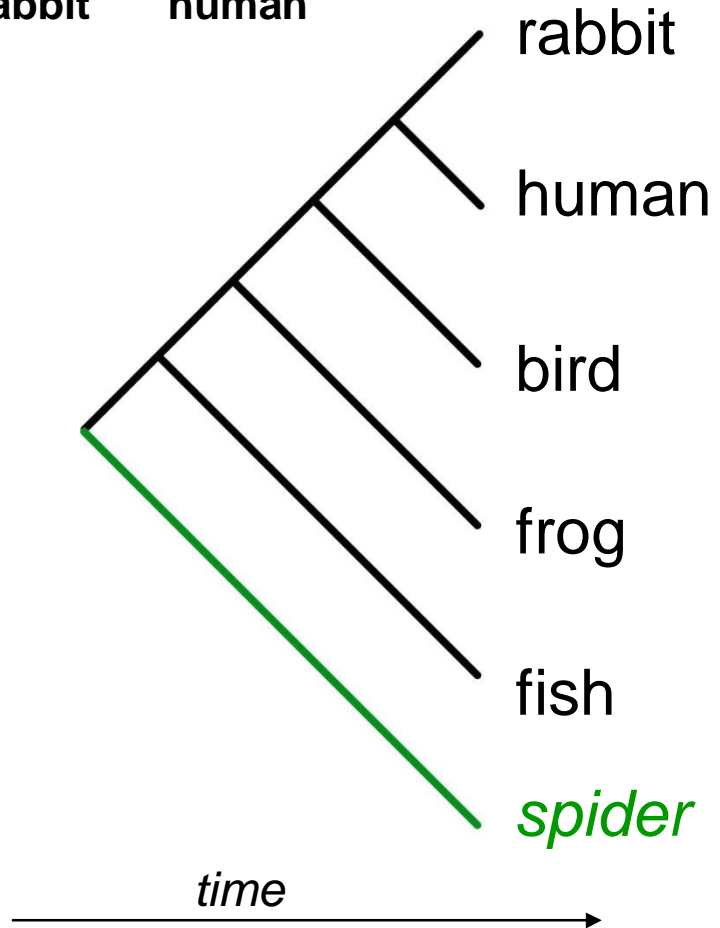
**we will use an invertebrate species to root the tree**

(since these species are all Craniata it would be better to use Urochordata (tunicates) to root the tree)

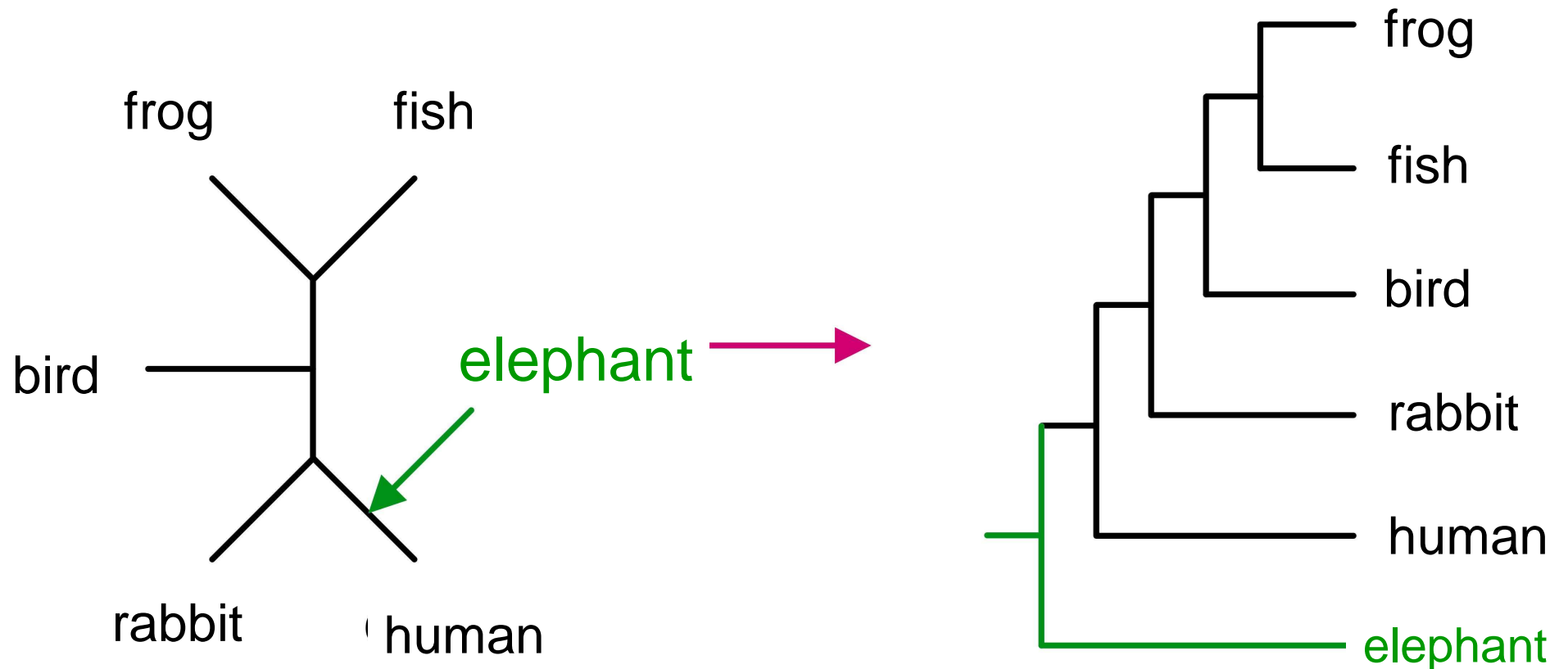
unrooted tree



we place the root to the middle of the branch connecting **outgroup** with the rest of the tree

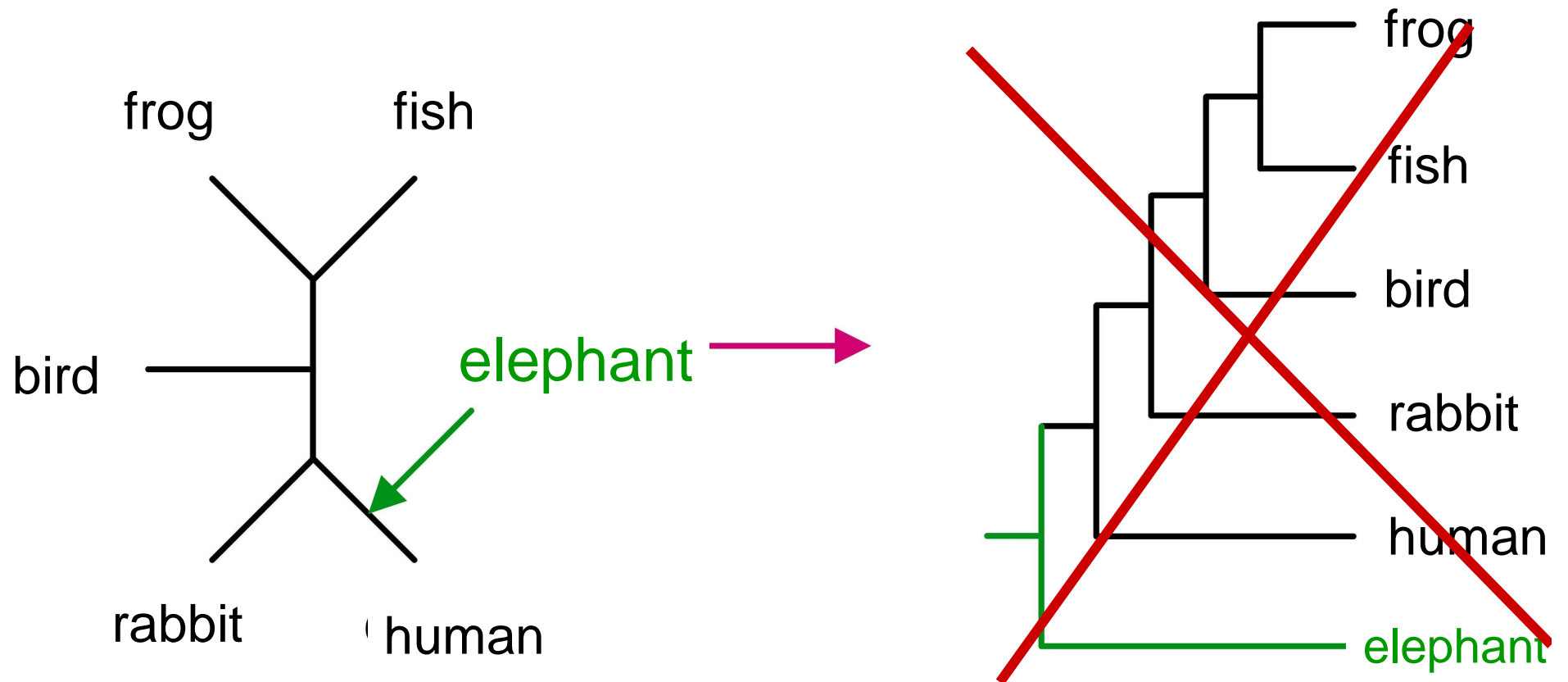


## But what if our outgroup is wrong?





## But what if our outgroup is wrong?



# rooted vs. unrooted phylogenetic trees

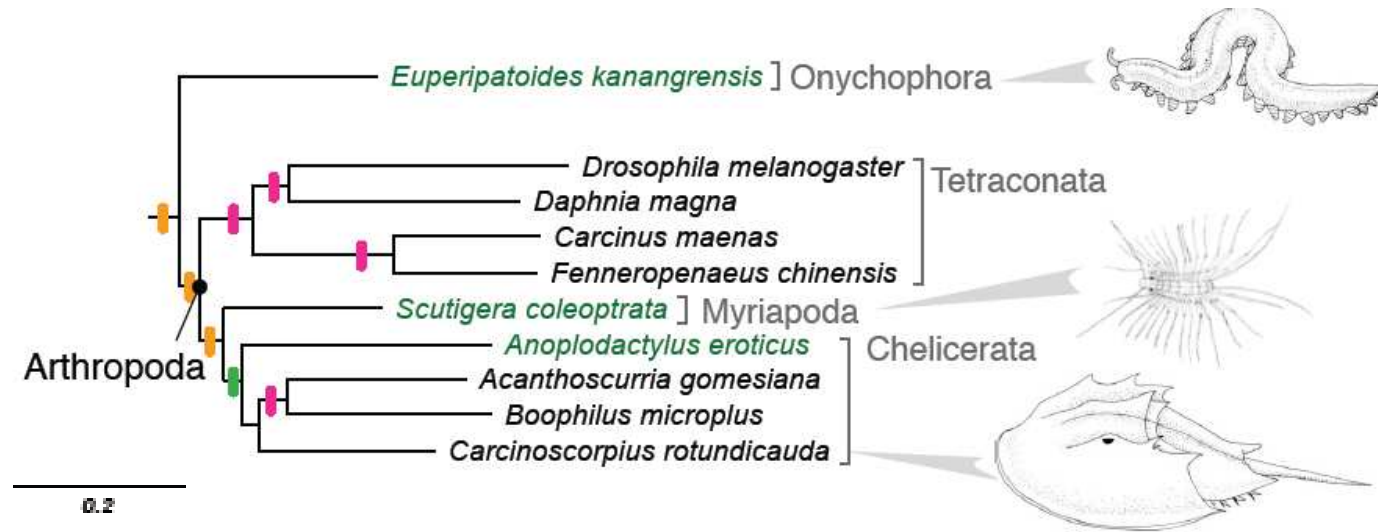
$$(2n - 3)!$$

$2^{n-2}(n-2)!$  Number of bifurcating rooted trees for  $n$  taxa (OTUs)

