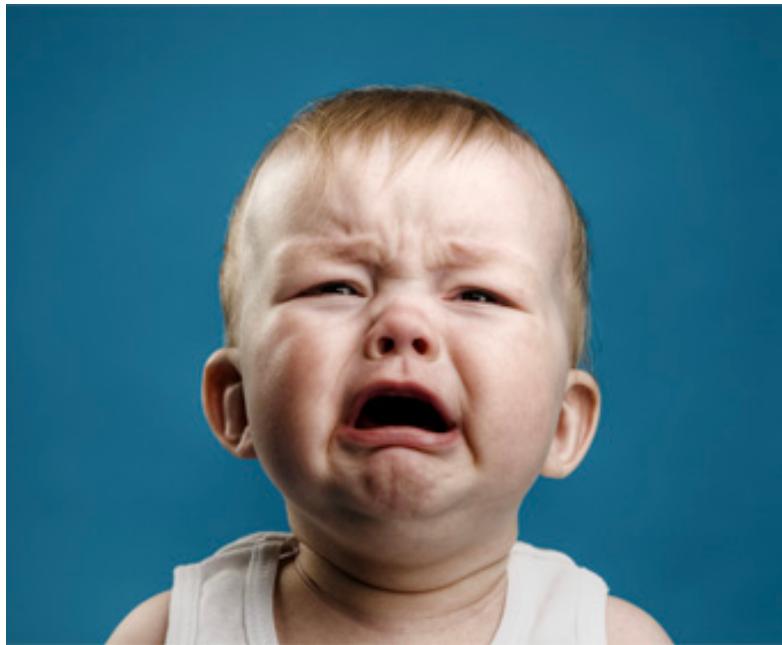




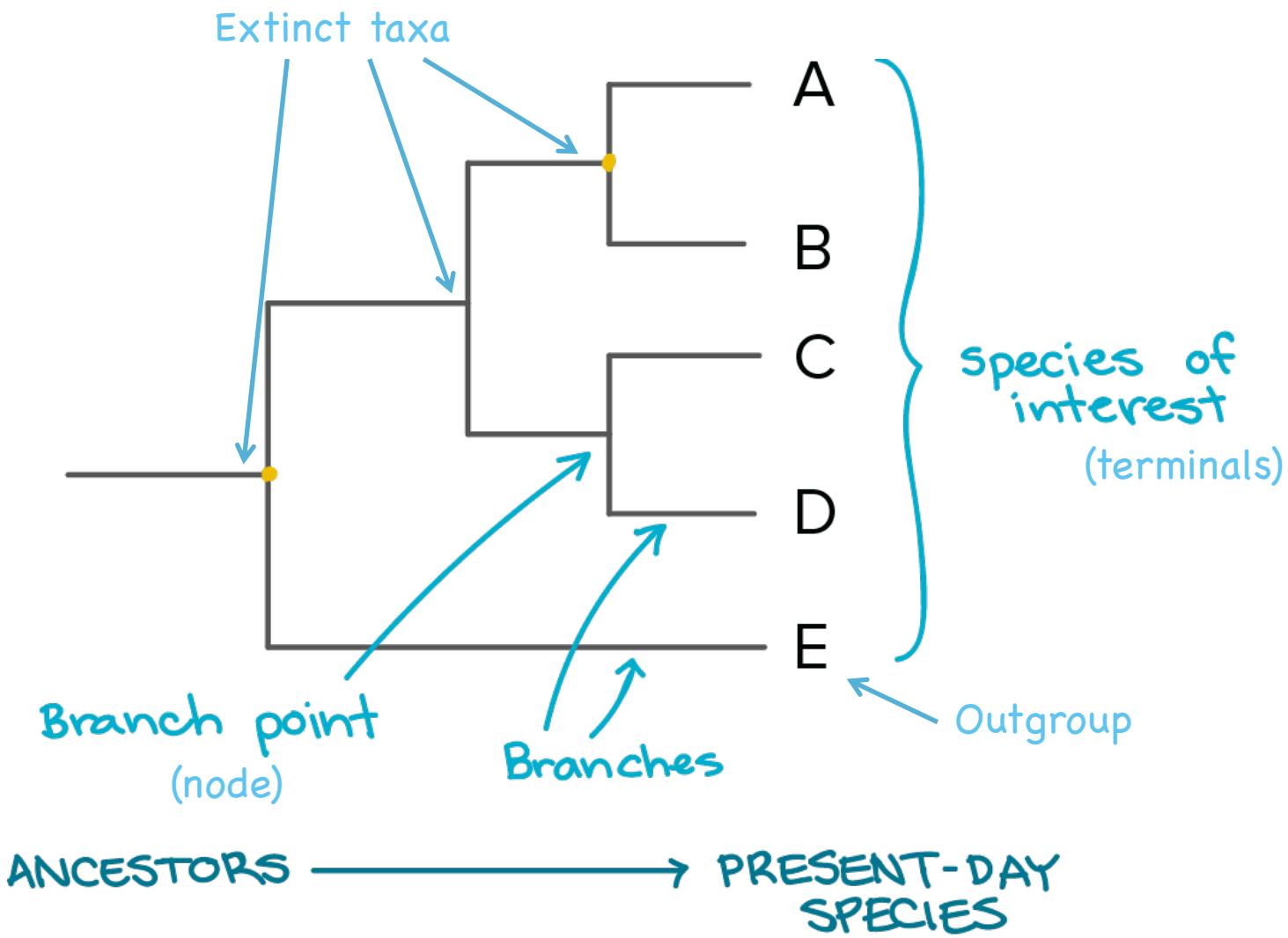
Methods in Phylogenetic Analysis

SUMMARY



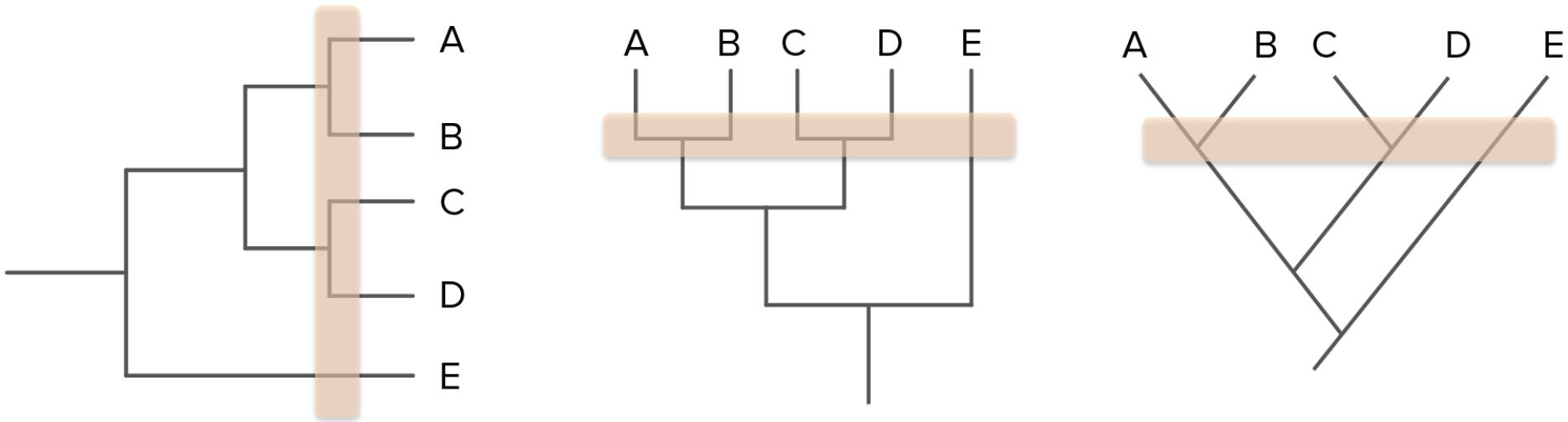
KEEP CALM
YOU
GOT
THIS

The phylogenetic tree



Tree types

Are these the same tree?

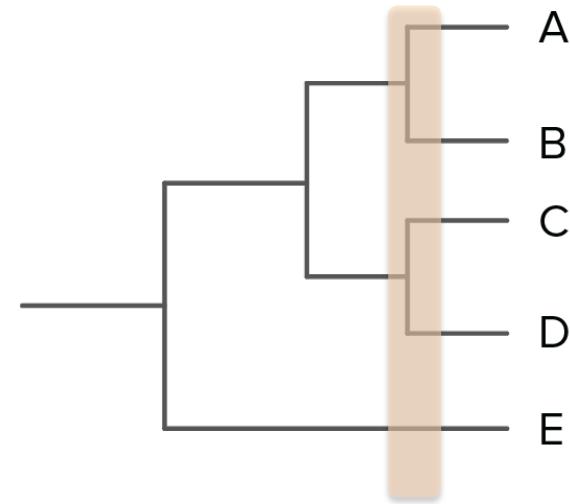


CLADOGRAM: used in cladistics **BUT** a cladogram is not an evolutionary hypothesis since it does not contain information about how ancestors and descendants are related or how much descendants have changed through time

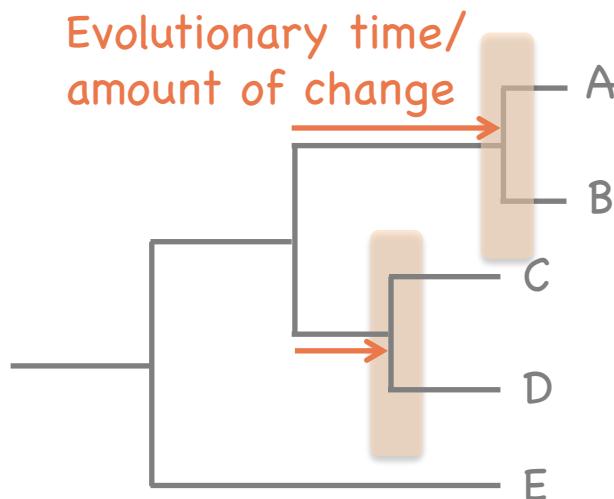
SIMPLISTIC REPRESENTATION
(OK for morphological matrices)

Tree types

Are these the same tree?



CLADOGRAM

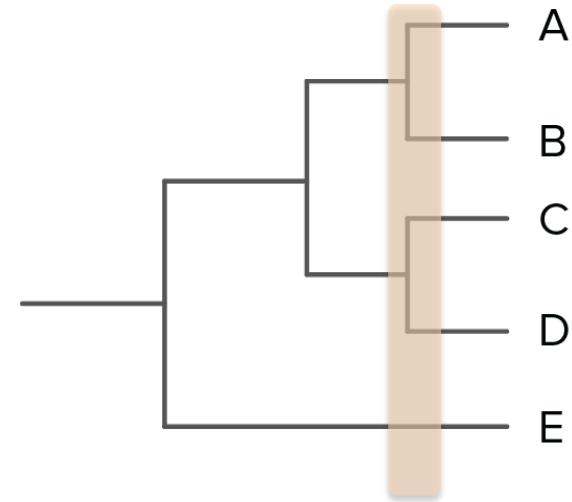


PHYLOGRAM: a branching tree where the branch lengths indicate the amount of **evolutionary change** inferred from the analysis.

