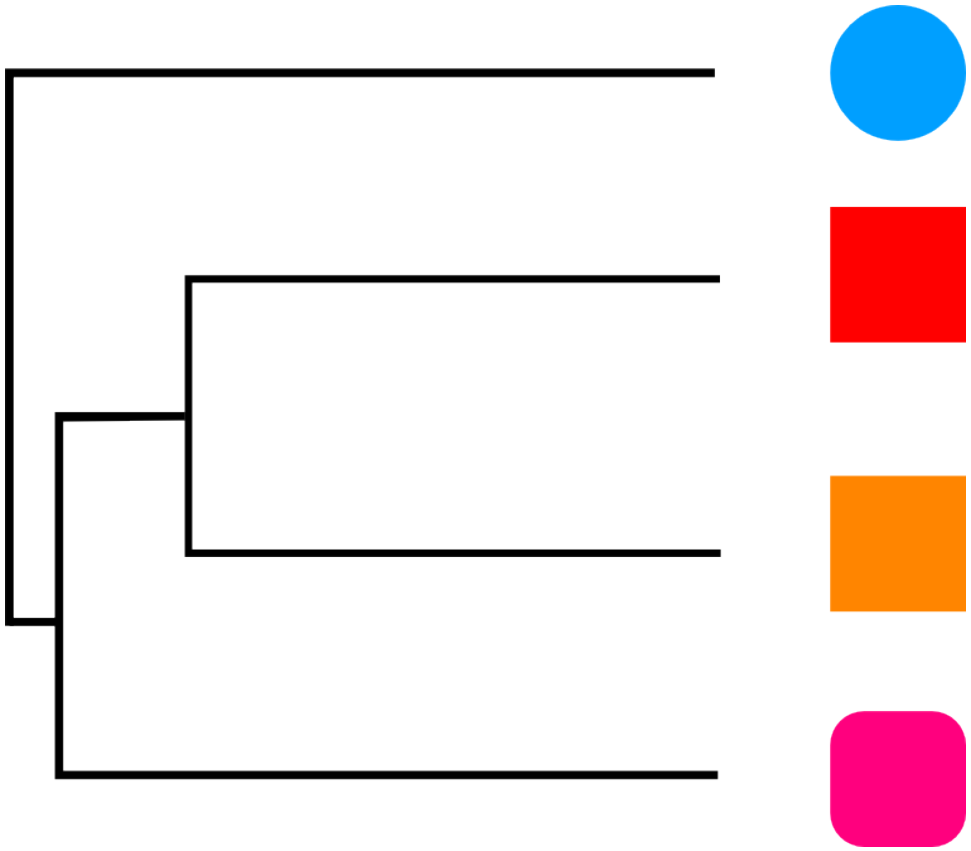
MDSC 308

Phylogenetics

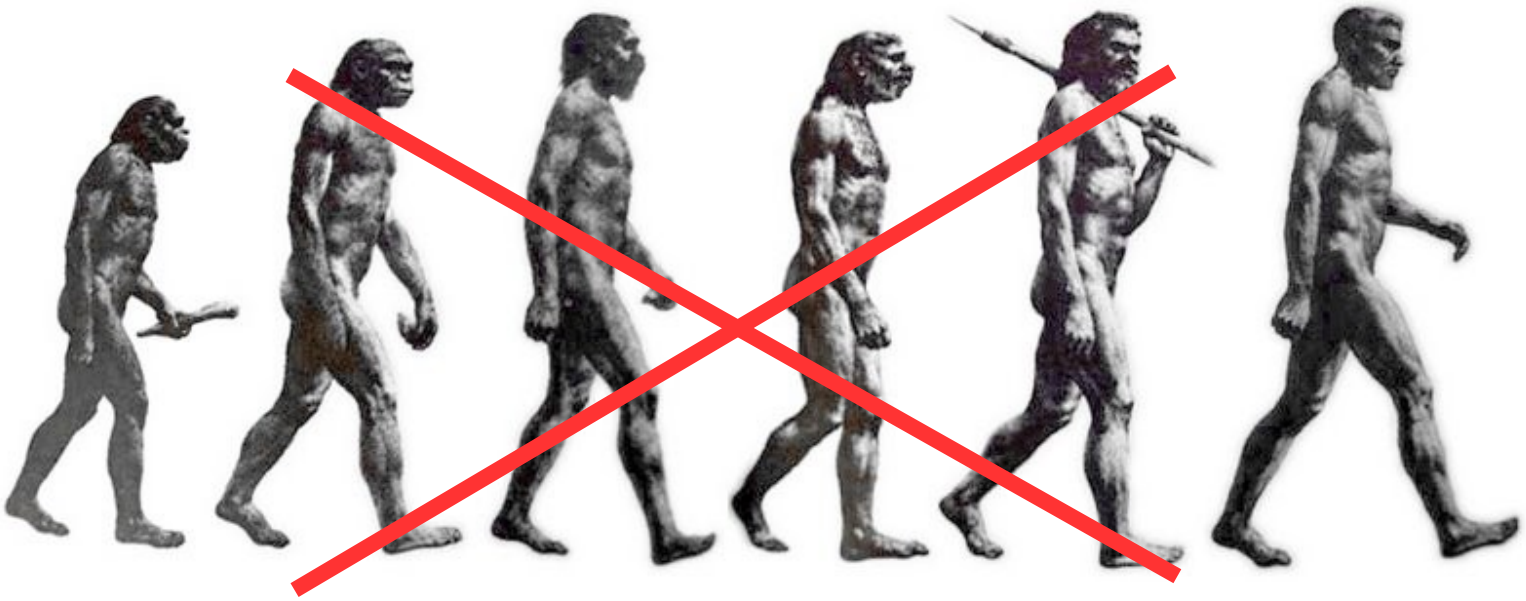
“Nothing in Biology Makes Sense Except in the Light of Evolution”