OTU = operational taxonomical unit

Number of bifurcating trees is increasing with number of taxa. The number of unrooted trees for  $n$  taxa is equal for the number of rooted trees for  $(n-1)$  taxa.

no. of taxa	no. of unrooted trees	no. of rooted trees
4	3	15
8	10 395	135 135
10	2 027 025	34 459 425
22	$3 \times 10^{23}$	
50	$3 \times 10^{74}$	
100	$2 \times 10^{182}$	



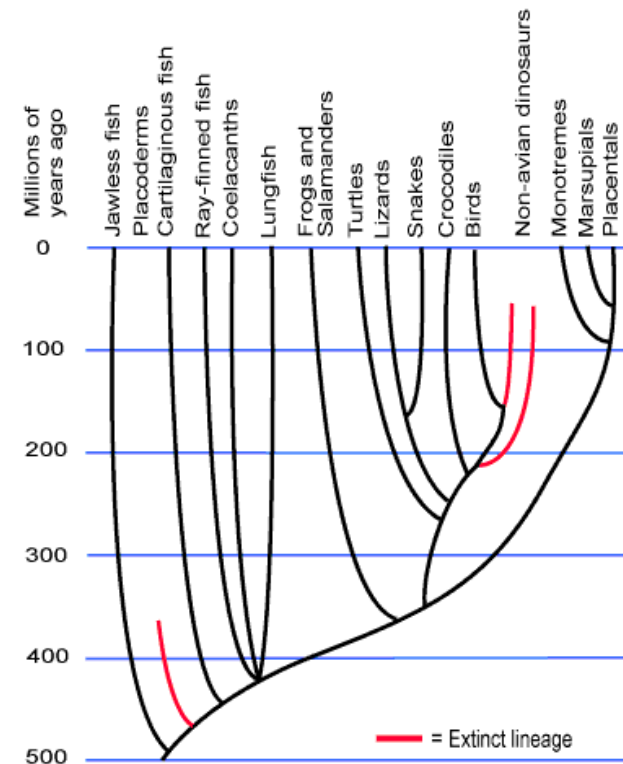
**branch length** = time of divergence from the ancestor

molecular clock: sequence divergence increases over time linearly

when molecular clock holds (accumulation is linear over time) – all lineages in the tree have accumulated substitutions at the same rate

evolutionary rate dependent on metabolic rate, generation time, bottleneck events,...

we need calibration points (fossils, geological events)





# How to make trees?

like family trees, phylogenetic trees represent patterns of ancestry

to reveal phylogenetic relationship, we have to compare characters which are **inherited from a common ancestor**

→ **homologous characters** ( x analogous characters are result of convergent evolution)

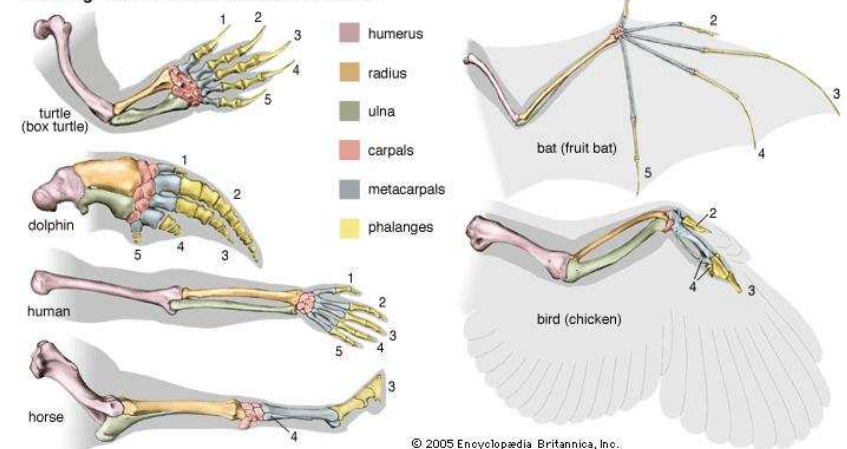
**character** = heritable traits that can be compared across organisms

two or more forms = **character state**

types: physical characteristics (morphology), genetic sequences, and behavioral traits

		10	20	30	40	50	60
Ael		GCCCAGTCATTAACTTCGAC	GAT	A	TACTACTCCCT	CGTCCGCCAGAGAAC	6
CB	9	GCTTAGCCCTAAACTAAGAT	AC	AT	TCAC G	TACA ATCCGCCAGAGCAC	
CE	9	GCTTAGCCCTAAACTTCGATTC	CCGA	ATAC A	AACA ATCCGCCAGAGCAC		
CM	9	GCTTAGCCCTAAACTTCGATATCTACC		ATA A	AATA TCCGCCAGAGAAC		
CV	9	GCTAGGCCATTAACTTAGATGTAC	AA	AAAC A	TACA ATCCGCCAGAGTAC		
Cyr		GCTTAGCCGTAAACCCCTGAT	AGT	CAA	ACAACT	ATCCGCCAGAGAAC	
FAI		GCTTAGGCCATTAACTTAGGC	ACA	ATA	TAAAC	CACA TGCCGCCAGAGAAC	
EAS		GCTTAGGCCATTAACTTAGGC	ACA	ATA	TAAAC	CACA TGCCGCCAGAGAAC	
EM	9	GCTTAGGCCATTAACTTAGGC	ACA	ACA	TAAAC	TGCA TGCCGCCAGAGAAC	
EMS		GCTTAGGCCATTAACTTAGGC	ACA	ACA	TAAAC	TGTA TGCCGCCAGAGAAC	
EMT		GCTTAGGCCATTAACTTAGGC	ATA	ACA	TAAAC	TGTA TGCCGCCAGAGAAC	
EMV		GCTTAGGCCATTAACTTAGGC	ACA	ACA	TAAAC	TGCA TGCCGCCAGAGAAC	
GA	9	GCTTAGGCCATTAACTTAGGC	ATA	TATC	CCA	TAC A AATA TATTCGCCAGAGAAC	
GG	9	GCCAGACGTAAACCTTGTATAGAAC	CA	TATTA	CAATC	TATTCGCCAGAGAAC	
GKKJ		GCTTAACCATTAACCAAGAT	AAT	TCA	AC A	CCTA CCTCGGCCAGAGAAC	
GKKS		GCTTAGCCATTAACTTAGGC	A	TCA	AC A	CCCA CCTCGGCCAGAGAAC	
GRO		GCTTAACCATTAACCAAGAT	AAT	TCA	AC A	CCTA CCTCGGCCAGAGAAC	
GKS		GCTTAGCCATTAACTTAGGC	ATT	TCA	AC A	CCCA ACTCGGCCAGAGAAC	
GLI		GCTTAGCCATTAACTTAGGC	TAT	TCTA	CAC	TAC A AATA TATTCGCCAGAGAAC	
GLU		GCTTAGCCATTAACTTAGGC	TAT	TCTA	TTC	TAC AC AATA TATTCGCCAGAGAAC	
Hem		GCTTAGCCATTAACTTAGGC	CCA	CTAC	A	CATG GGCCGCCAGAGAAC	
hemta		GCTTAGCCATTAACTTAGGC	CCA	CTAC	A	AAGC GGCCGCCAGAGAAC	
Hol		GCTTAGCCATTAACTTAGGC	ACC	TAAAA	CTACA	CATA GGCCGCCAGAGAAC	
Rha		GCTTAGCCATTAACTTAGGC	ATT	TAAA	ACCA	CACAA TGCCGCCAGAGAAC	

Homologies of the forelimb in six vertebrates



© 2005 Encyclopedia Britannica, Inc.



## Pros and cons of molecular characters (mainly sequence data)

- molecular data is much more abundant (human genome 3,1 Gb, *E. coli* 4.6 Mb)
- independent (one position in the sequence of nucleotides x an eye – the eye is missing, but it means that also cornea, retina, etc. are missing)
- easy to describe (A, C, T, G at the position 175 in the cytochrome b gene) x some structure on the bone more pointed
- can resolve relationships at all different levels of organization, from species and populations to phyla and kingdoms
- less subjective
- neutral – number of shared characters mirrors phylogenetic relationship not just the same environmental selection pressures



## Pros and cons of molecular characters (mainly sequence data)

- expensive
- technically demanding
- also non-experts on a given group can use them (lack of insight)
- no information about phenotype

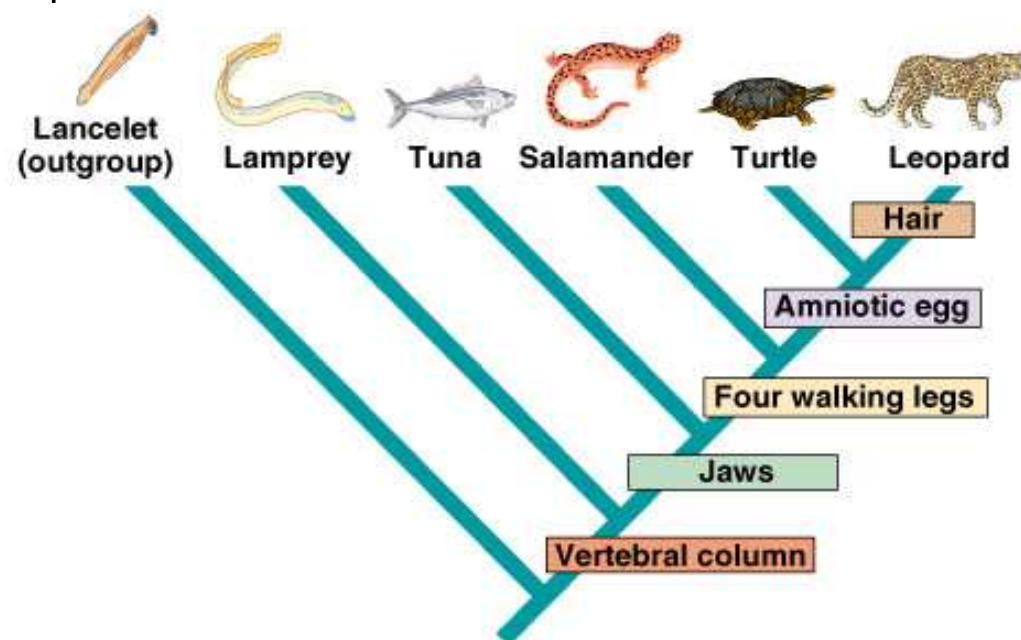


# How to make a tree - morphological example

- selected species have shared primitive and derived characters
- for reconstructing phylogeny derived shared characters = **synapomorphies** are important
- create a character table (matrix) with variable characters
- group the taxa based on synapomorphies - the more shared characters, the more closely related are the species
- for morphological data we use usually maximum parsimony method which prefers the simplest explanation of observed data

CHARACTERS	TAXA					
	Lancelet (outgroup)	Lamprey	Tuna	Salamander	Turtle	Leopard
Hair	0	0	0	0	0	1
Amniotic (shelled) egg	0	0	0	0	1	1
Four walking legs	0	0	0	1	1	1
Jaws	0	0	1	1	1	1
Vertebral column (backbone)	0	1	1	1	1	1