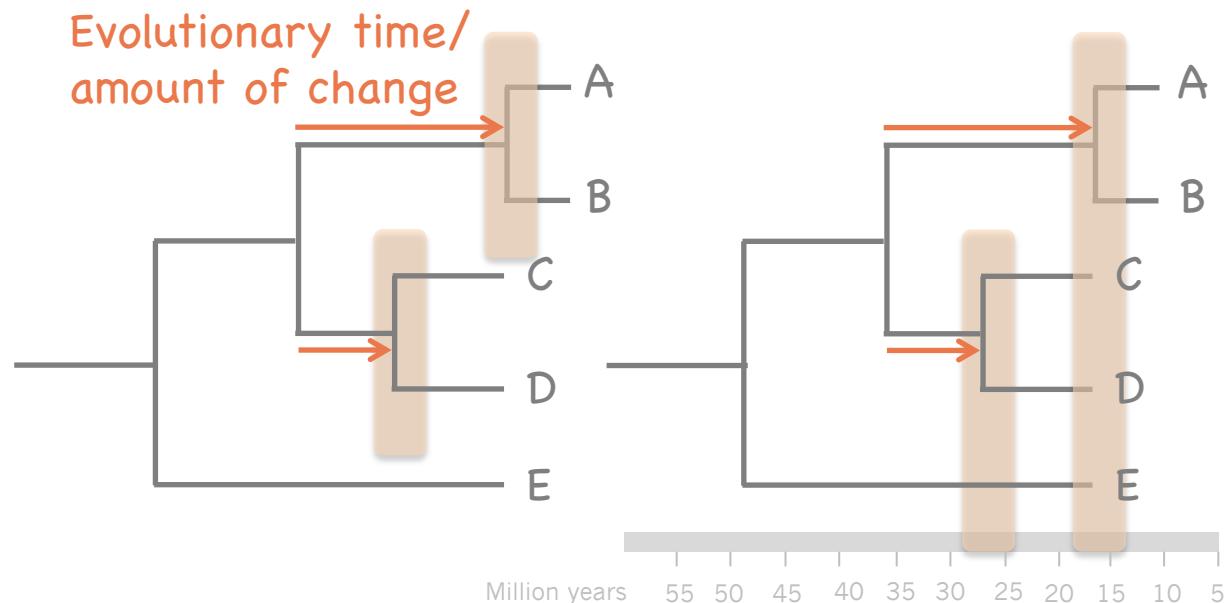


Tree types

Are these the same tree?



CLADOGRAM

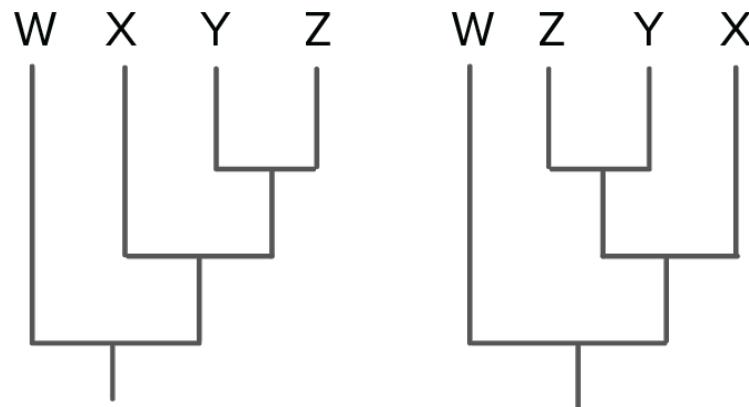


PHYLOGRAM

PHYLOGRAM (ultrametric):
ages assigned to each node
using molecular clocks

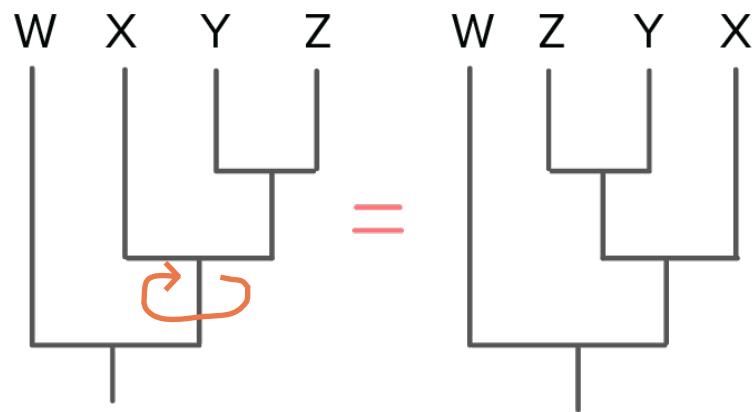
Tree types

Are these the same tree?



Tree types

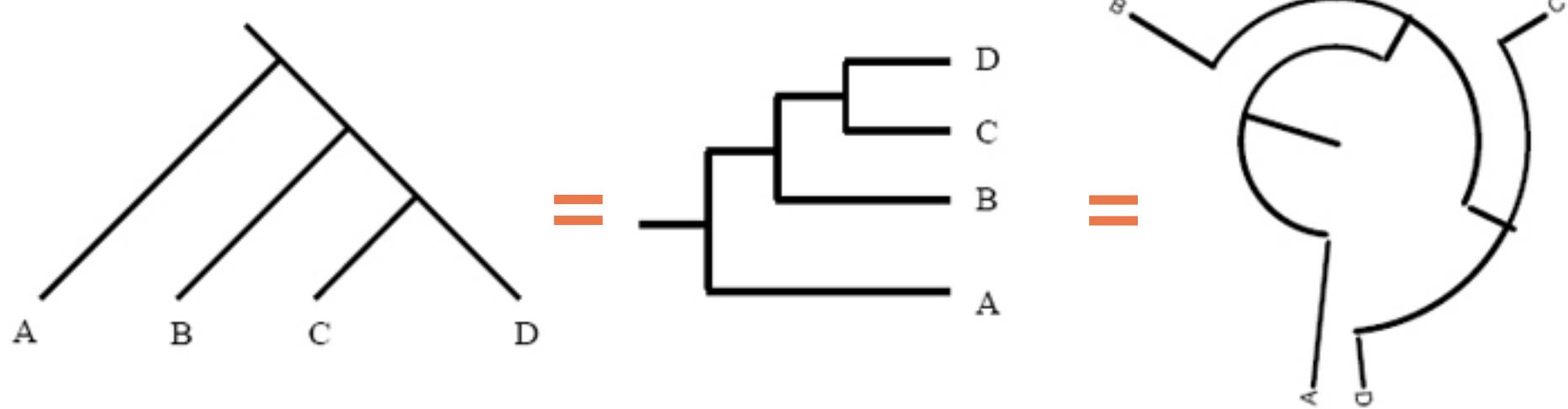
Are these the same tree?



Rotated branches

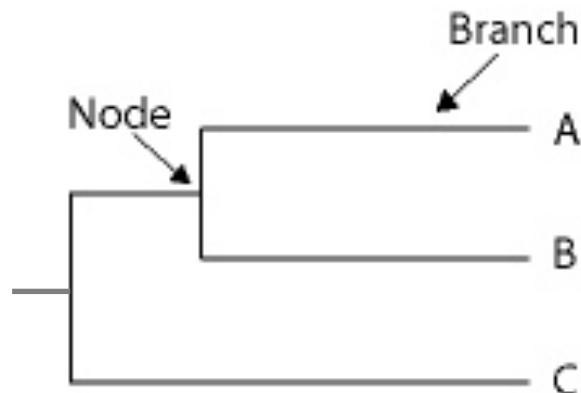
Tree types

Are these the same tree?

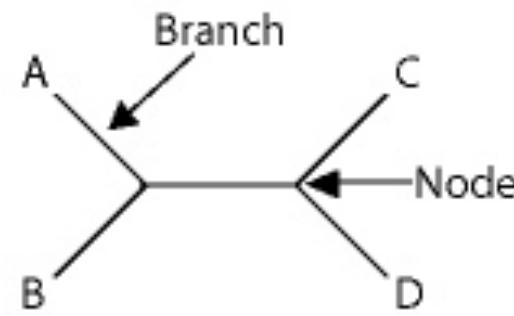


Tree types

Are these the same tree?