Theodosius Dobzhansky, 1973

Comparing things

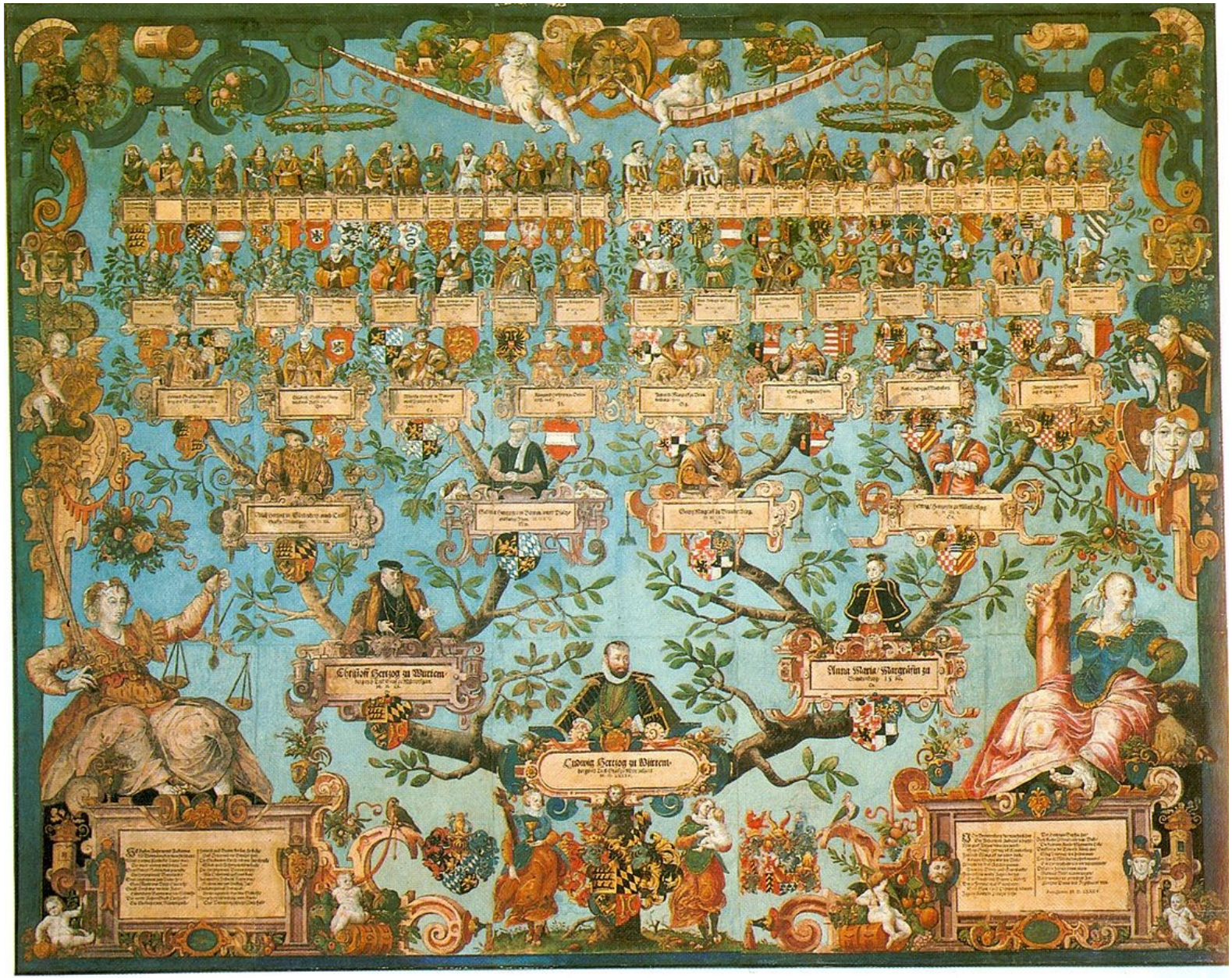


Evolution?