(a) Character table



(b) Cladogram

# molecular characters

Outgroup

Taxon A

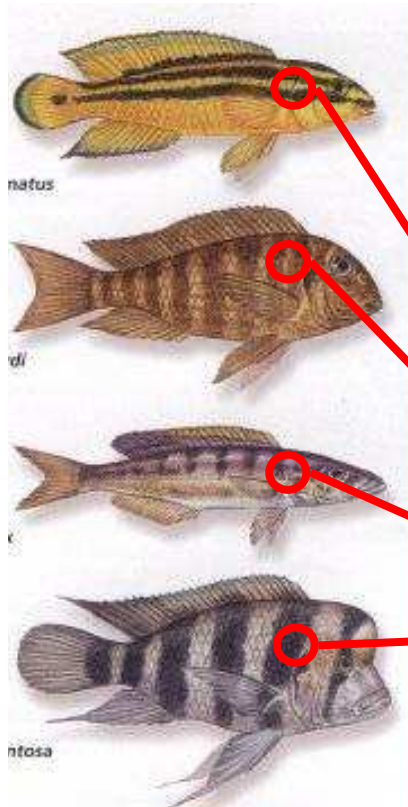
Taxon B

Taxon C

```

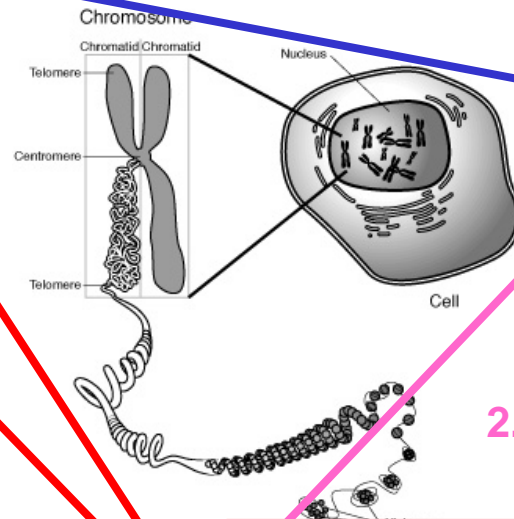
AAGCTTCATAGGAGCAACCATTCCTAATAATAAGCCTCATAAAGCC
AAGCTTCACCGGCGCAGTTATCCTCATAATATGCCTCATAATGCC
GTGCTTCACCGACGCAGTTGTCCTCATAATGTGCCTCACTATGCC
GTGCTTCACCGACGCAGTTGCCCTCATGATGAGCCTCACTATGCA
    
```

3. alignment

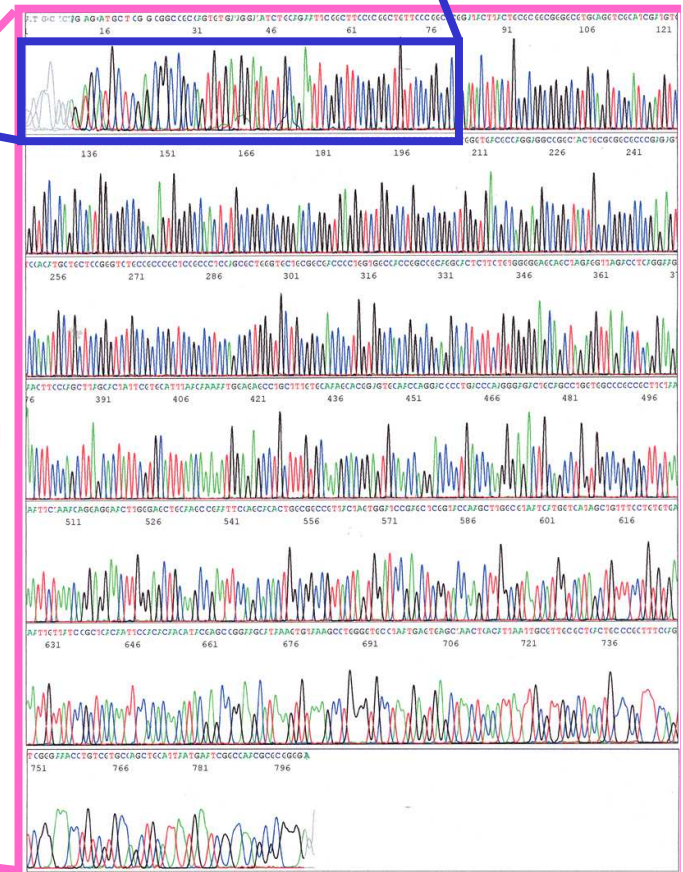
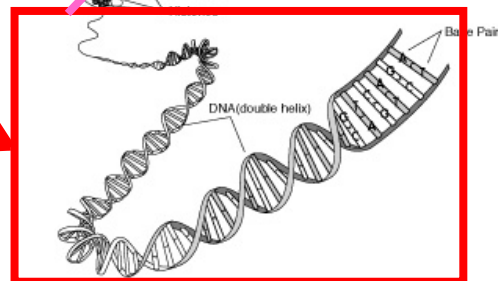


1. DNA  
isolation

<http://www.access Excellence.org/AB/GG/chromosome.html>



2. sequencing



# molecular characters

Outgroup

Taxon A

Taxon B

Taxon C

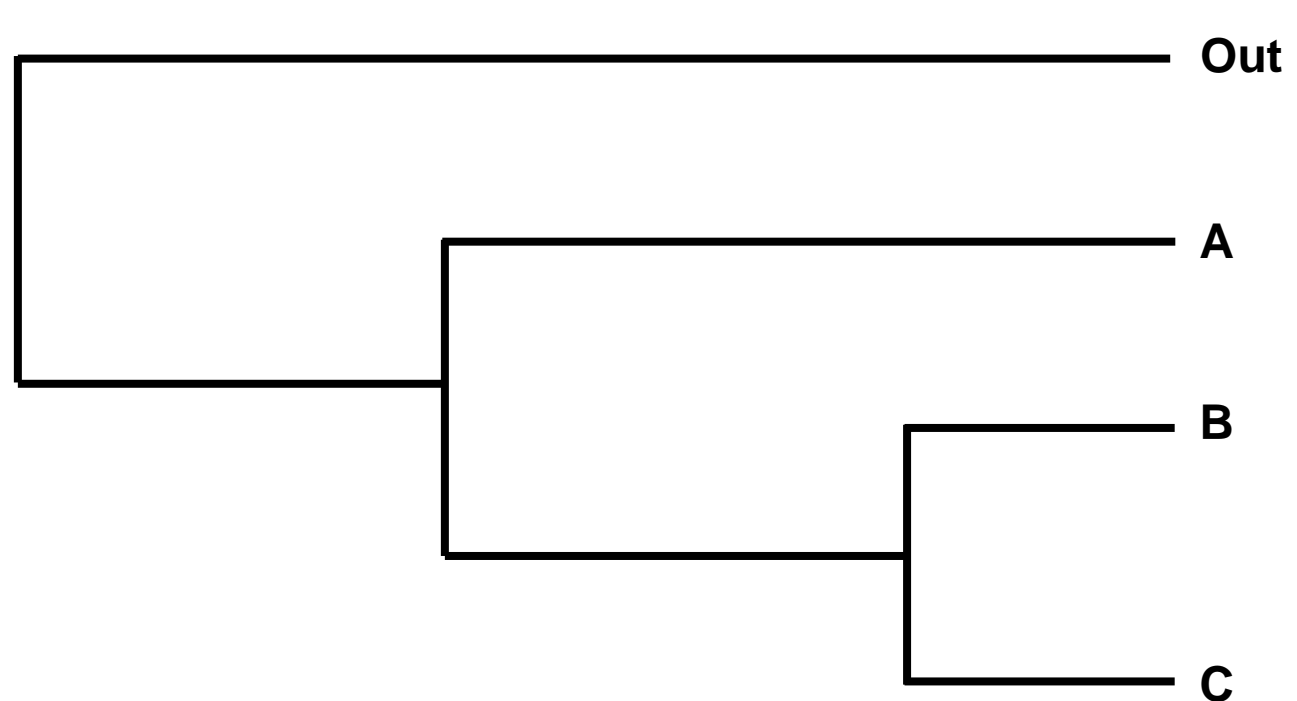
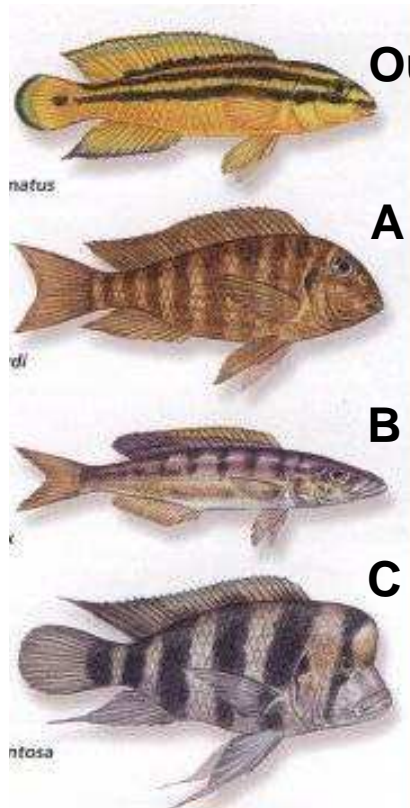
AAGCTTCA~~TA~~

invariant sites

GAGCTTCA~~CA~~

GTGCTTCA~~CG~~

GTGCTTCA~~CG~~



# molecular characters

Outgroup

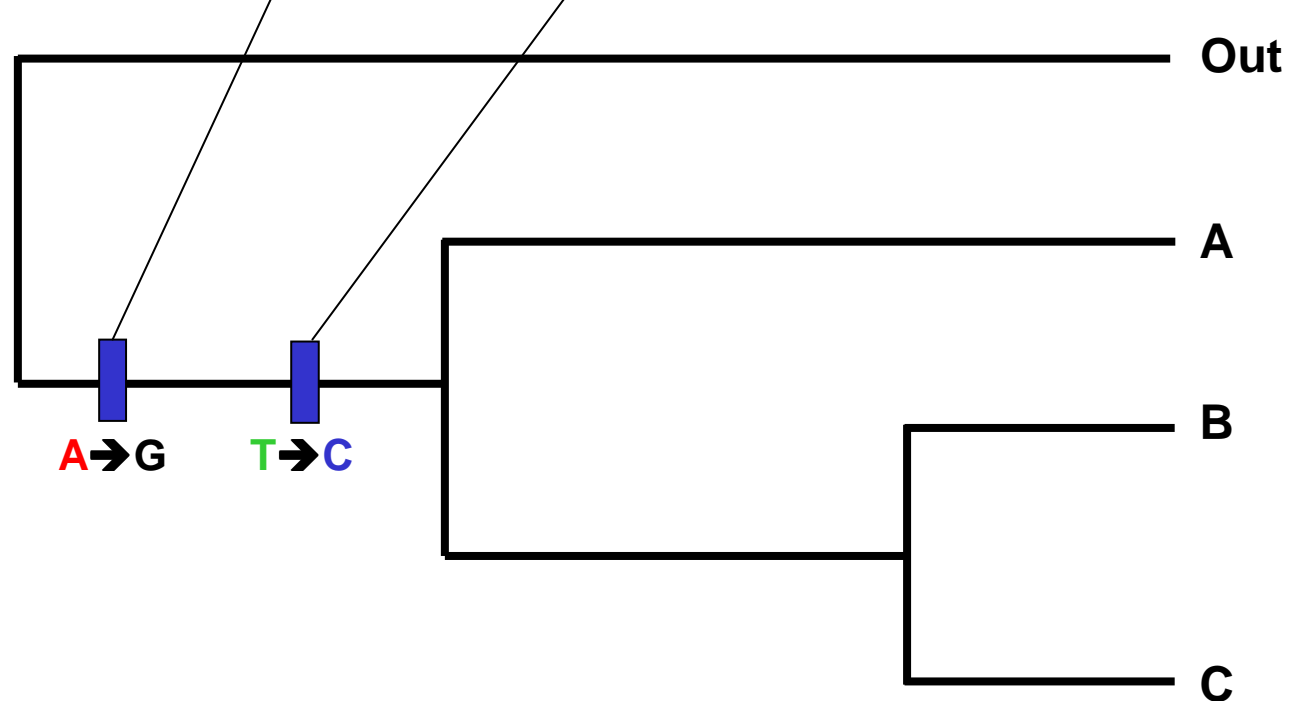
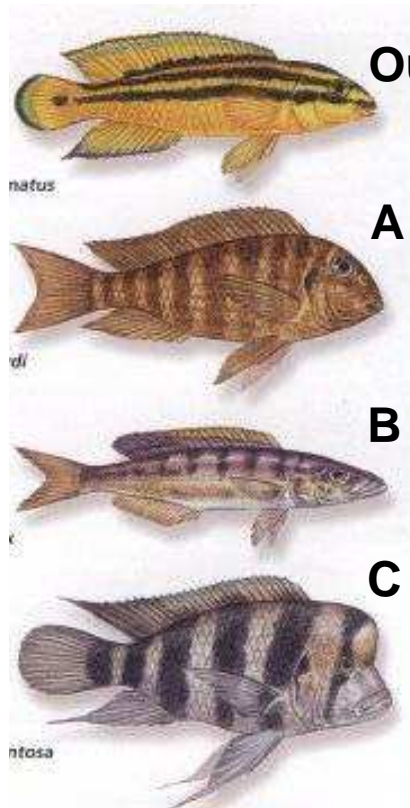
Taxon A