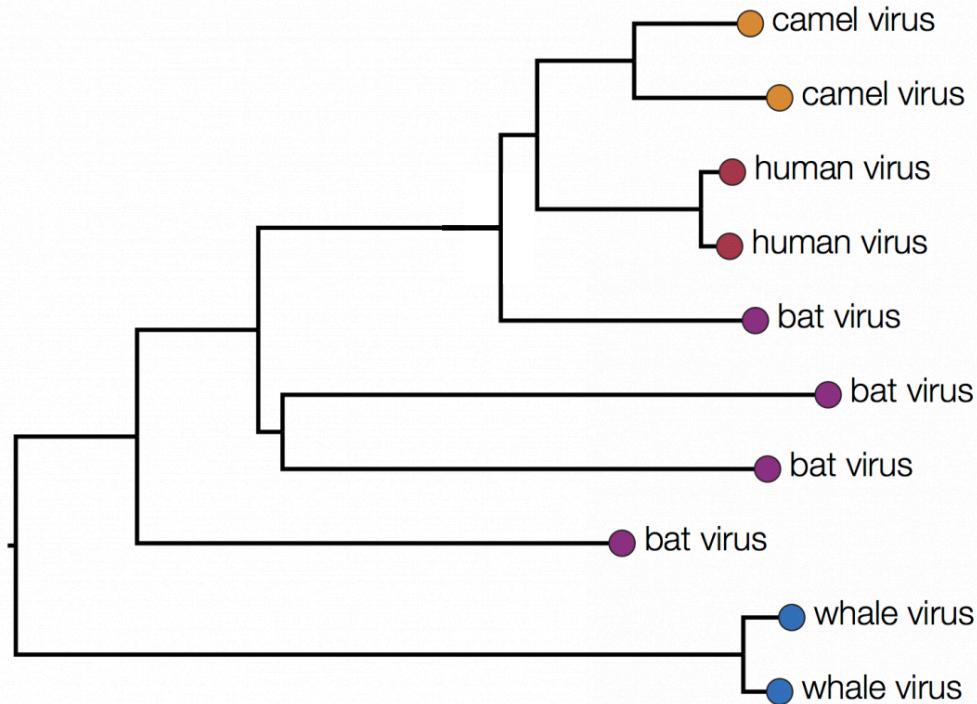


Rooted



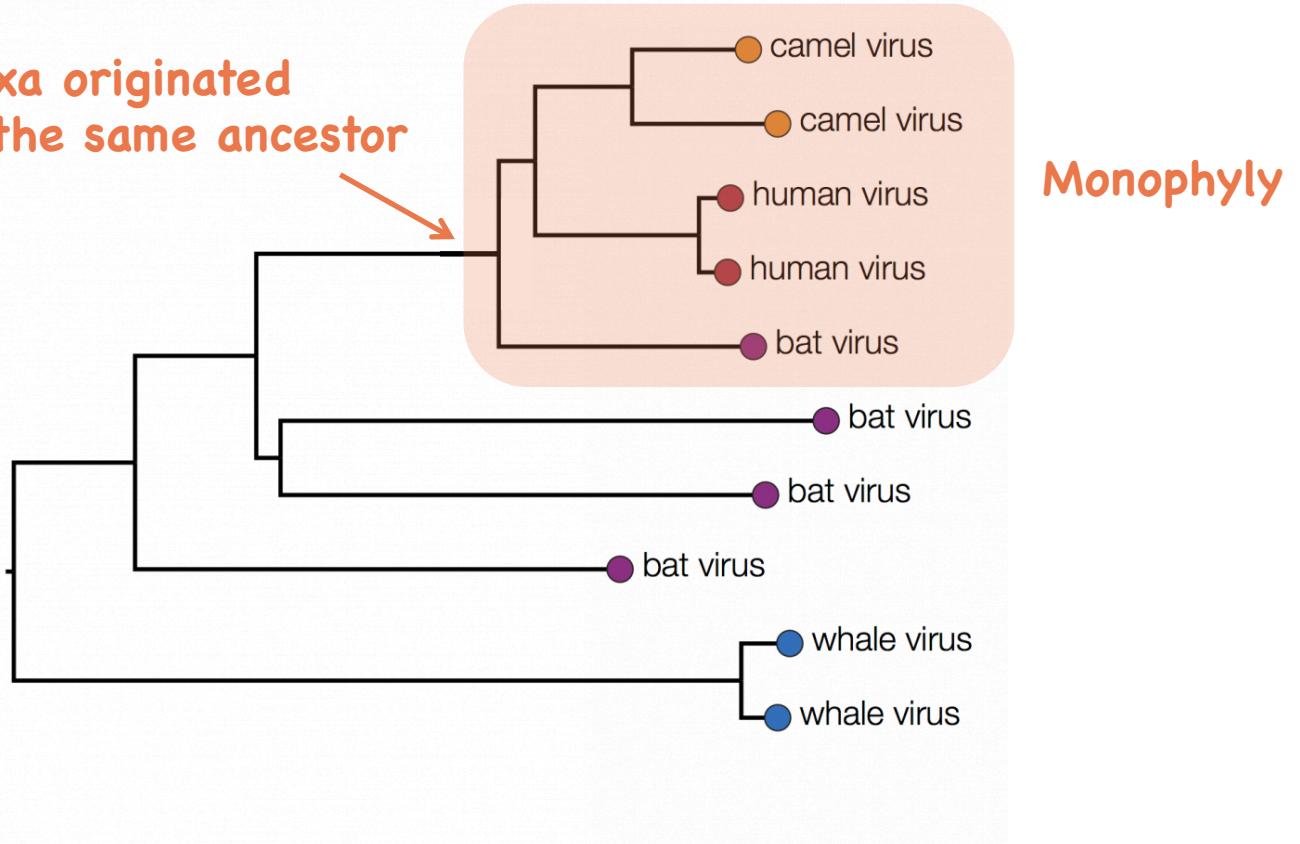
Unrooted

Anatomy of the phylogenetic tree



Anatomy of the phylogenetic tree

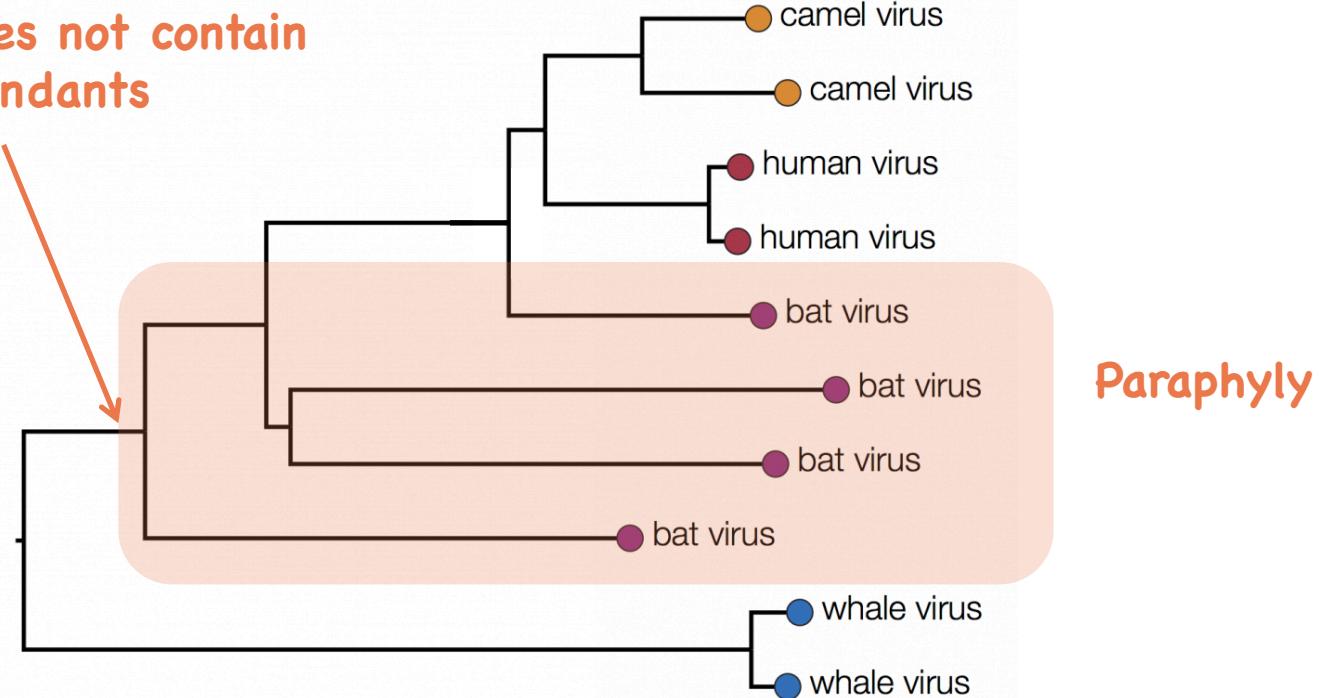
All taxa originated from the same ancestor



Monophyly

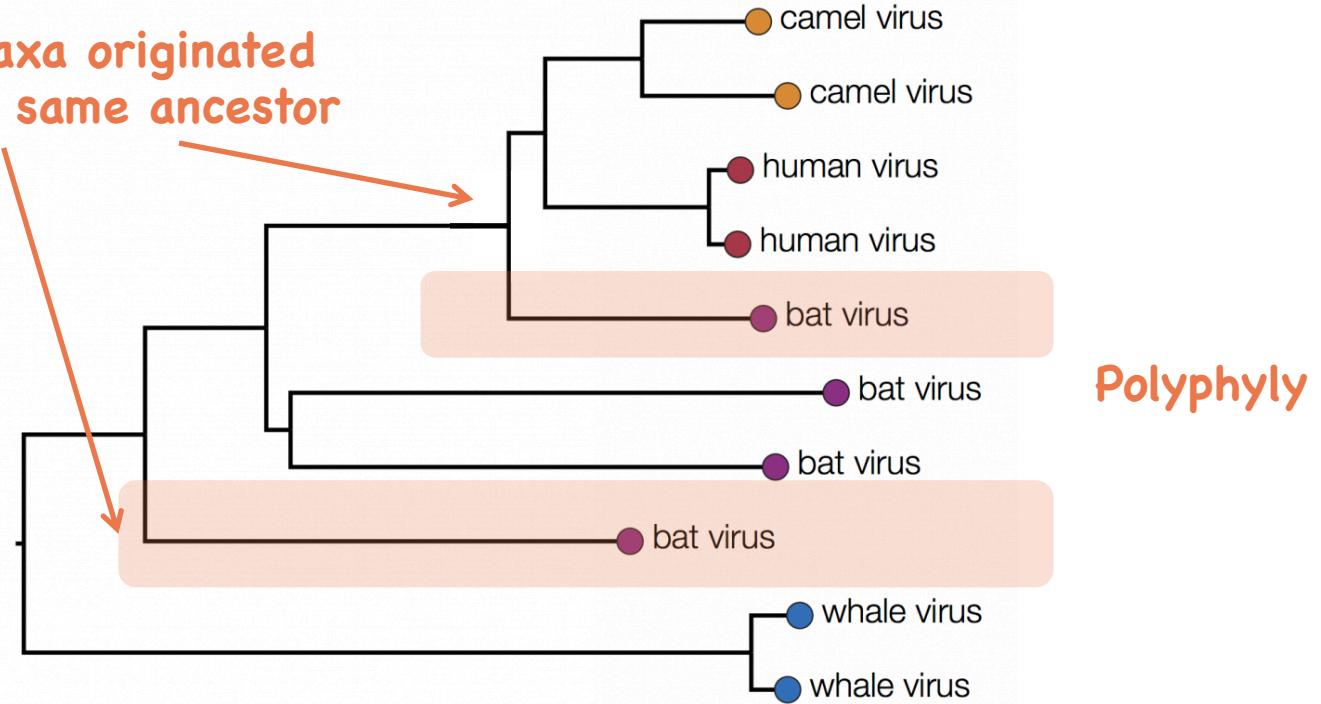
Anatomy of the phylogenetic tree

Single ancestor but the clade does not contain all descendants



Anatomy of the phylogenetic tree

Not all taxa originated from the same ancestor



Polyphyly

What do we need to build a phylogenetic tree

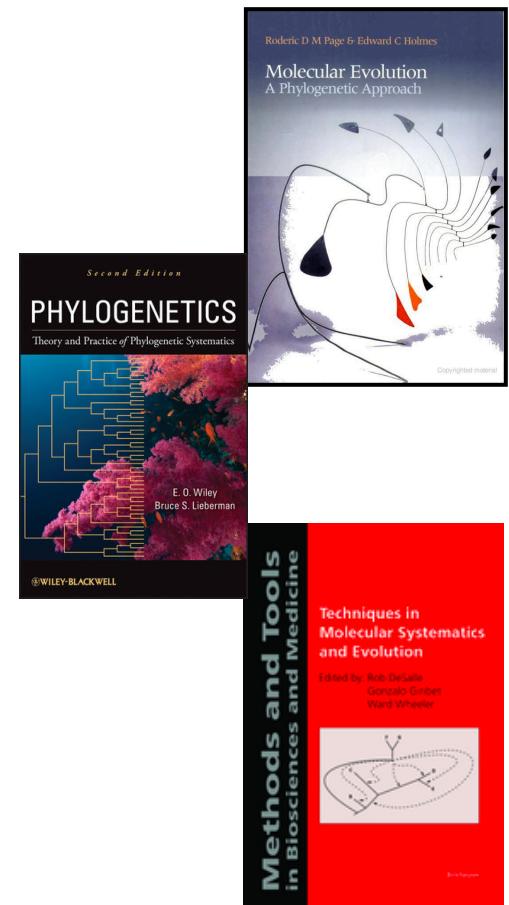
1. **What taxa?** → Your decision, but don't be biased!
2. **What type of data?** → Your decision, but based on what is available!
3. **How many markers?** → Budget! ...also, PCR success...as many as possible
4. **What alignment program?** → Muscle? MAFFT?
5. **What model of evolution?** → jModelTest
6. **What phylogenetic method?** → Your decision...mine is always ML + BI
7. **How to assess confidence?** → Bootstrapping, post. prob., convergence, etc

What do we need to build a phylogenetic tree



What do we need to build a phylogenetic tree

- Page & Holmes, 1998, *Molecular evolution: a phylogenetic approach* (Wiley)
- Wiley & Lieberman, 2011, *Phylogenetics: theory and practice of phylogenetic systematics* (Wiley)
- DeSalle, Giribet & Wheeler, 2002, *Techniques in molecular systematics and evolution* (Birkhäuser Verlag)

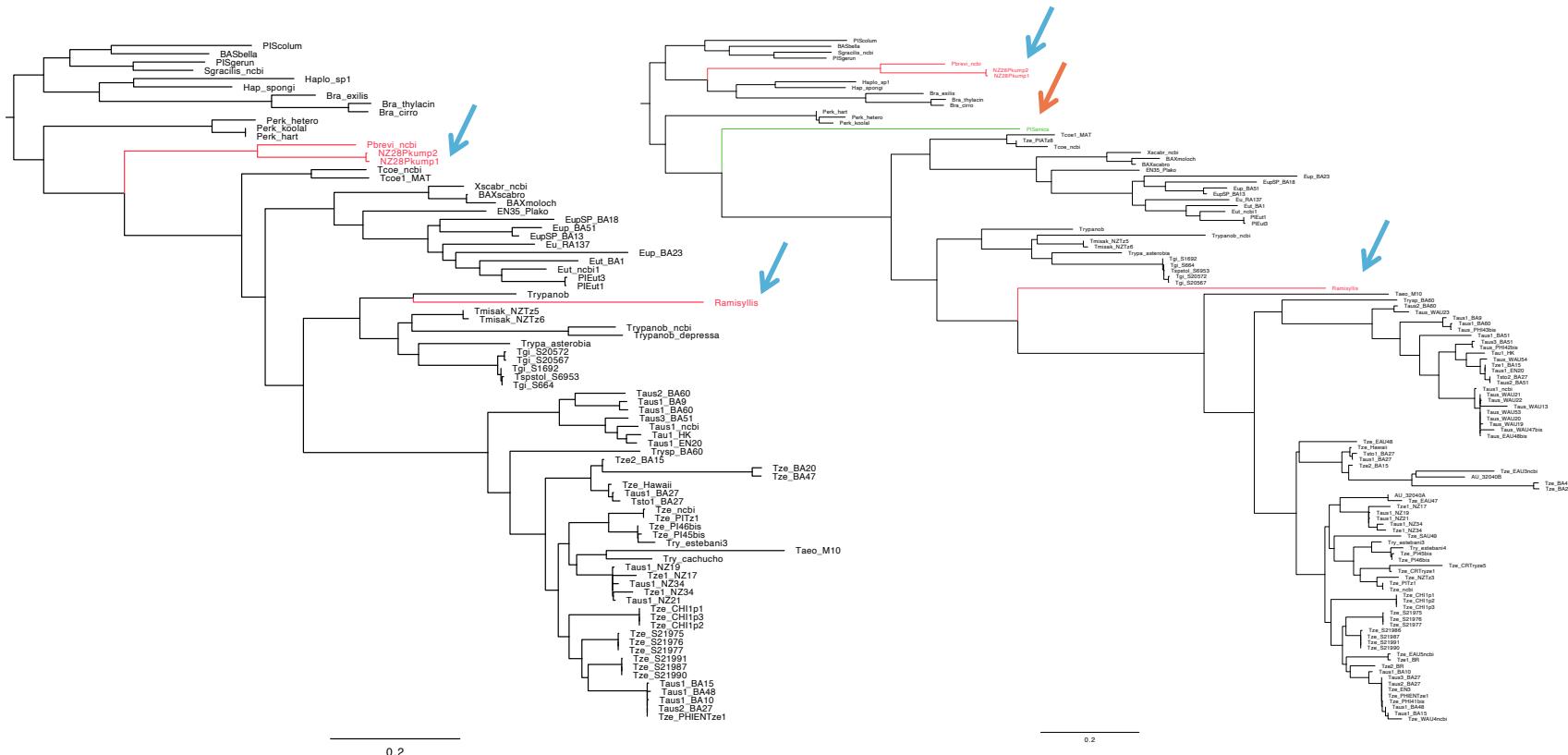


What do we need to build a phylogenetic tree

1. What taxa?

Your decision, but don't be biased!

The impact of **outgroup** and ingroup taxa selection

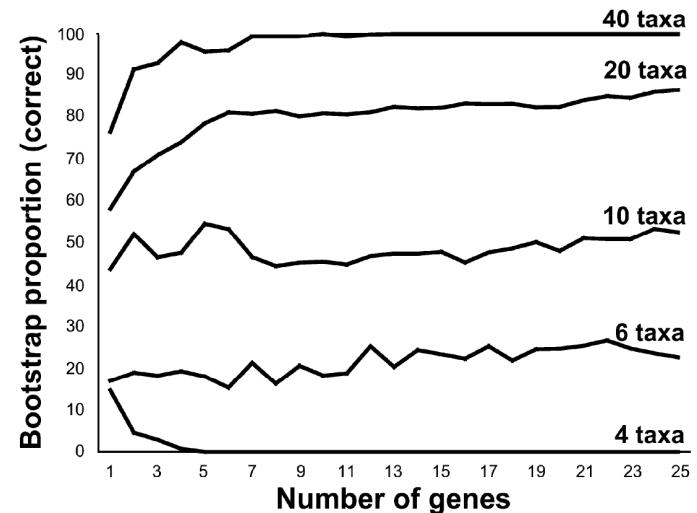
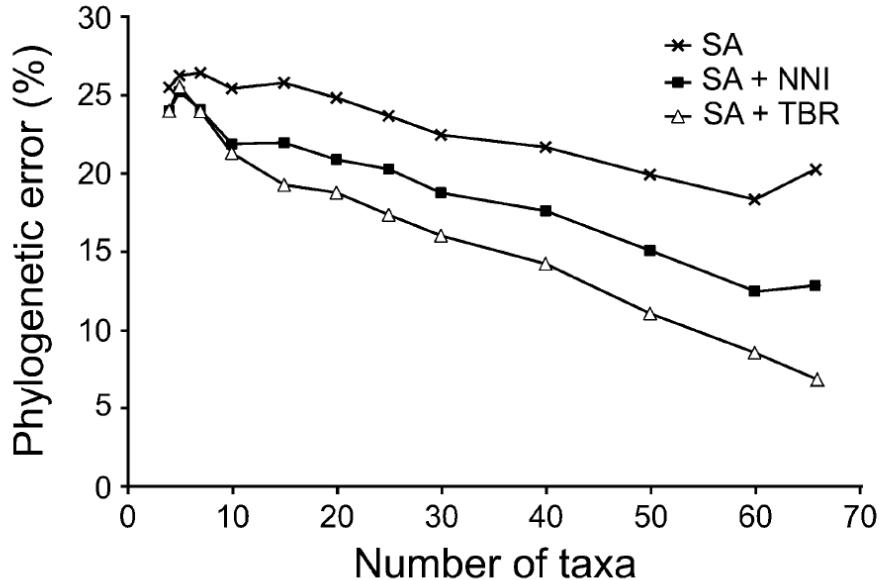


What do we need to build a phylogenetic tree

1. What taxa?

Your decision, but don't be biased!

The impact of outgroup and **ingroup** taxa selection



What do we need to build a phylogenetic tree

2. What type of data?

Your decision, but based on what is available!

3. How many markers?

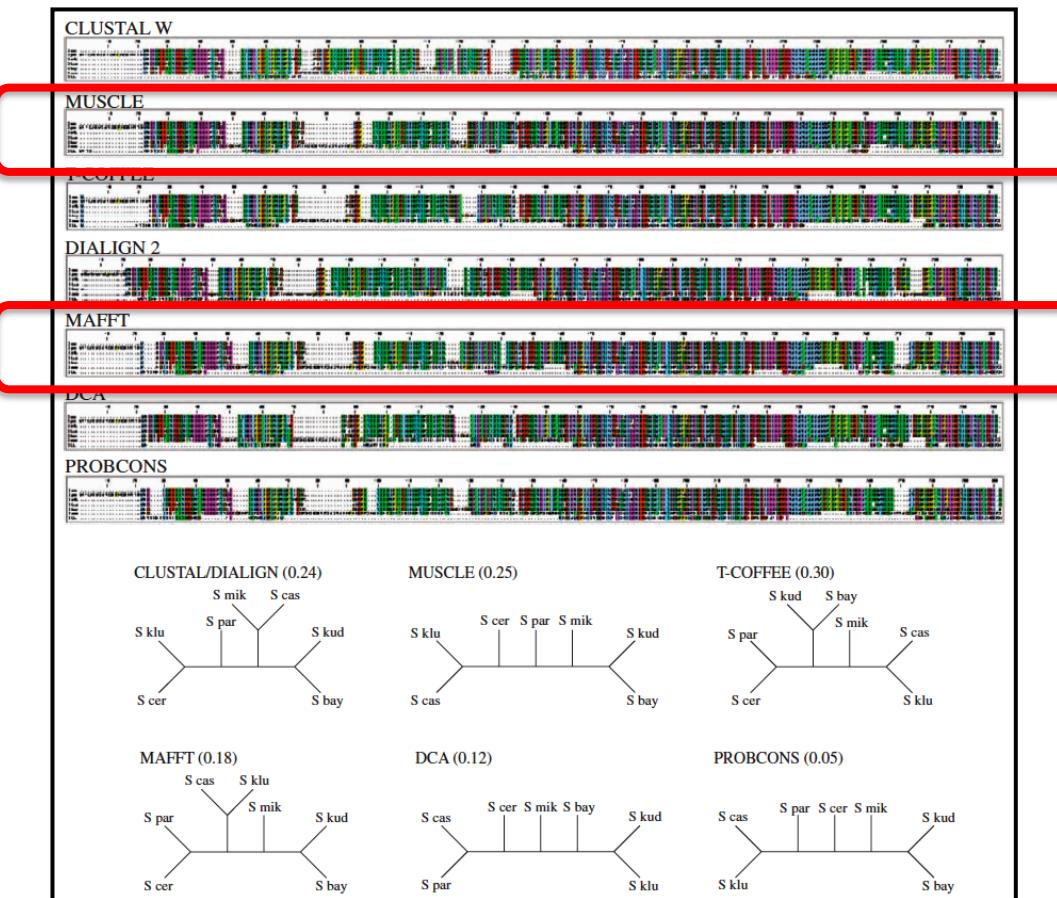
Variability vs. Conservation: Depending on your question, but most (small scale) phylogenetic studies now in **metazoans** use:

- Ribosomal genes (18S, 28S, 16S, ITS1&2) which are more conserved and/or
- Protein coding genes (COI, COB, ALG11, etc) which are more variable

Do your research...and check ncbi first to see what is available already

What do we need to build a phylogenetic tree

4. What alignment program?



Wong et al., 2008. Science

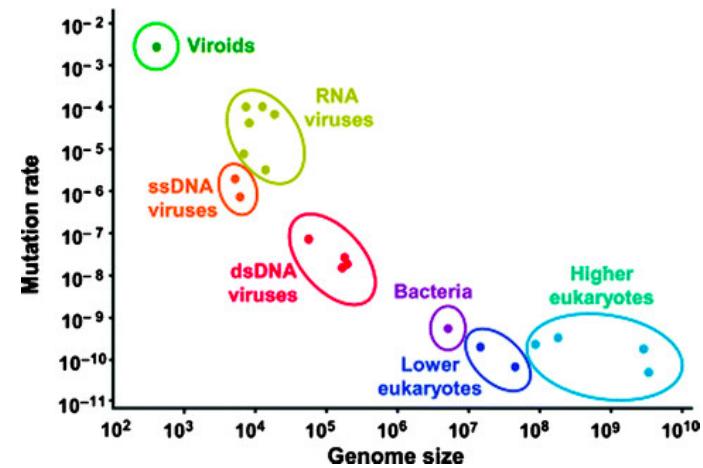
What do we need to build a phylogenetic tree

5. What model of evolution?

Not all genes evolve (mutate) at the same pace and in the same manner

Their mutation rates depend on many parameters:

- GC genome content
- Genome size
- Generation time
- Expression levels (usage of the gene)
- Protein coding genes
- Position in the genome
- Linkage disequilibrium
- etc



Each gene will have its own substitution model

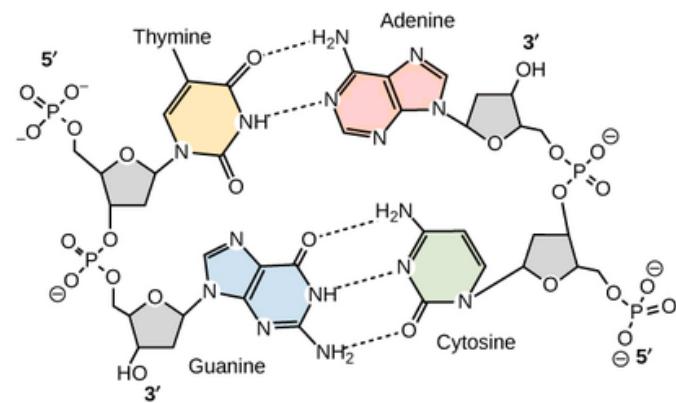
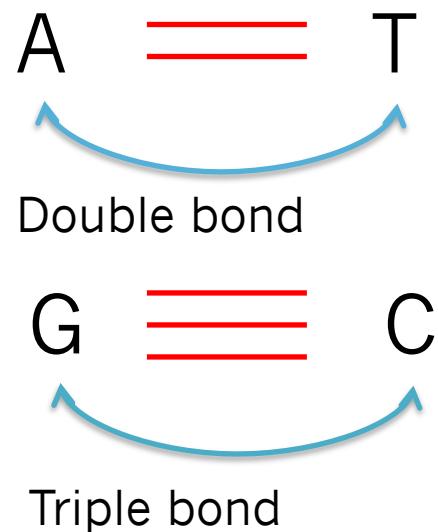


These substitution models differ in terms of the parameters used to describe the rates at which one nucleotide replaces another during evolution