Taxon B

Taxon C

AAGCTTCATA  
GAGCTTCACA  
GTGCTTCACG  
GTGCTTCACG

Synapomorphies  
supporting A+B+C





# molecular characters

Outgroup

Taxon A

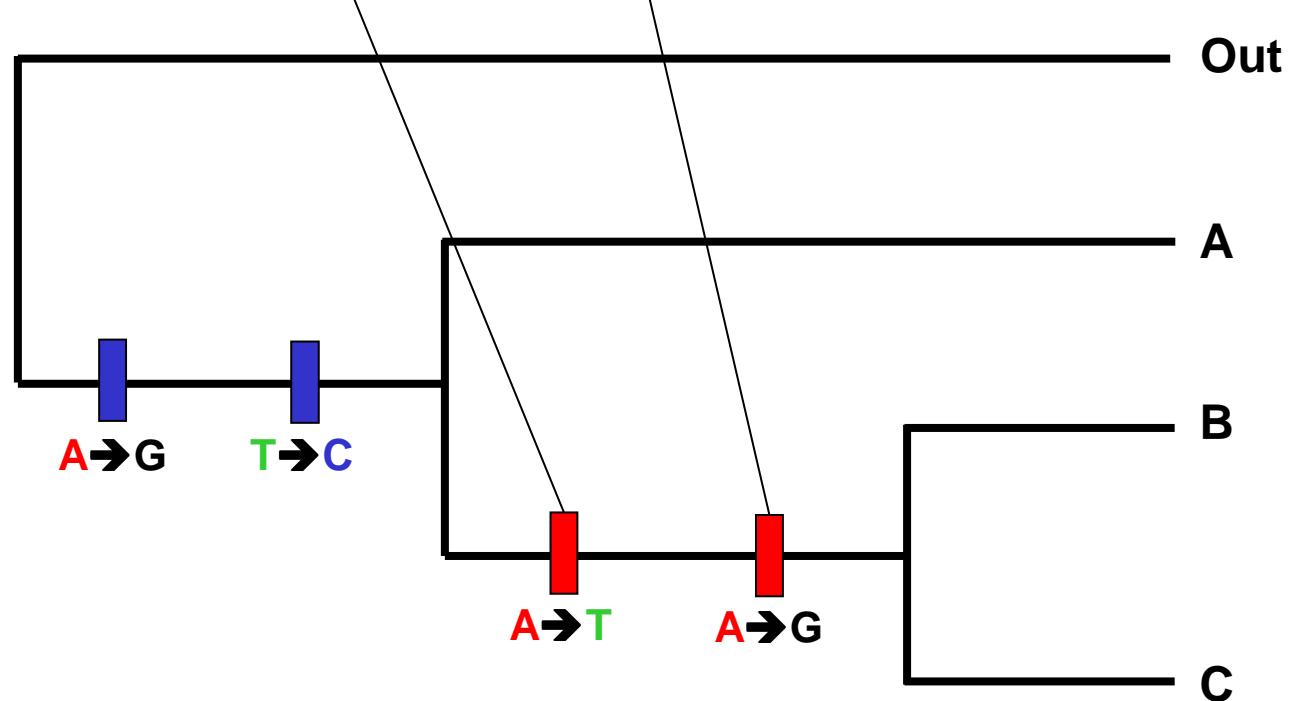
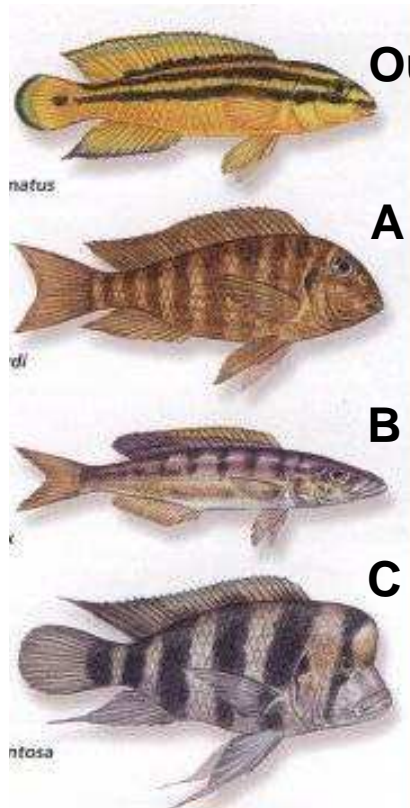
Taxon B

Taxon C

AAGCTTCATA  
GAGCTTCACA  
GTGCTTCACG  
GTGCCTCACG

Synapomorphies  
supporting A+B+C

Synapomorphies  
supporting B+C



# molecular characters

Outgroup

Taxon A

Taxon B

Taxon C

AAGCTTCATA

GAGCTTCACA

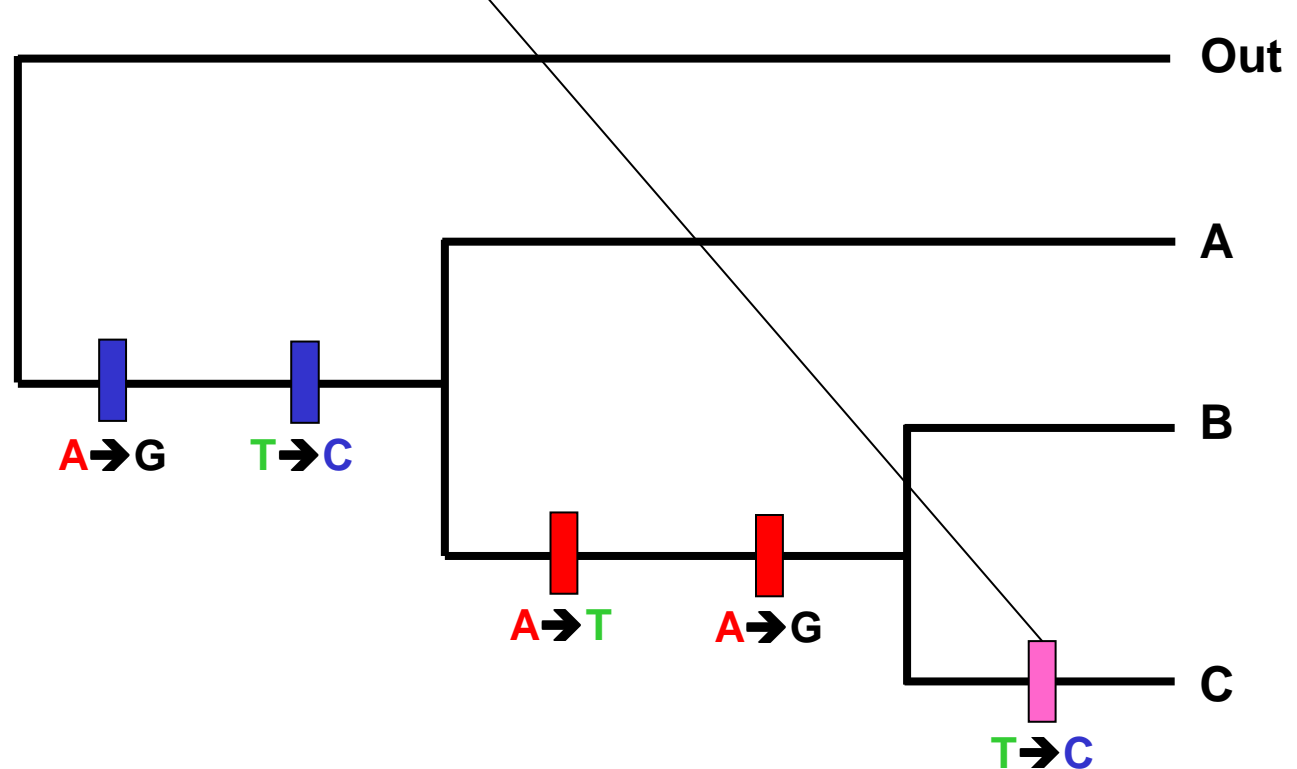
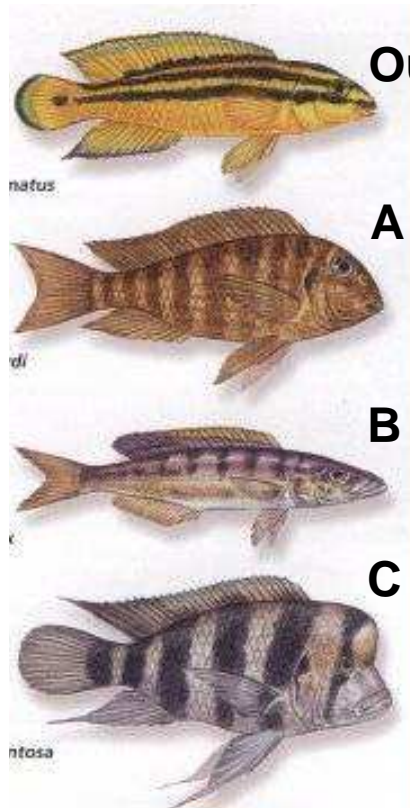
GTGCTTCACG

GTGCTTCACG

Synapomorphies  
supporting A+B+C

Synapomorphies  
supporting B+C

Apomorphy for C



# How to make phylogenetic trees?

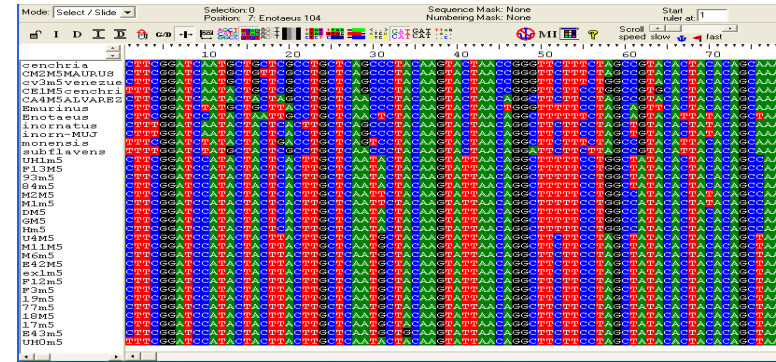
## Workflow:

---

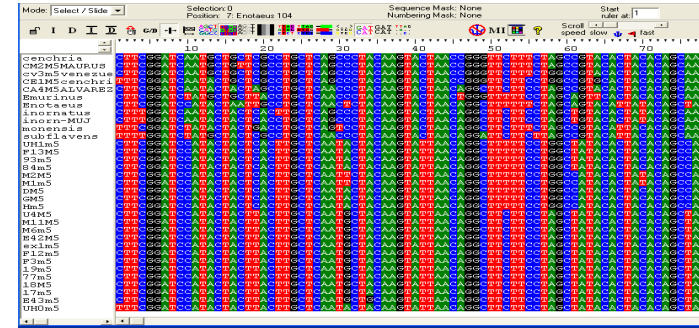
- ✓ obtain DNA sequence
- ✓ quality check
  - sequence alignment
  - calculating genetic distances
  - phylogeny estimation – topology and branch length
  - reliability test (bootstrap)
  - tree visualization

Where sequences differ and  
where are the same?