What do we need to build a phylogenetic tree

5. What model of evolution?

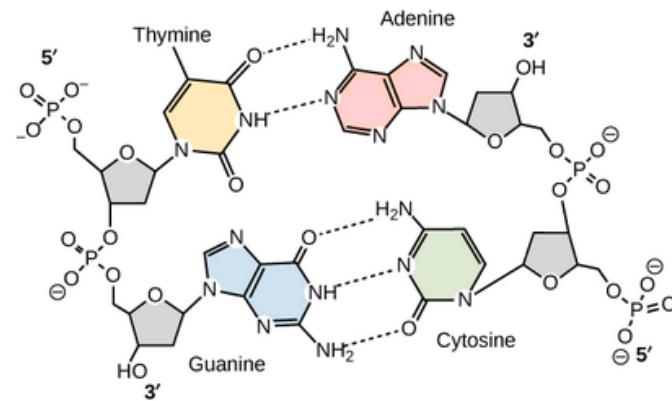
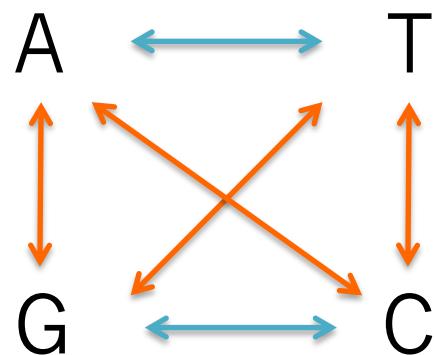
These substitution models differ in terms of the parameters used to describe the rates at which one nucleotide replaces another during evolution



What do we need to build a phylogenetic tree

5. What model of evolution?

These substitution models differ in terms of the parameters used to describe the rates at which one nucleotide replaces another during evolution



All mutations possible but with different frequencies

↔ transitions

↔ transversions

What do we need to build a phylogenetic tree

5. What model of evolution?

These substitution models differ in terms of the parameters used to describe the rates at which one nucleotide replaces another during evolution

Jukes & Cantor 1969: it assumes the same base frequencies and mutation rates

$$\left(\pi_A = \pi_G = \pi_C = \pi_T = \frac{1}{4} \right)$$

K80 (Kimura 1980): different rates for transitions/transversions

....(many more)

GTR (Tavaré 1986): the most complex and most used

$$\alpha = r(A \rightarrow G) = r(G \rightarrow A)$$

$$\beta = r(A \rightarrow C) = r(C \rightarrow A)$$

$$\gamma = r(A \rightarrow T) = r(T \rightarrow A)$$

$$\delta = r(G \rightarrow C) = r(C \rightarrow G)$$

$$\epsilon = r(G \rightarrow T) = r(T \rightarrow G)$$

$$\eta = r(C \rightarrow T) = r(T \rightarrow C)$$

substitution rates

$$\Pi = (\pi_A, \pi_G, \pi_C, \pi_T), \quad Q = \begin{pmatrix} -(\alpha\pi_G + \beta\pi_C + \gamma\pi_T) & \alpha\pi_G & \beta\pi_C & \gamma\pi_T \\ \alpha\pi_A & -(\alpha\pi_A + \delta\pi_C + \epsilon\pi_T) & \delta\pi_C & \epsilon\pi_T \\ \beta\pi_A & \delta\pi_G & -(\beta\pi_A + \delta\pi_G + \eta\pi_T) & \eta\pi_T \\ \gamma\pi_A & \epsilon\pi_G & \eta\pi_C & -(\gamma\pi_A + \epsilon\pi_G + \eta\pi_C) \end{pmatrix}$$

equilibrium vectors

What do we need to build a phylogenetic tree

5. What model of evolution?

jModelTest 2.1.4

File Edit Analysis Results Tools Help About

Main PhyML-log

```
--- jModeltest 2.1.4 ---
(c) 2011-onwards D. Darriba, G.L. Taboada, R. Doallo and D. Posada,
(1) Department of Biochemistry, Genetics and Immunology
    University of Vigo, 36310 Vigo, Spain.
(2) Department of Electronics and Systems
    University of A Coruna, 15071 A Coruna, Spain.
e-mail: ddarriba@udc.es, dposada@uvigo.es
-----
Sat Feb 08 13:22:06 EET 2014
Windows NT (unknown) 6.2, arch: x86, bits: 32, numcores: 4
-----
Citation: Darriba D, Taboada GL, Doallo R and Posada D. 2012.
    "jModelTest 2: more models, new heuristics and parallel computing".
    Nature Methods 9(8), 772.
-----
Reading data file "aP6.fas"... OK.
    number of sequences: 6
    number of sites: 631
Reading data file "gusanosCOI.mafft.fas"... OK.
    number of sequences: 44
    number of sites: 561
Reading data file "primate-mtDNA.nex"... OK.
    number of sequences: 12
    number of sites: 898

```

Likelihood scores not available

primate-mtDNA.nex



What do we need to build a phylogenetic tree

5. What phylogenetic method?



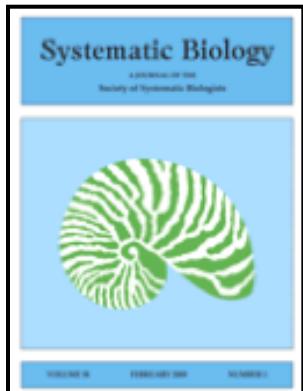
This is the most important decision of your life

What do we need to build a phylogenetic tree

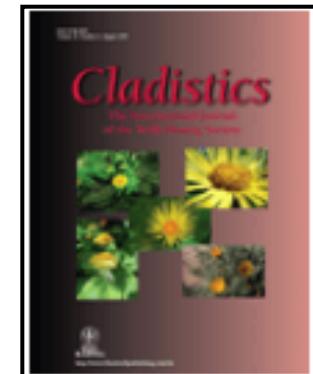
5. What phylogenetic method?



- Statistical phylogenetics
- Probabilistic methods
- Society for Systematic Biology



- Cladistics
- Parsimony
- Willi Hennig Society



What do we need to build a phylogenetic tree

5. What phylogenetic method?

- Statistical
- Probabilistic
- Society for
Biology

Society



What do we need to build a phylogenetic tree

5. What phylogenetic method?

purportedly sophisticated.

Phylogenetic data sets submitted to this journal should be analysed using parsimony. If alternative methods are also used and there is no difference among the results, the author should defer to the principles of the Society and present the tree obtained by parsimony. Unless there is a pertinent reason to include multiple trees from alternative methods, a tree based on parsimony is sufficient as an intelligible, informative and repeatable hypothesis of relationships, and articles should not be cluttered with multiple, often redundant, trees produced from other methods.

 **Christoph Bleidorn** 
@C_Blei

A new Editorial from Cladistics. Back to the Cold War of systematists...onlinelibrary.wiley.com/doi/10.1111/cl.12345/full
#phylogeny

1:14 PM - Jan 13, 2016

 19  52 people are talking about this

This is the most important decision of your life...not really

 **Christian Ray** 
@jcjray

Replying to @C_Blei
@C_Blei Bbb..but the nature of historical individuality precludes quantitative statistical analysis in biology! Parsimony cures cancer.

3:26 PM - Jan 15, 2016

 2  See Christian Ray's other Tweets 

What do we need to build a phylogenetic tree

5. What phylogenetic method?

Distance vs discrete data
(distance vs MP, ML, Bayesian)

Or
MP vs model-based methods
(MP vs distance, ML, Bayesian)

What do we need to build a phylogenetic tree

5. What phylogenetic method? DISTANCE METHODS

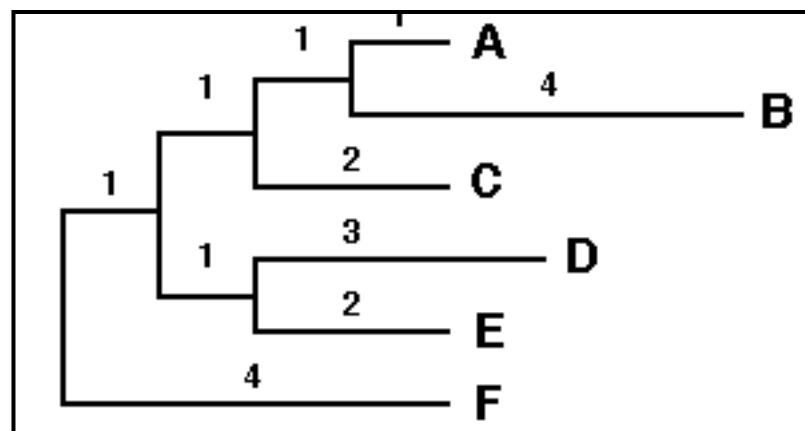
Human	aactc	
Chimp	aagtc	
Orang	tagtt	
becomes		
H	C	O
H	-	1 3
C	1	- 2
O	3	2 -

Distance matrix

	A	B	C	D	E
B	2				
C	4	4			
D	6	6	6		
E	6	6	6	4	
F	8	8	8	8	8

With or without evolutionary model behind

NJ (model corrected)



Superceded by probabilistic methods
DO NOT USE IT

What do we need to build a phylogenetic tree

5. What phylogenetic method? MAXIMUM PARSIMONY

The assumption is that the ‘true’ tree will contain the least number of mutations possible i.e., the most parsimonious solution

Given a set of sequences (partial evidence) we need to find ancestral sequences, build a rooted tree, and estimate the smallest number of changes contained in the branches

It poses an impossible computational problem (no algorithm known) and therefore all models use simplified versions

Weighted parsimony: different scoring for each change (transition/transversion) and different scoring for each position

What do we need to build a phylogenetic tree

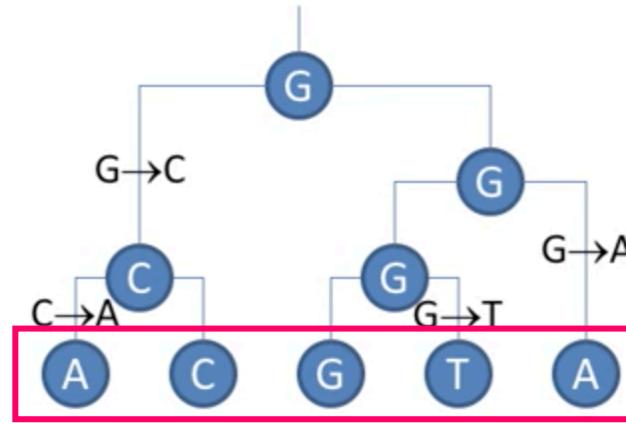
5. What phylogenetic method? MAXIMUM PARSIMONY

	1	2	3	4	5	6	7	8	9	10
Species 1 -	A	G	G	G	T	A	A	C	T	G
Species 2 -	A	C	G	A	T	T	A	T	T	A
Species 3 -	A	T	A	A	T	T	G	T	C	T
Species 4 -	A	A	T	G	T	T	G	T	C	G

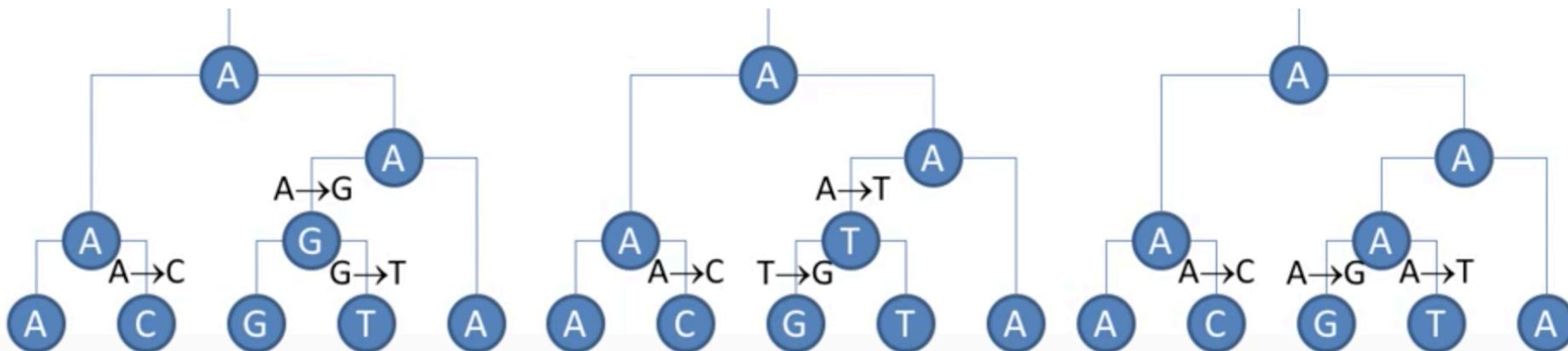
Informative sites

What do we need to build a phylogenetic tree

5. What phylogenetic method? MAXIMUM PARSIMONY

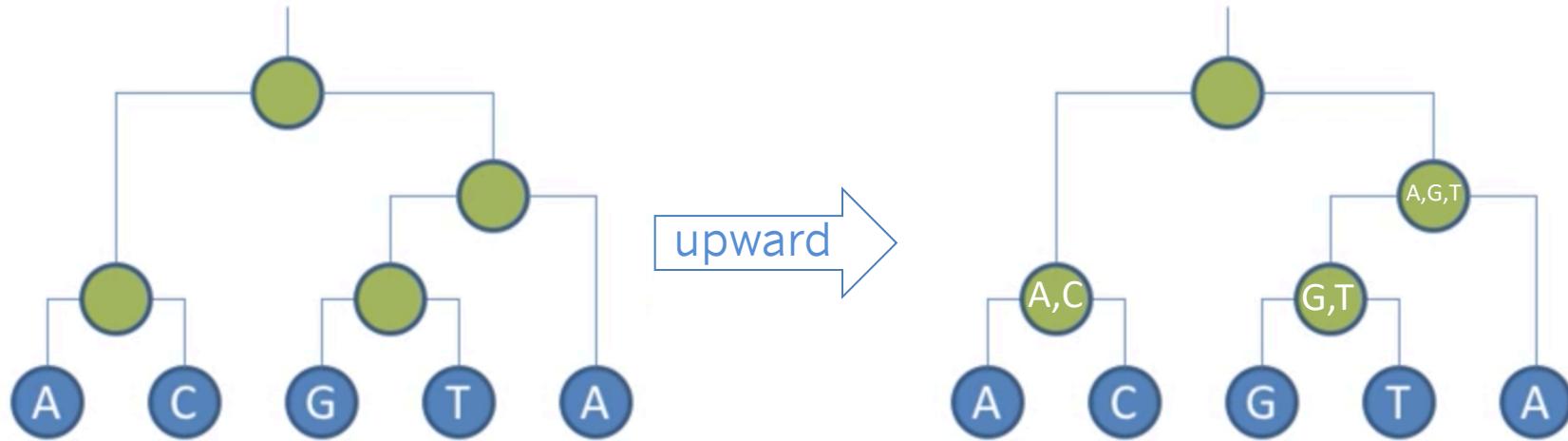


Fitch's algorithm
to find the
ancestral states:
upward and
downward phase



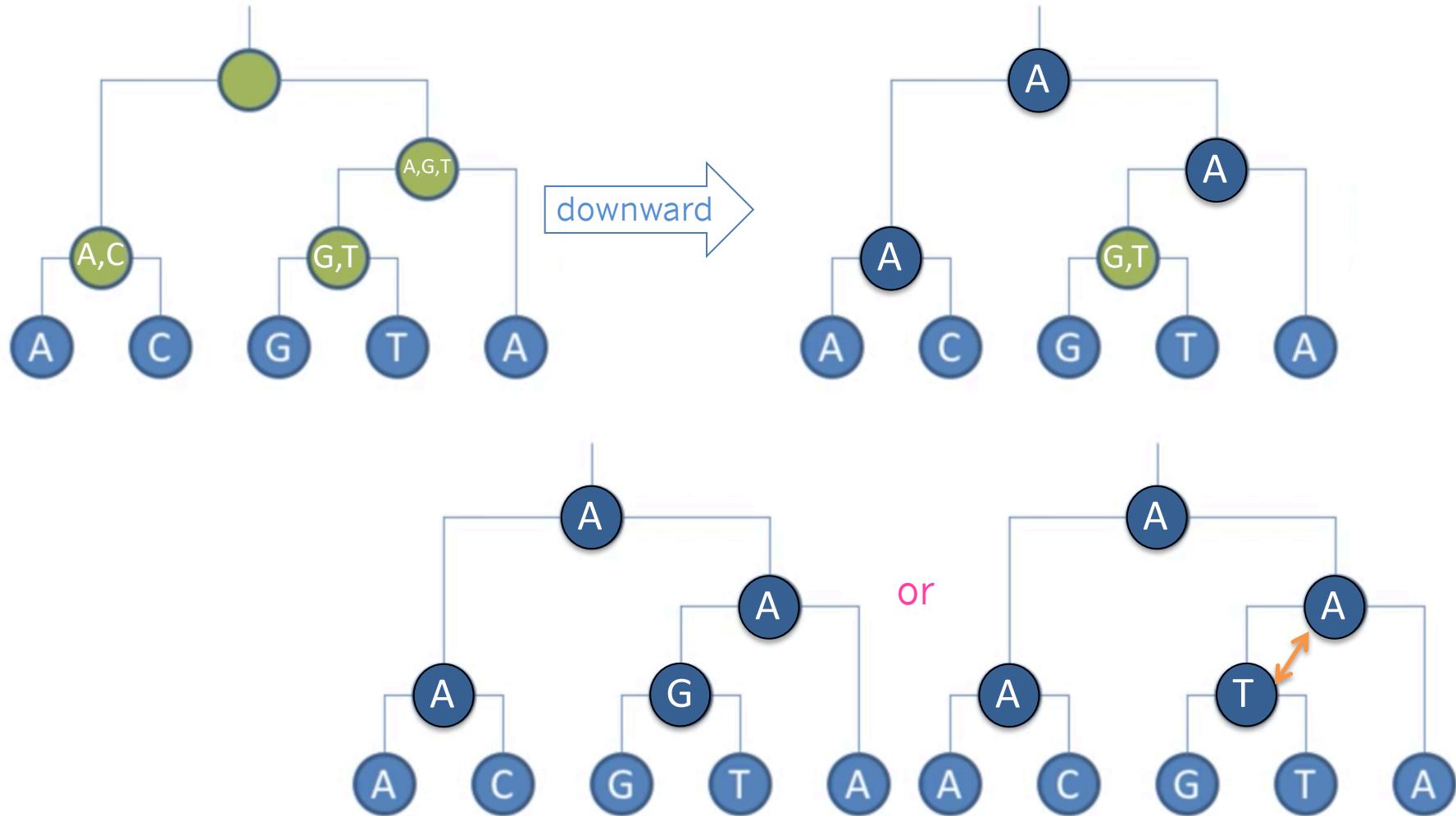
What do we need to build a phylogenetic tree

5. What phylogenetic method? MAXIMUM PARSIMONY



What do we need to build a phylogenetic tree

5. What phylogenetic method? MAXIMUM PARSIMONY



What do we need to build a phylogenetic tree

5. What phylogenetic method? MAXIMUM LIKELIHOOD

Likelihood: the probability of producing the observed data by a model given the model parameters, $L_D = \Pr(X|\Theta)$

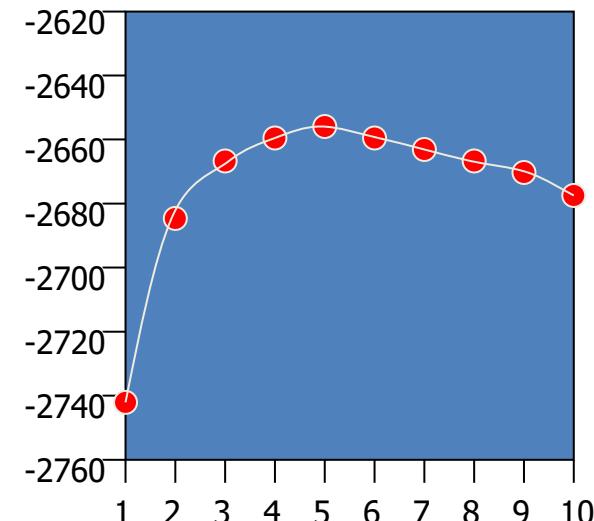
X : data already aligned, each site mutates independently

Θ : parameters of the model (topology, model of substitution)

In big datasets the computation is extremely difficult and long, BUT...

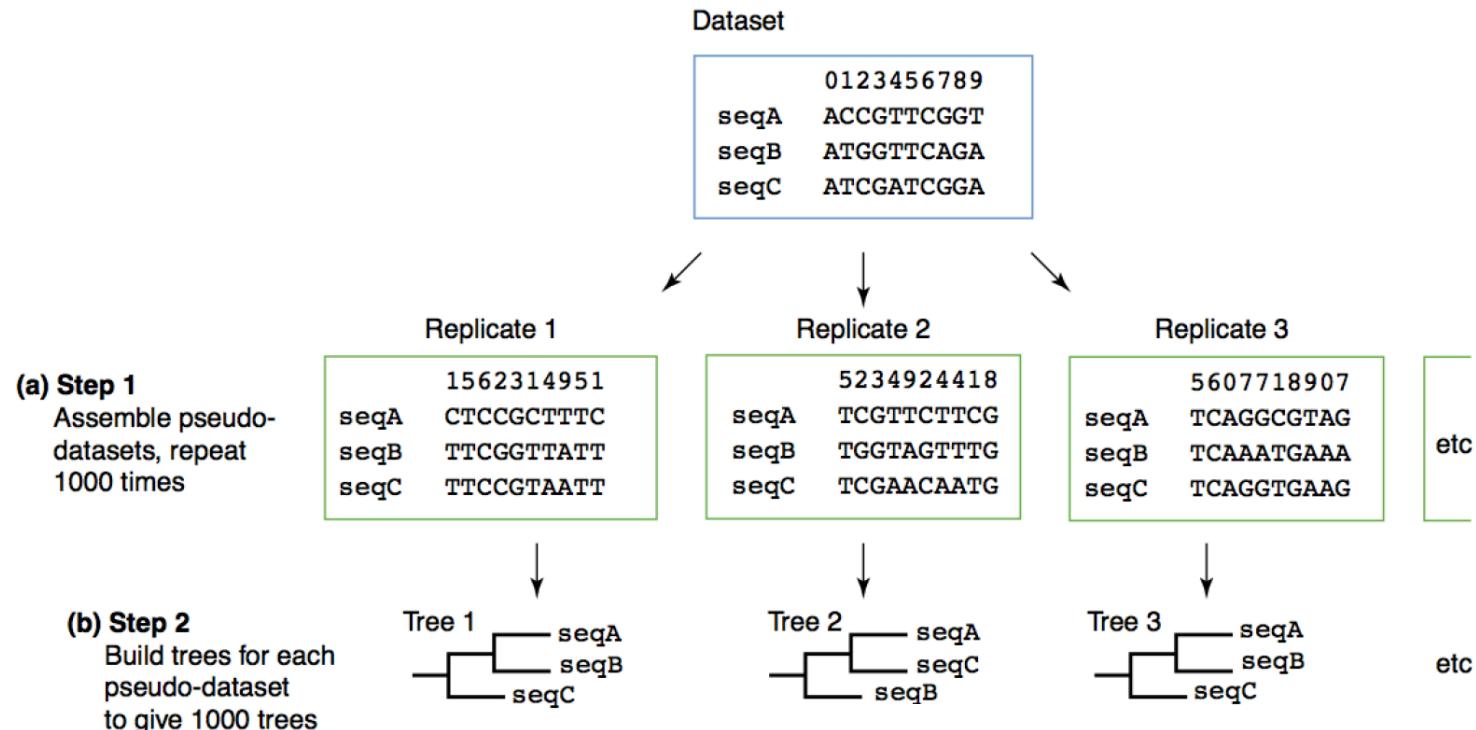
...it is one of the **most powerful** methods because:

- it uses models of substitution (**evolutionary models**)
- it corrects multiple substitutions
- it allows estimation of branch length (=amount of change from ancestor)
- fixing all but one parameter it allows finding the max. Likelihood for that parameter