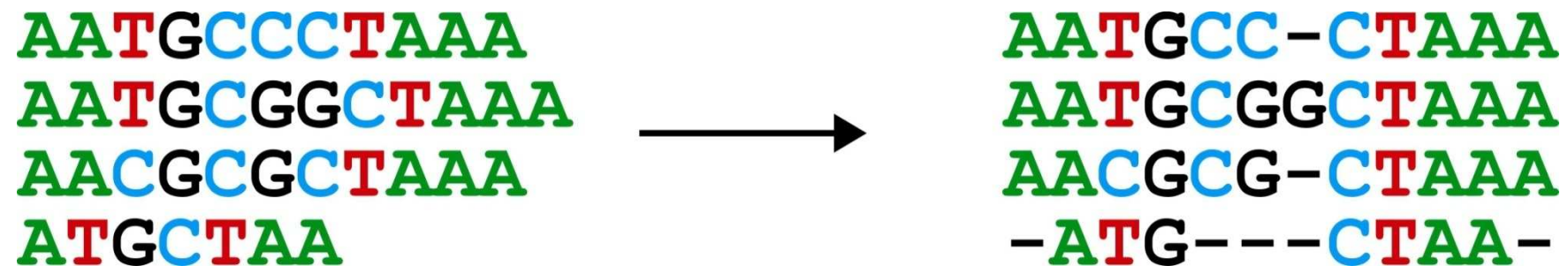
# Alignment



- a way of arranging the sequences of DNA, (RNA, protein) to identify regions of similarity
  - start of every phylogenetic analysis
  - assessing of position homology of each base in the sequence
  - each position (column in the alignment) in the sequence represents character potentially useful for the phylogenetic analysis
  - different programs for calculating and editing alignments
    - manual: BioEdit, Macaw
    - automatic – different algorithms
- Clustal X, PileUp, Multalin, Mafft – often online



Alignment - pairwise alignment (two sequences)  
 - multiple alignment (more sequences)



gaps are inserted between the bases so that identical or similar characters are aligned in successive columns



during alignment: algorithm is searching for the best position for each base and **receives points**:

**Plus points** for every match,

**Minus points** for inserting or extension of the gap,

**The goal is to obtain the best score**

(A)

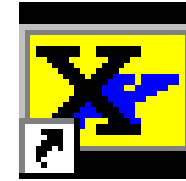
C	9																			
S	-1	4																		
T	-1	1	5																	
P	-3	-1	-1	7																
A	0	1	0	-1	4															
G	-3	0	-2	-2	0	6														
N	-3	1	0	-2	-2	0	6													
D	-3	0	-1	-1	-2	-1	1	6												
E	-4	0	-1	-1	-1	-2	0	2	5											
Q	-3	0	-1	-1	-1	-2	0	0	2	5										
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8									
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5								
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5							
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5						
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4					
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4				
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	3	2	1	3	1	4			
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6		
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7	
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W

**Figure 4.4**

**Amino acid substitution scoring matrices.** (A) The BLOSUM-62 matrix and (B) the PAM120 substitution matrix. Each cell represents the score given to a residue paired with another residue (row  $\times$  column). The values are given in half-bits, as discussed in Section 5.1. The colored shading indicates different physicochemical properties of the residues (see Figure 2.3): small and polar, yellow; small and nonpolar, white; polar or acidic, red; basic, blue; large and hydrophobic, green; aromatic, orange.

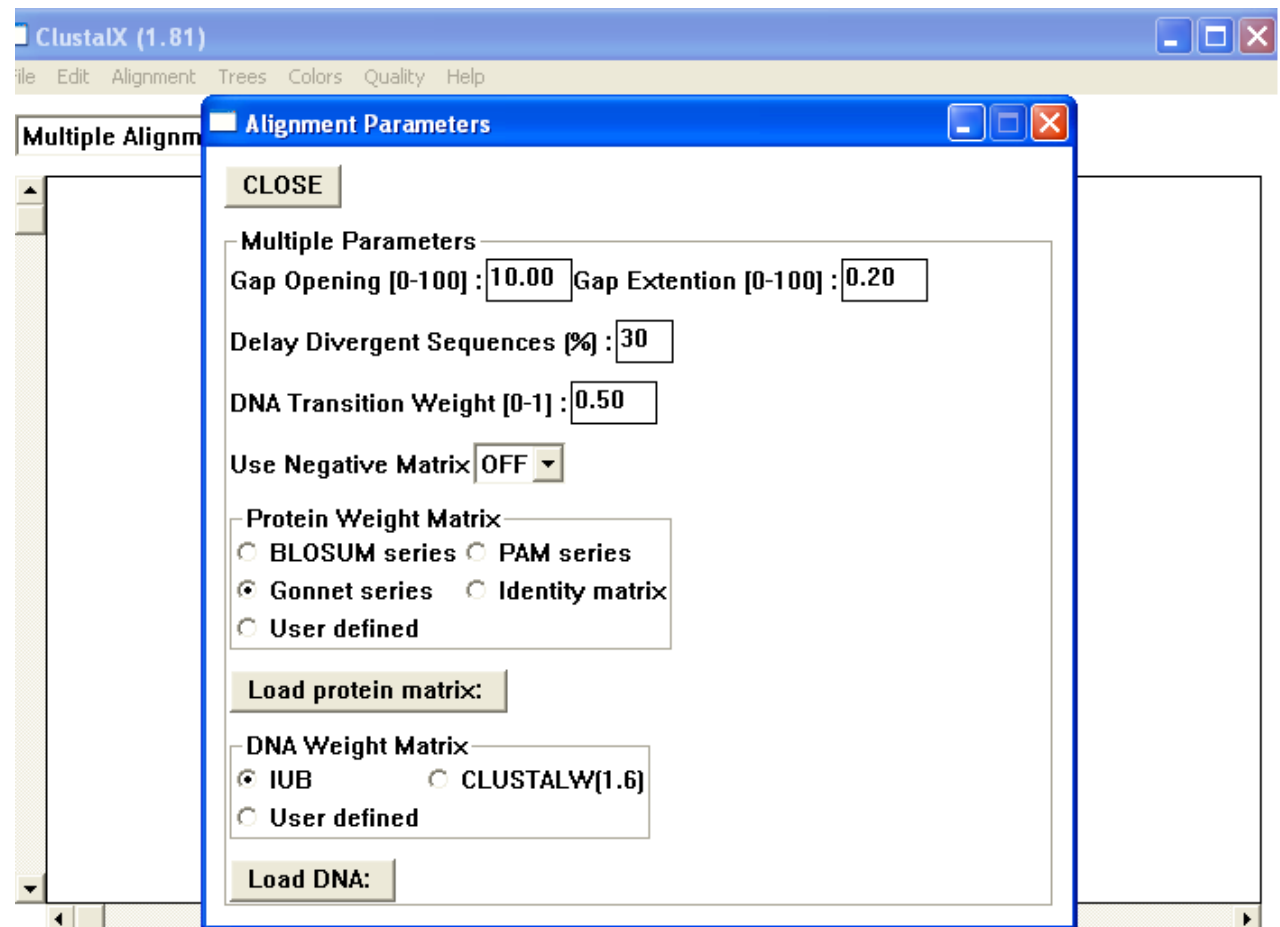
scoring matrices

# Alignment:



Clustal W (Clustal X)

[www.clustal.org](http://www.clustal.org) – frequently used software (Thompson et al. 1994 a 1997)



# What can help:

- first use default parameters, later increase penalty for gap opening and decrease penalty for gap extension
- coding genes align as amino acids
- use information about sequence secondary structure (genes for 12S a 16S RNA) – database with alignments with secondary structure in mind (<http://www.arb-silva.de/>)
- delete ambiguously align positions

MOLECULAR PHYLOGENETICS AND EVOLUTION  
Vol. 4, No. 1, March, pp. 1–9, 1995

**Elision: A Method for Accommodating Multiple Molecular  
Sequence Alignments with Alignment-Ambiguous Sites**

WARD C. WHEELER,\* JOHN GATESY,† AND ROB DeSALLE‡

MOLECULAR PHYLOGENETICS AND EVOLUTION  
Vol. 2, No. 2, June, pp. 152–157, 1993

**Alignment-Ambiguous Nucleotide Sites and the Exclusion of  
Systematic Data**

JOHN GATESY,\* ROB DeSALLE,† AND WARD WHEELER\*





# Phylogeny estimation

character based (maximum parsimony, maximum likelihood, Bayesian analysis)

two types of methods

distance based (Neighbour-joining, UPGMA)

Two different approaches:

**algorithm** – number of specific steps resulting in one best tree  
methods: UPGMA, Neighbour-joining

**optimality criterion** – consider and compare all theoretically possible trees  
based on selected criteria- number of evolutionary steps,  
likelihood value