What do we need to build a phylogenetic tree

5. What phylogenetic method? MAXIMUM LIKELIHOOD



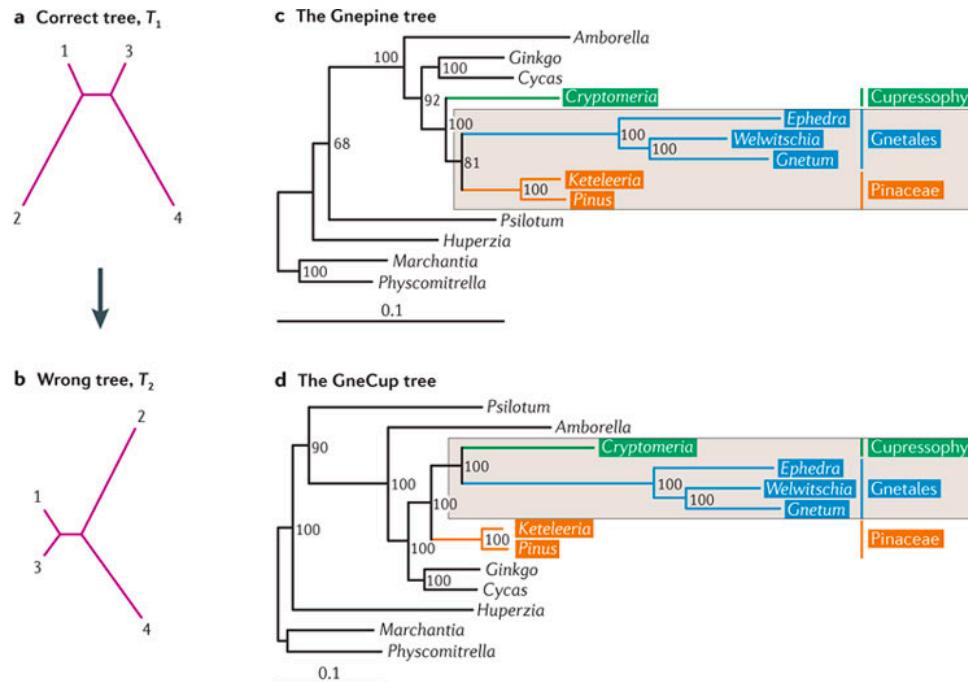
What do we need to build a phylogenetic tree

5. What phylogenetic method? MAXIMUM LIKELIHOOD

Problems with maximum likelihood (and parsimony)

Long branch attraction

is a phenomenon when **rapidly evolving lineages** are inferred to be closely related, regardless of their true evolutionary relationships



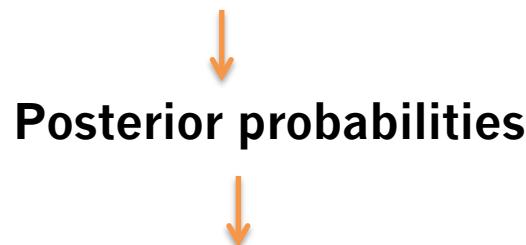
What do we need to build a phylogenetic tree

5. What phylogenetic method? BAYESIAN INFERENCE

Reverend Thomas Bayes 1763: inverse probability

Based on Bayes' theorem, the bayesian approach combines the prior probability of a tree $P(A)$ with the likelihood of the data (B) to produce a posterior probability distribution on trees $P(A|B)$.

Likelihood gives you the probability of the data given the hypothesis and Bayesian gives you the probability of the hypothesis given the data



It uses MCMC (Markov chain-Montecarlo) algorithms

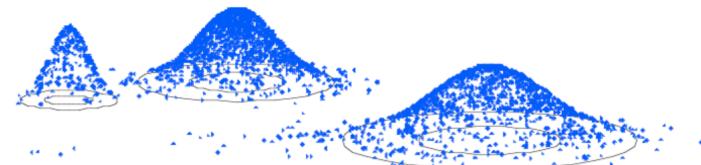
What do we need to build a phylogenetic tree

5. What phylogenetic method? BAYESIAN INFERENCE

How does it work?

- ✓ Start somewhere, that somewhere will have a likelihood and a prior
- ✓ Randomly propose a new state (maybe adjust one branch length), if the change has a better likelihood \times prior, the chain goes there
- ✓ Calculate the posterior probability ratio between the current and previous state. It should be between 0 and 1
- ✓ Choose a random number between 0 and 1 and if that number has a better likelihood than the ratio of states, the change is accepted (sometimes if slightly worse as well)
- ✓ That is how the chain crosses likelihood valleys

<http://setosa.io/ev/markov-chains/>

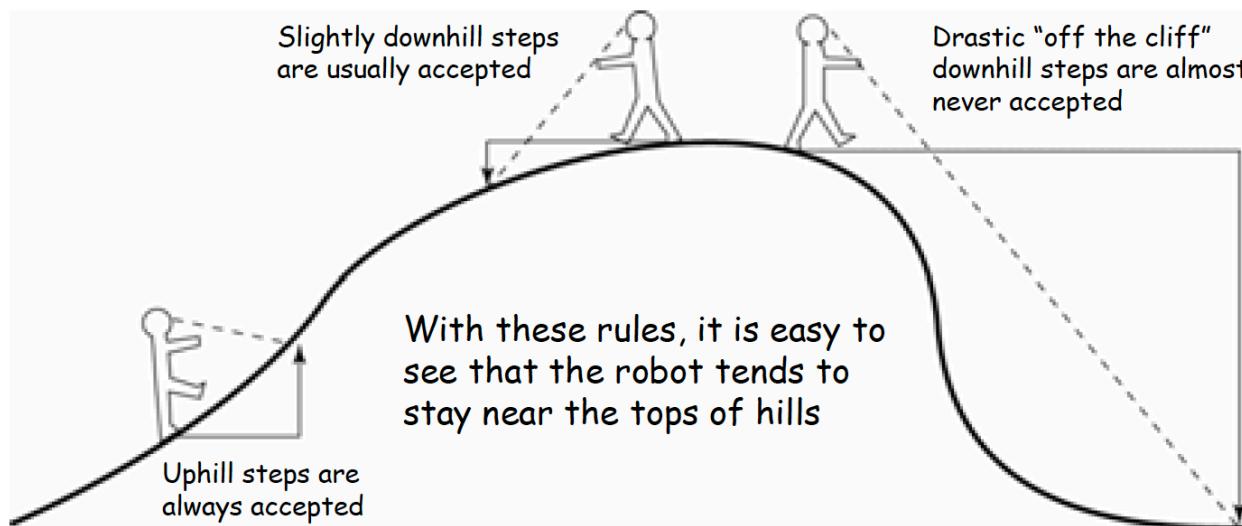


What do we need to build a phylogenetic tree

5. What phylogenetic method? BAYESIAN INFERENCE

How does it work?

MCMC robot's rules

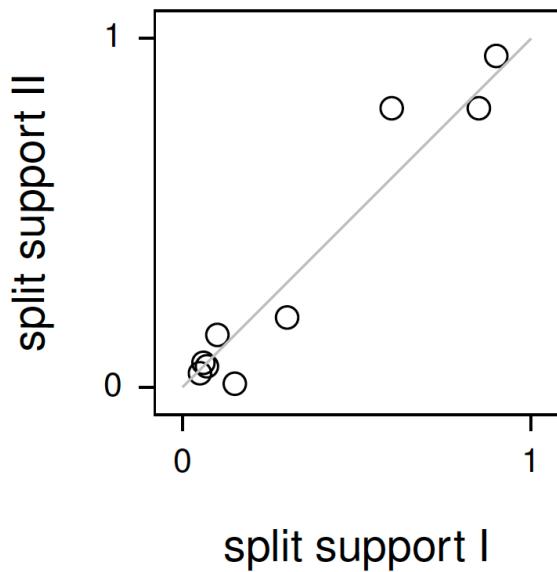


What do we need to build a phylogenetic tree

5. What phylogenetic method? BAYESIAN INFERENCE

Two runs

Check the split frequencies between both runs



Topological agreement between 2 runs in the form of a single number
It should get to <0.01

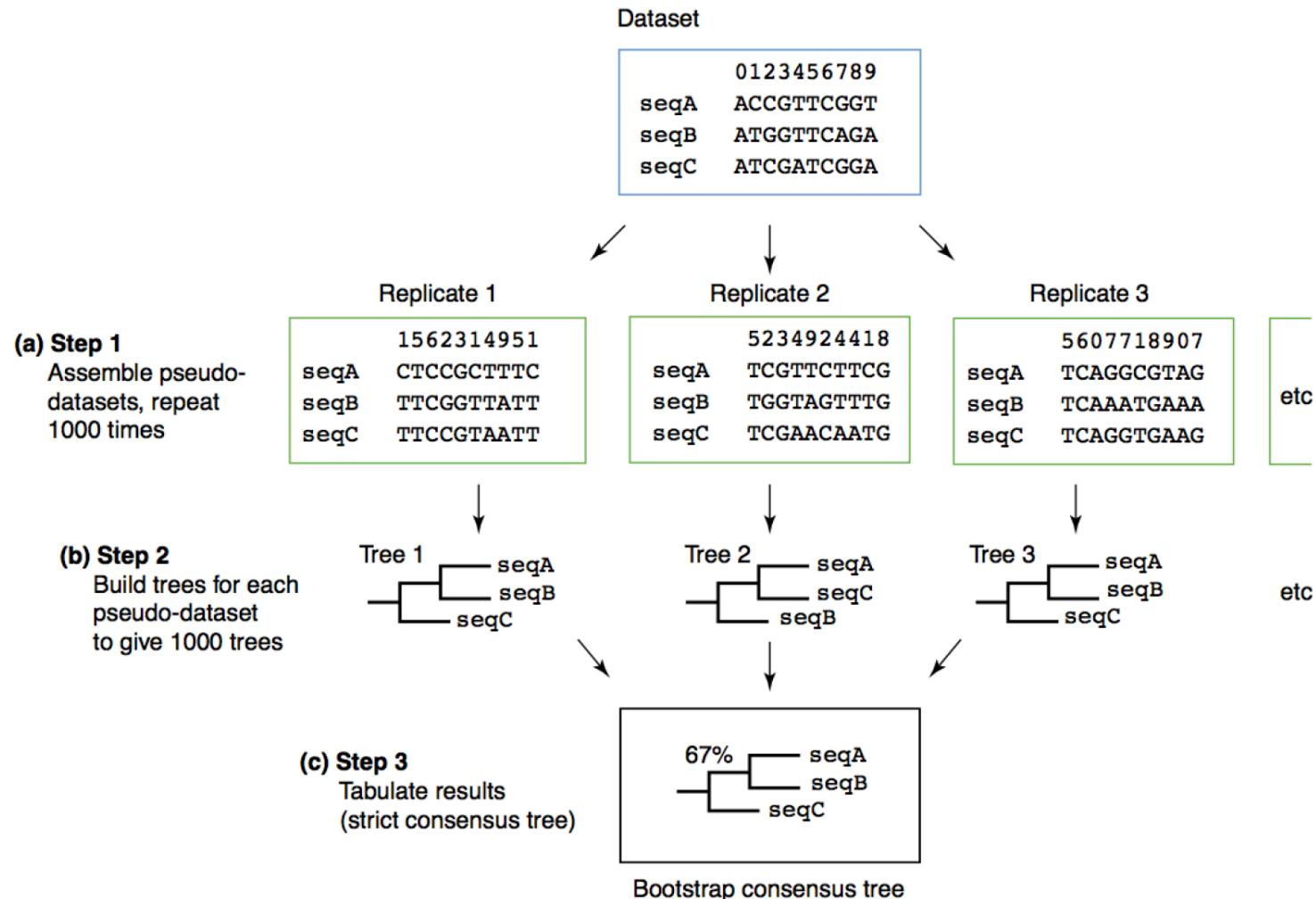
What do we need to build a phylogenetic tree

6. How to assess confidence? Evaluating support

- Bremer support: difference in the branch lengths when clades are removed
(only in parsimony) **over 70%**
- Jackknife: probability of a clade observed in all the trees (parsimony)
over 70%
- Bootstrap: probability of a clade observed in all the trees (ML) **over 70%**
- Posterior probability: probability of a clade being assigned under the conditions sampled (BI) **over 0.95**
- Convergence: assessing whether all chains (MCMC) converged in the same solution (BI)

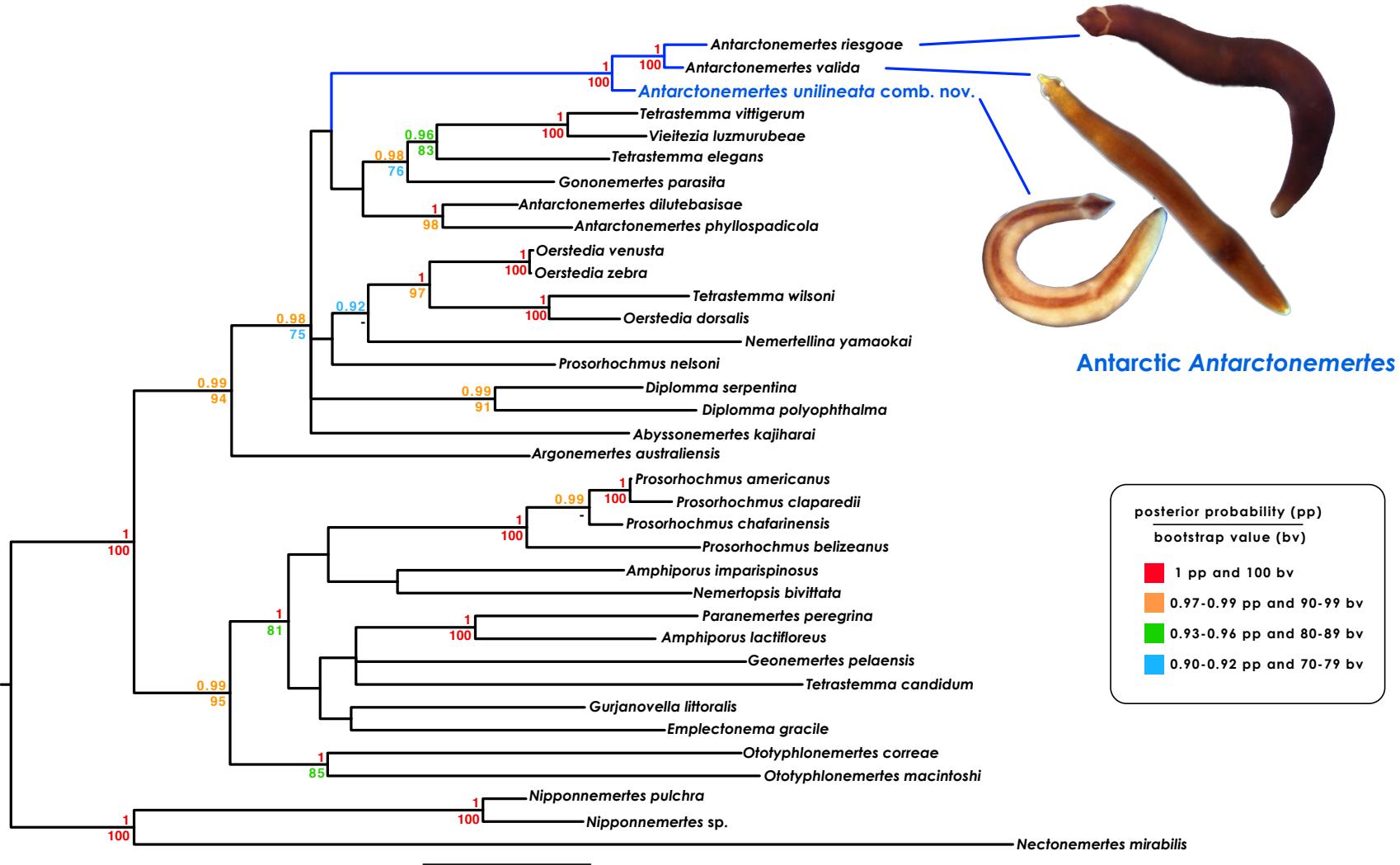
What do we need to build a phylogenetic tree

6. How to assess confidence? Evaluating support



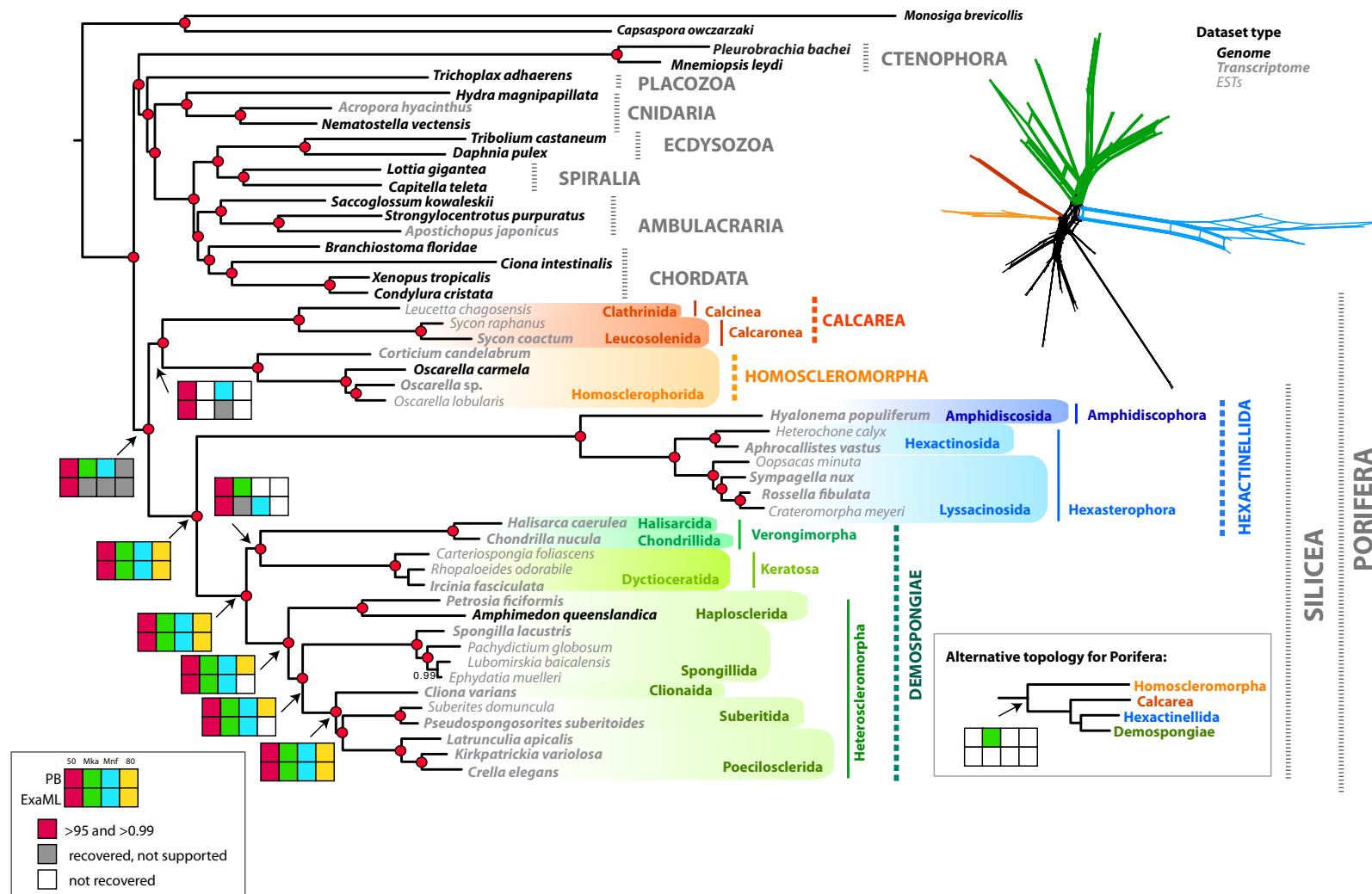
What do we need to build a phylogenetic tree

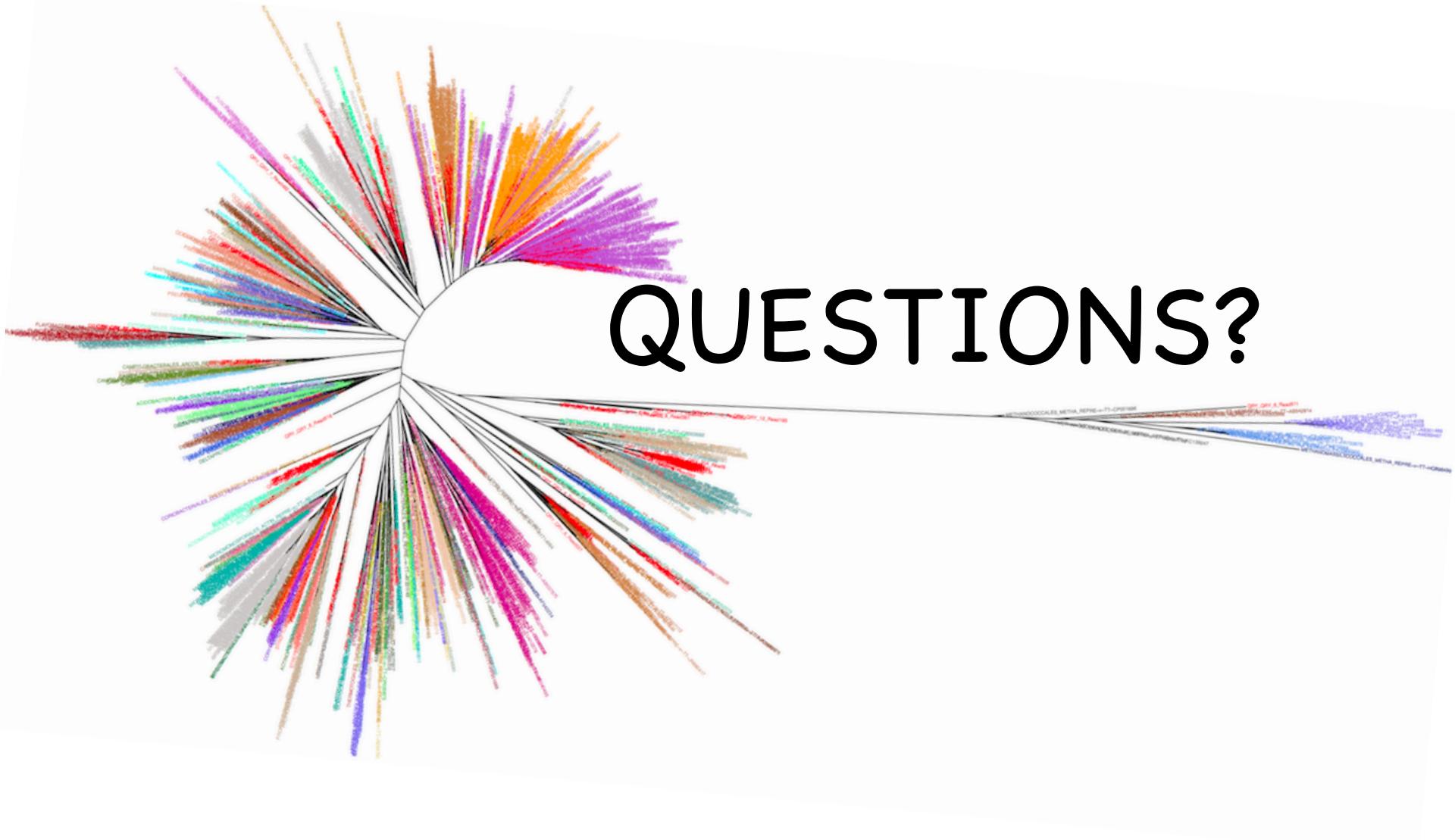
6. How to assess confidence? Evaluating support



What do we need to build a phylogenetic tree

6. How to assess confidence? Evaluating support





QUESTIONS?