# distances

input is a matrix of distances between species

		taxon			
taxon	I	---	0.1	0.4	0.6
	II	---	---	0.5	0.5
	III			---	0.6
	IV				---

# proportional (p) distance

number of substitutions between sequences

p = total number of base differences / total no. of available sites

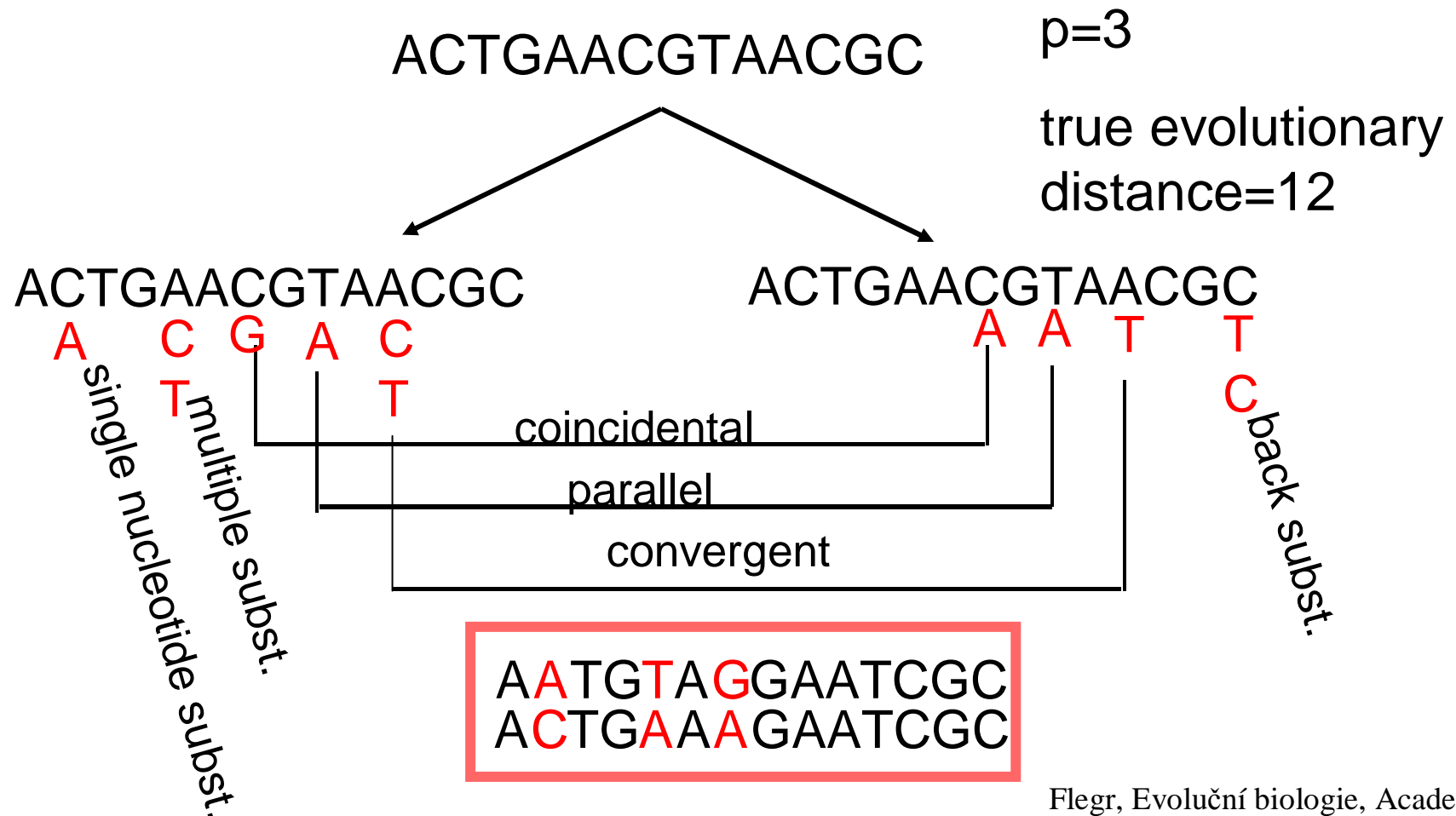
$$p = n_d / n$$

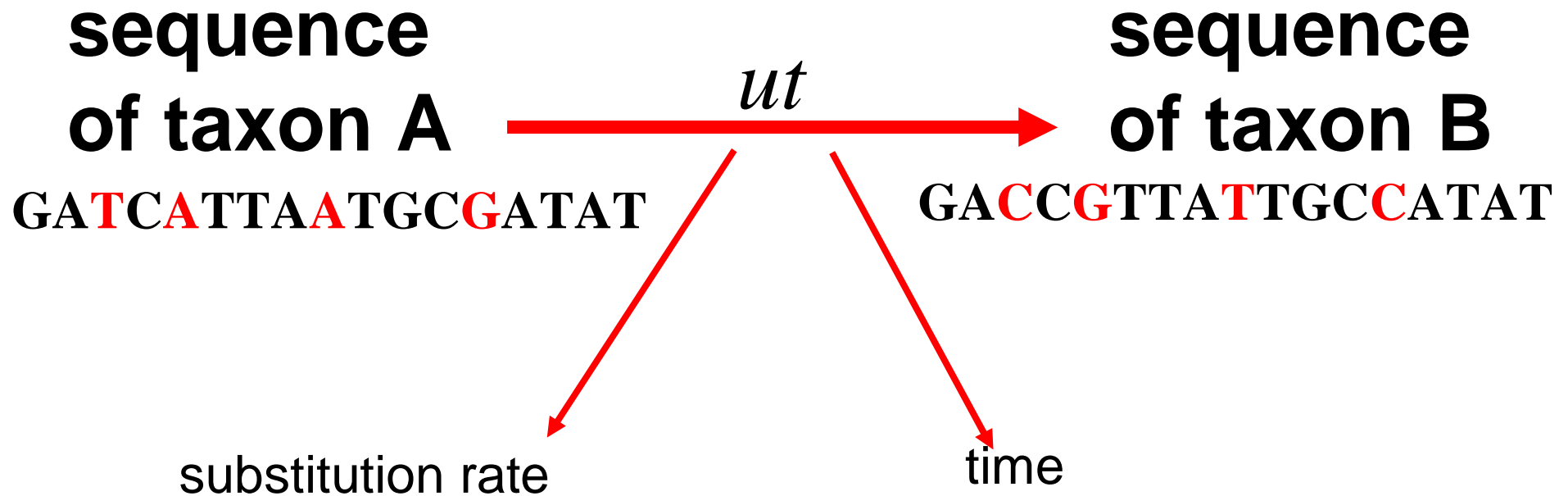
GATCATTATGCGATAT  
GACCGTTATTGCCATAT

$$p = 4/17 = 0.23$$

# real number of substitutions in the sequence over time is usually higher than observed p distance

we can see just 3 differences (p), but in fact there was 12 substitutions





→ in phylogenetic analyses we use “correction” of observed distances to estimate number of hidden changes (multiple mutations etc.)

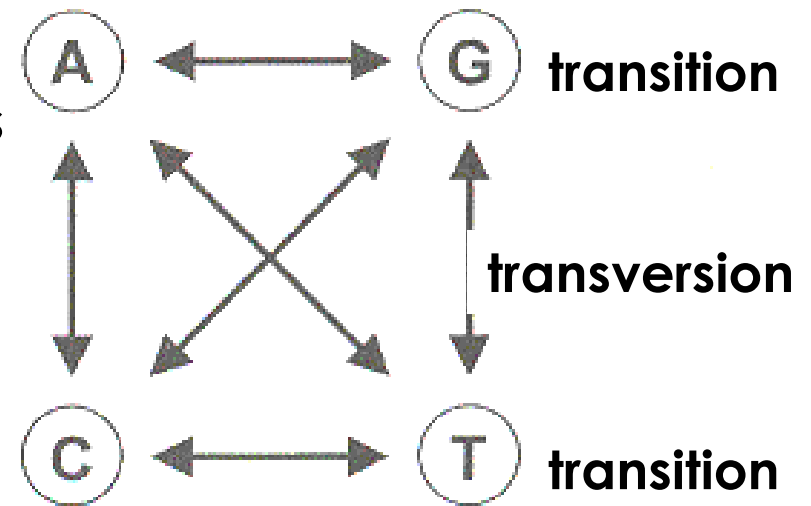
correction based on different substitution type (Ts, Tv),  
different substitution rate, frequencies of nucleotides

## Jukes-Cantor model (distance)

all substitution types and base frequencies are presumed equal

### JC distance

$$d_{JC} = -\frac{3}{4} \ln\left(1 - \frac{4}{3} p\right)$$



## Kimura 2-parameter model (K2P):

transitions are more likely than transversions,  
equal base frequencies

K2P distance

$$P = n_{TS} / n$$

$$Q = n_{TV} / n$$

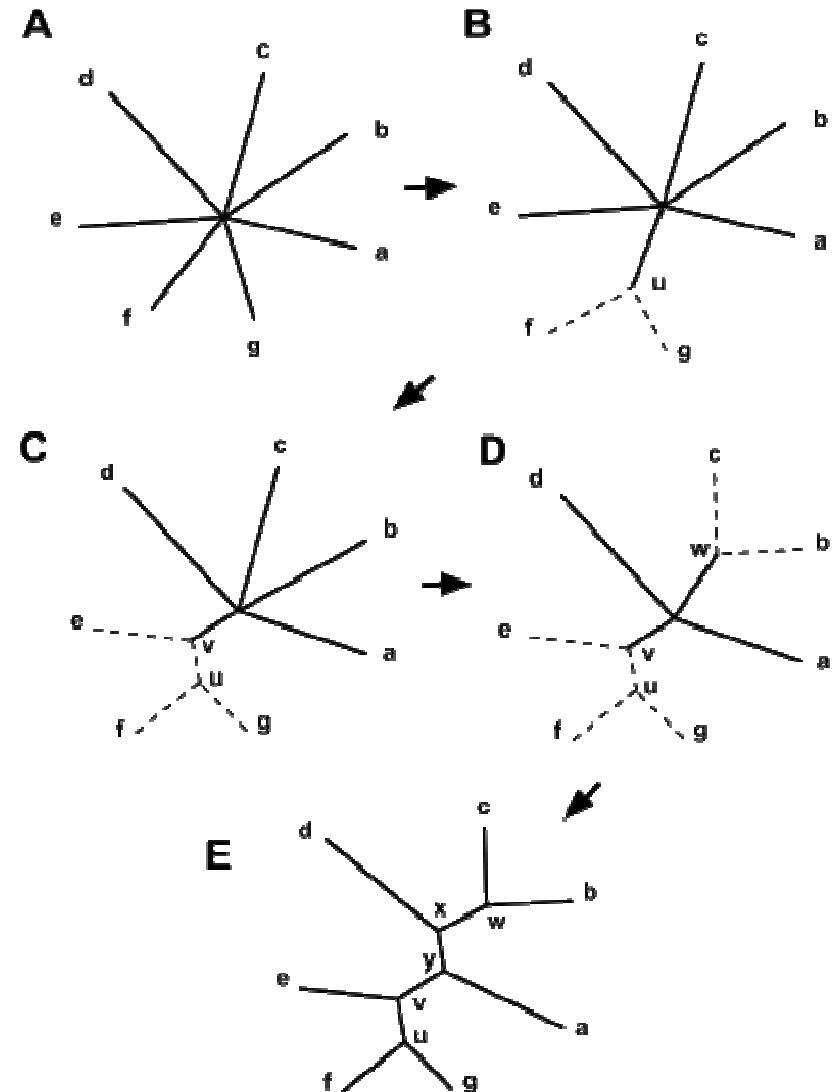
$$d_{K2P} = 0.5 \ln\left(\frac{1}{1 - 2P - Q}\right) + 0.25 \ln\left(\frac{1}{1 - 2Q}\right)$$



# methods

**Neighbour-joining (NJ)** - the fully resolved tree is decomposed from a fully unresolved star tree by successively inserting branches between a pair of closest neighbors and the remaining terminals in the tree - result is one tree

- **other methods: UPGMA**  
(Unweighted Pair Group Method using Arithmetic means), **Minimal evolution**



## conclusion, pros and cons

distance methods rely on evolutionary models to estimate the numbers of multiple substitutions – their result is dependent on how well the accepted models match the actual evolutionary properties of the sequences

only one tree is derived

discards the primary character data

problem with interpretation of branch lengths

very fast, ideal for the first insight

# Maximum parsimony:

optimality criterion - parsimony score = minimum number of events (steps) required by a tree to explain the variation in the data

search for topologies that minimize the total tree length assuming a minimum number of base changes

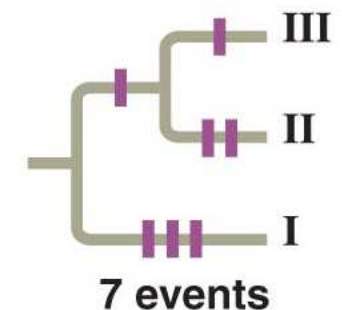
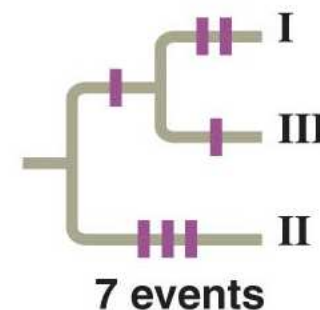
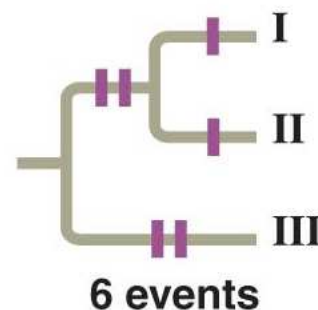
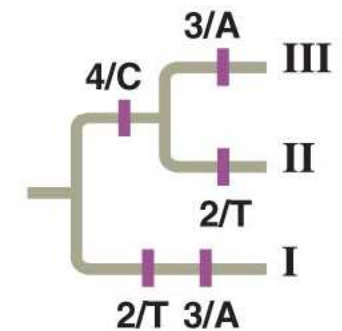
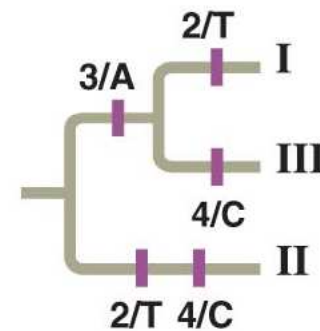
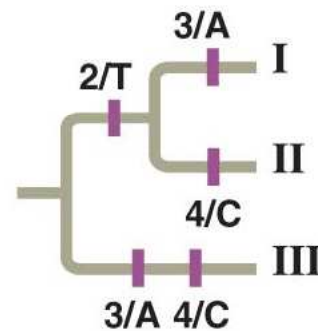
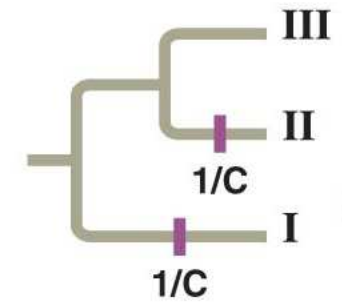
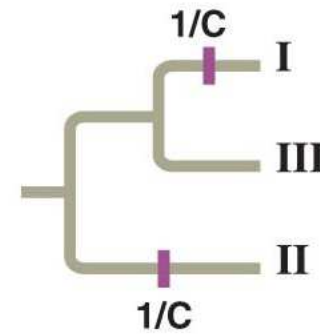
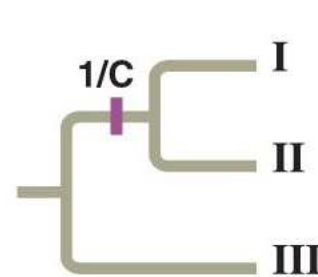
“Occam’s Razor” – “keep it simple“

	Site			
	1	2	3	4
Species I	C	T	A	T
Species II	C	T	T	C
Species III	A	G	A	C
Ancestral sequence	A	G	T	T

## not all characters are good for parsimony:

the alignment is checked for **informative positions**

= a site must have the same sequence characters in at least two taxa and must favor one topology over another



6 events

7 events

7 events

# Maximum parsimony:

optimality criterion - parsimony score = minimum number of events (steps) required by a tree to explain the variation in the data

search for topologies that minimize the total tree length assuming a minimum number of base changes

“Occam’s Razor” – “keep it simple”

**(2n – 3)!** We already know that there are a lot of possible trees- in most cases we can not compare all of them

**$2^{n-2}(n-2)!$**

no. of taxa	no. of unrooted trees	no. of rooted trees
4	3	15
8	10 395	135 135
10	2 027 025	34 459 425
22	$3 \times 10^{23}$	
50	$3 \times 10^{74}$	
100	$2 \times 10^{182}$	

**no. of trees  
exponentially  
increases**

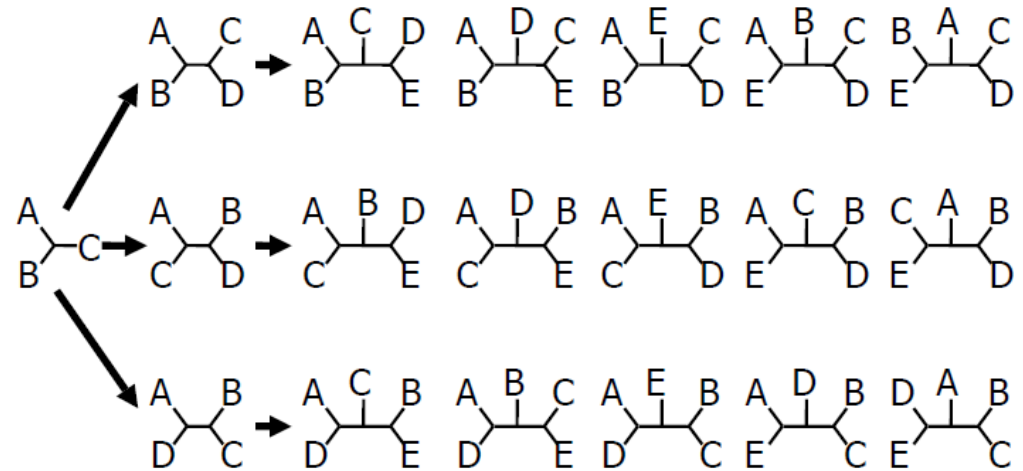




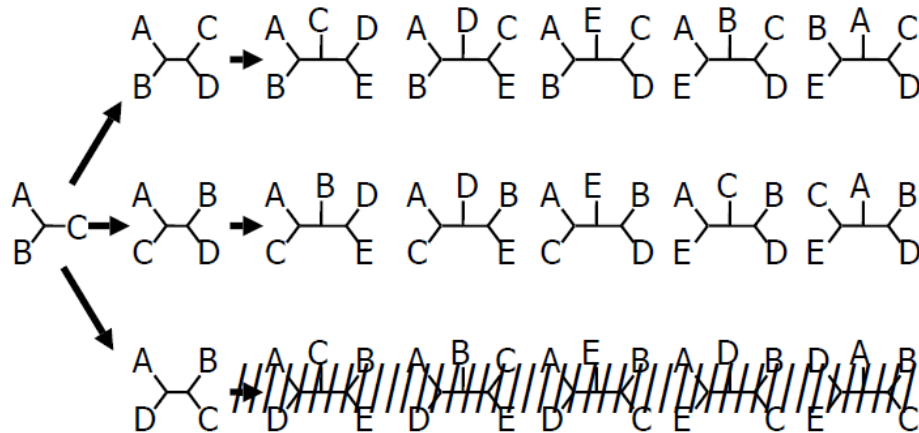
# Tree searching



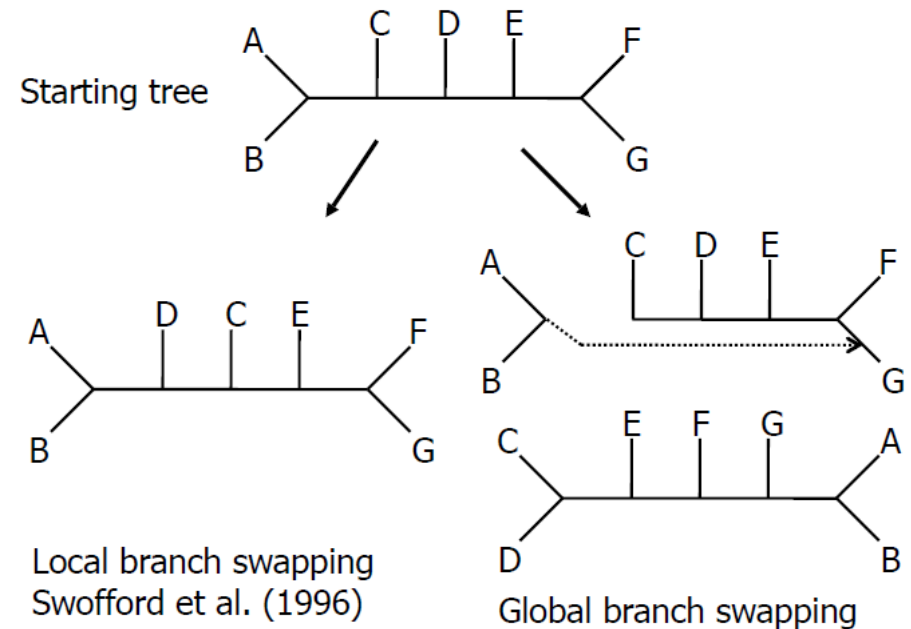
## Exhaustive Searching



## Branch and Bound Searching



## Heuristic Searching



# Maximum parsimony

in most cases we can not compare all trees

⇒ heuristic search

- create random tree
- calculate parsimony score
- rearranging of the tree,
- calculate parsimony score
- further the method works with the better (shorter) tree
- repeated rearranging and calculating scores
- at the end shortest tree

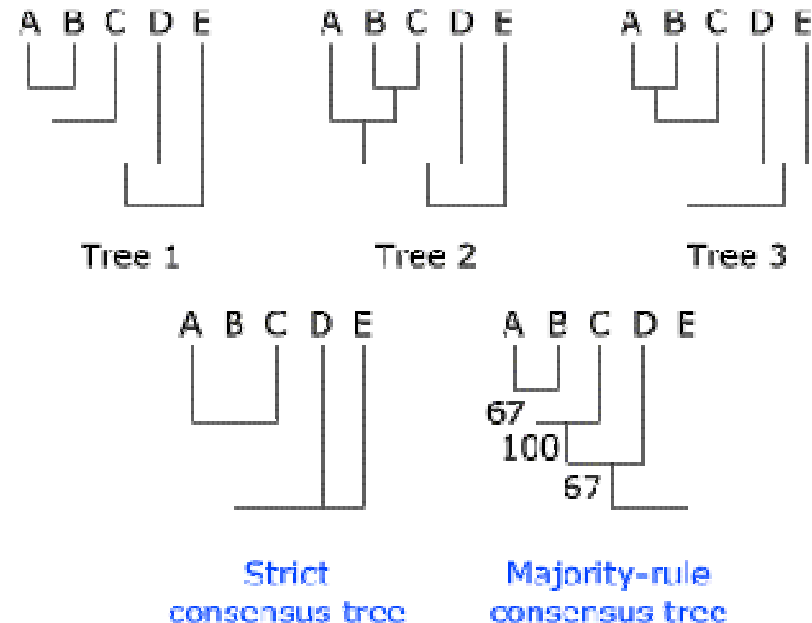
**Sometimes (quite often) we find more equal trees**



# Consensus tree:

**when multiple phylogenies are supported** - a consensus tree shows only those relationships common to all trees (based on our settings)

- **strict consensus** (only relationships common to all trees)
- **majority-rule consensus** (relationships common to 50 or 70% of trees are shown)



# Parsimony: pros and cons

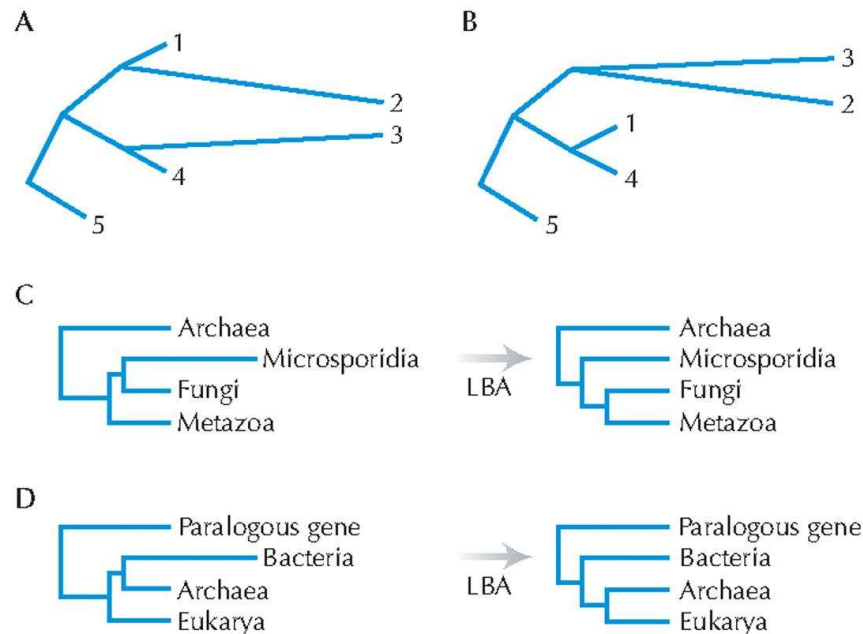


- works directly on the data
- straightforward, well understood principle
- relatively fast
- does not need a model of evolution ( but not really model free – change is rare)



- performs weakly on distantly related data
- prone to “false identities” (multiple changes)
- long branch attraction
- can produce many trees with the same parsimony score

# long branch attraction (LBA)



**FIGURE 5.22.** Long-branch attraction is a methodological artifact that can cause phylogenetic trees to inaccurately portray evolutionary history. The phenomenon causes errors in phylogenetic reconstruction when two (or more) of the entities being studied lie on the end of long branches in their "real" tree but are not sister taxa. (A) In this hypothetical "real" tree of five species, species 2 and 3 (which are not sister taxa, as indicated) have undergone higher rates of evolution than the other three, and thus sit at the end of longer branches. Many phylogenetic reconstruction methods used to infer the evolution of species will cause the long branches to appear to be closely related and thus produce an incorrect tree (as shown in B). (C) In studies of the evolution of microsporidia (a relative of fungi, *left tree*), long-branch attraction (LBA) is believed to have erroneously identified them as deeply branching eukaryotes (*right tree*). (The evolution of microsporidia is discussed in more detail on p. 198.) (D) In trees of anciently duplicated genes, long-branch attraction might have pulled bacteria down to the paralogs used to root the tree, because the paralogs are at the end of a long branch (*right tree*). This would occur if bacteria evolved at a higher rate than archaea and eukaryotes (as suggested in the *left tree*).

5.22C,D, redrawn from Forterre P. et al., *Bioessays* 21: 871–879, © 1999 Wiley-Liss, Inc.

# How to make phylogenetic trees?

## Workflow:

---

- ✓ obtain DNA sequence
- ✓ quality check
- ✓ sequence alignment
- ✓ calculating genetic distances
- ✓ phylogeny estimation – topology and branch length
  - reliability test (bootstrap)
  - tree visualization



# reliability tests

-nonparametric resampling methods - bootstrapping, jackknifing

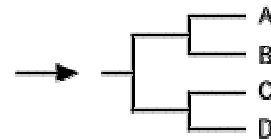
→ new data sets are created from the original data set by sampling columns of characters by random with replacement

- each site can be sampled again with the same probability as any of the other sites

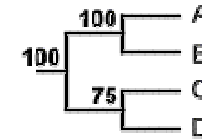
### Box 3

Bootstrap Analysis (Felsenstein, 1985)

s100 ..1010220112..  
 ...  
 ...  
 s3 ..0120401200..  
 s2 ..1000222003..  
 s1 ..1310110012..  
 A ..AGGCUCCAAA..  
 B ..AGGUUCGAAA..  
 C ..AGCCCGGAAA..  
 D ..AUUUCGGAAC..

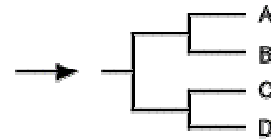


Tree based  
on original  
sequence  
alignment



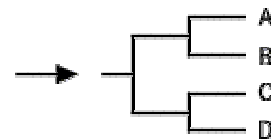
Bootstrap values  
superimposed on  
original tree  
(2)

sample 1 (s1)  
 A ..AGGGGUCAAA..  
 B ..AGGGGUCAAA..  
 C ..AGGGCCCAAA..  
 D ..AUUUUCCACC..



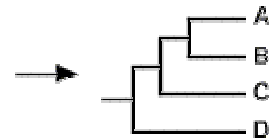
Bootstrap  
tree 1

sample 2 (s2)  
 A ..AUUCCCGGAAA..  
 B ..AUUCCCGGAAA..  
 C ..ACCCCGGAAA..  
 D ..ACCCCGGCC..



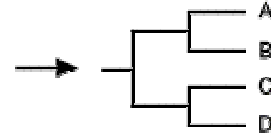
Bootstrap  
tree 2

sample 3 (s3)  
 A ..GGGUUUUCAA..  
 B ..GGGUUUUGAA..  
 C ..GCCCCCGGAA..  
 D ..UUUCCCCGAA..



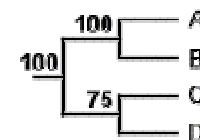
Bootstrap  
tree 3

sample 100 (s100)  
 A ..AGUCCCAAAA..  
 B ..AGUCCCAAAA..  
 C ..AUCCCCAAAA..  
 D ..AUCCCCAACC..



Bootstrap  
tree 100

sample n (100 < n < 2000)



Bootstrap  
consensus tree  
(1)

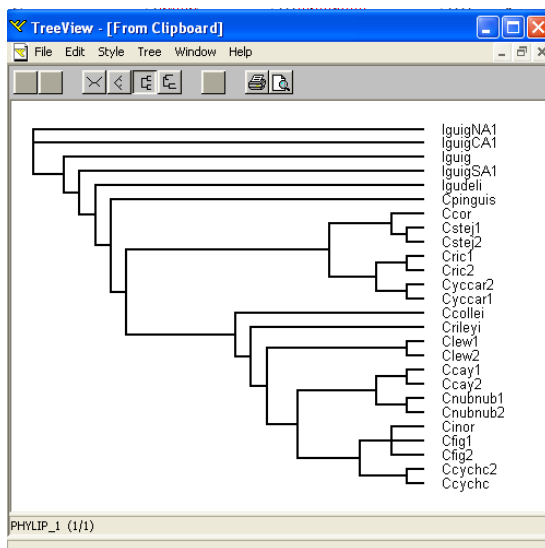
Bootstrap values:  
 < 50% - no - just by chance  
 > 75% MP ok  
 95-100% great

# Tree visualization:

- Newick format

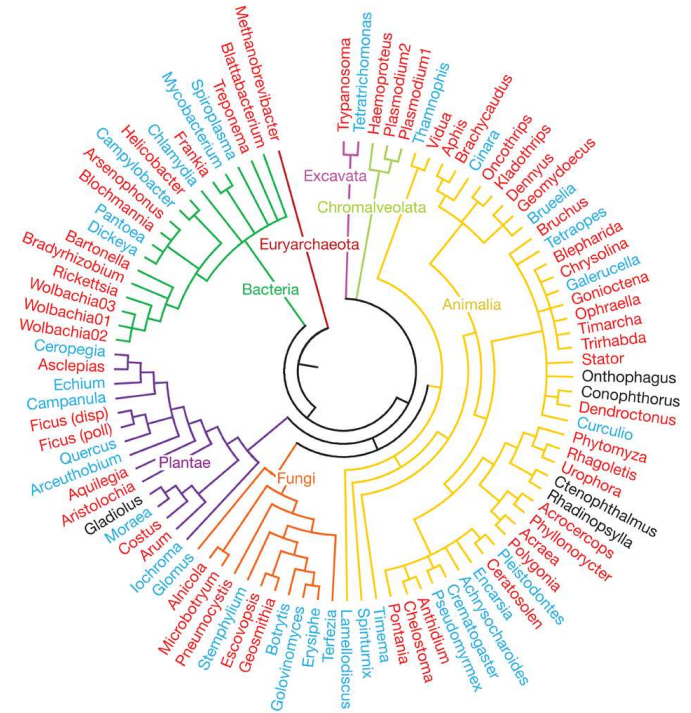
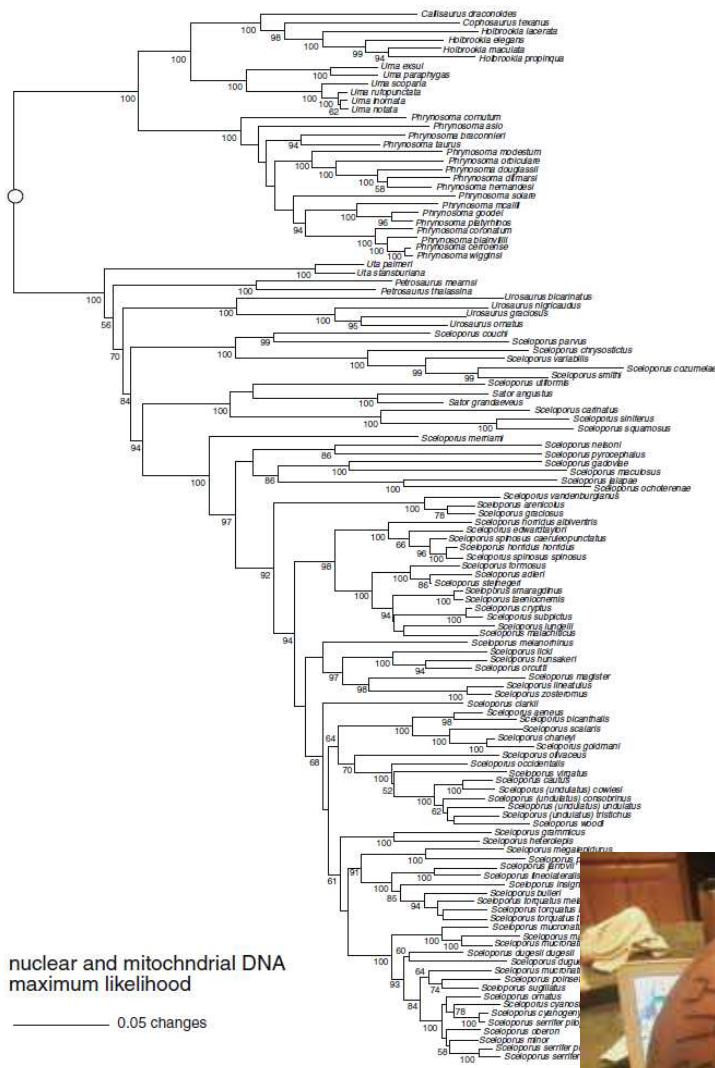
(A,(B,(C,(D,E))))

```
(IguigNA1_:0.00221,(Iguig:0.01733,(((Cpinguis:0.05228,(((Cstej1:0.00012,
Cstej2:0.00098):0.00354,Ccor:0.00543):0.03863,((Cric1:0.00184,
Cric2:0.00184):0.01853,(Cyccar2:0.00636,Cyccar1:0.00498):0.01702):0.03298):0.01722,
(Ccollei:0.04462,(Crileyi:0.01300,((Clew1:0.00221,Clew2:0.00221):0.00885,
(((Ccay1:0.00442,Ccay2:0.00111):0.00111,(Cnubnub1:0.00111,Cnubnub2:0.00000):0.00332):0.
00885,((Cinor:0.00000,Cfig1:0.00111,Cfig2:0.00332):0.00221,(Ccychc2:0.00000,
Ccychc:0.00000):0.00221):0.01051):0.00442):0.01023):0.02222):0.01447):0.02213):0.06215,Igu
deli:0.05379):0.02871,IguigSA1_:0.02231):0.01069):0.02471,IguigCA1_:0.00553);
```



**Different programs for tree  
visualization: TreeView,  
FigTree, Dendroscope**

## Tree visualization:



# Take -Home Message!

- there are more methods how to calculate a tree
- a phylogenetic tree is a hypothesis
- we have to test the reliability
- obtaining a good alignment is one of the most crucial steps towards a good phylogenetic tree

More methods and some examples next week ....



Some useful links for more information about molecular phylogenetics:

[http://evolution.berkeley.edu/evolibrary/article/phylogenetics\\_01](http://evolution.berkeley.edu/evolibrary/article/phylogenetics_01)

<http://www.ncbi.nlm.nih.gov/Class/NAWBIS/Modules/Phylogenetics/phylo1.html>