WRONG

Kingdom of Wurttemberg



Trees

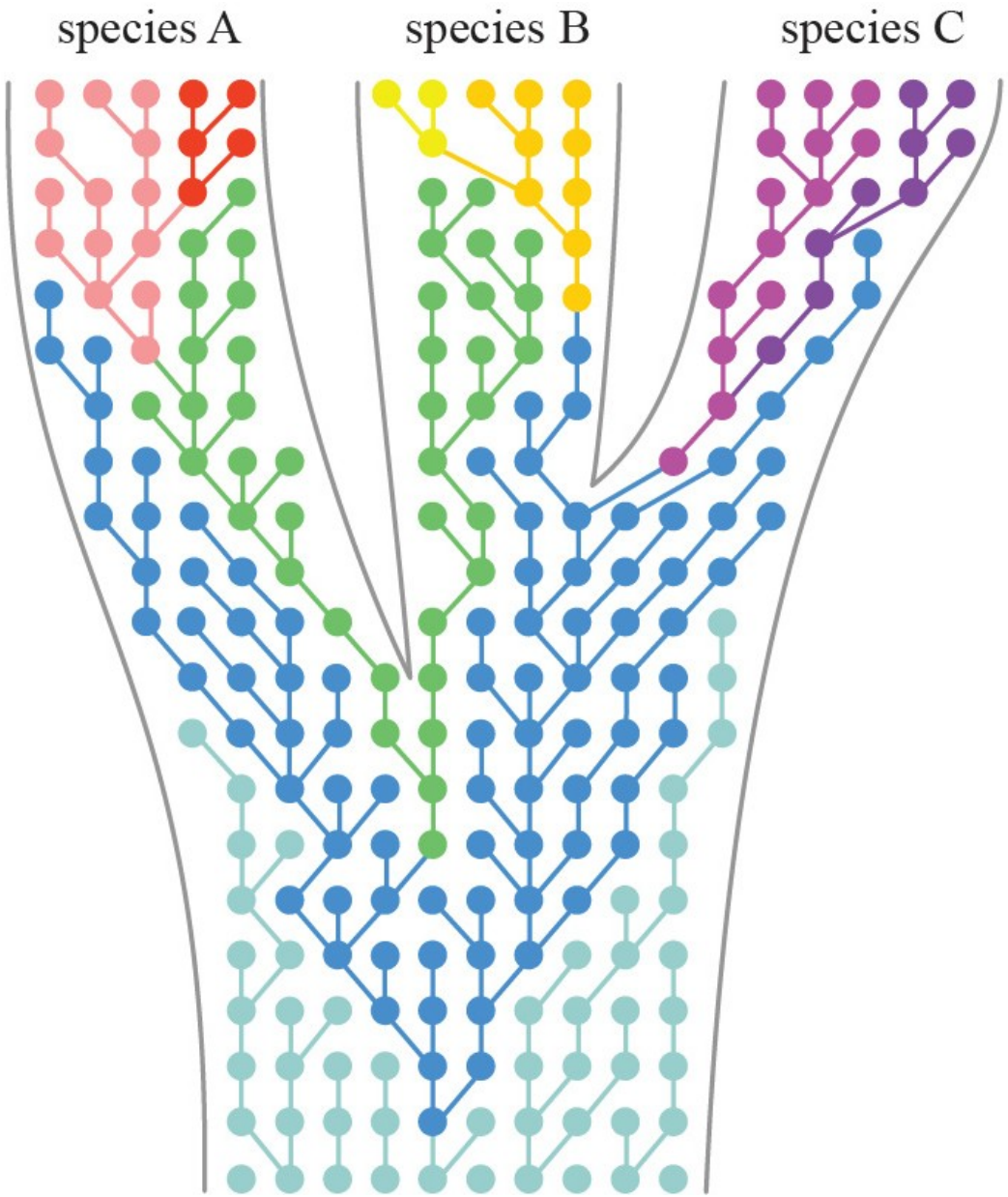
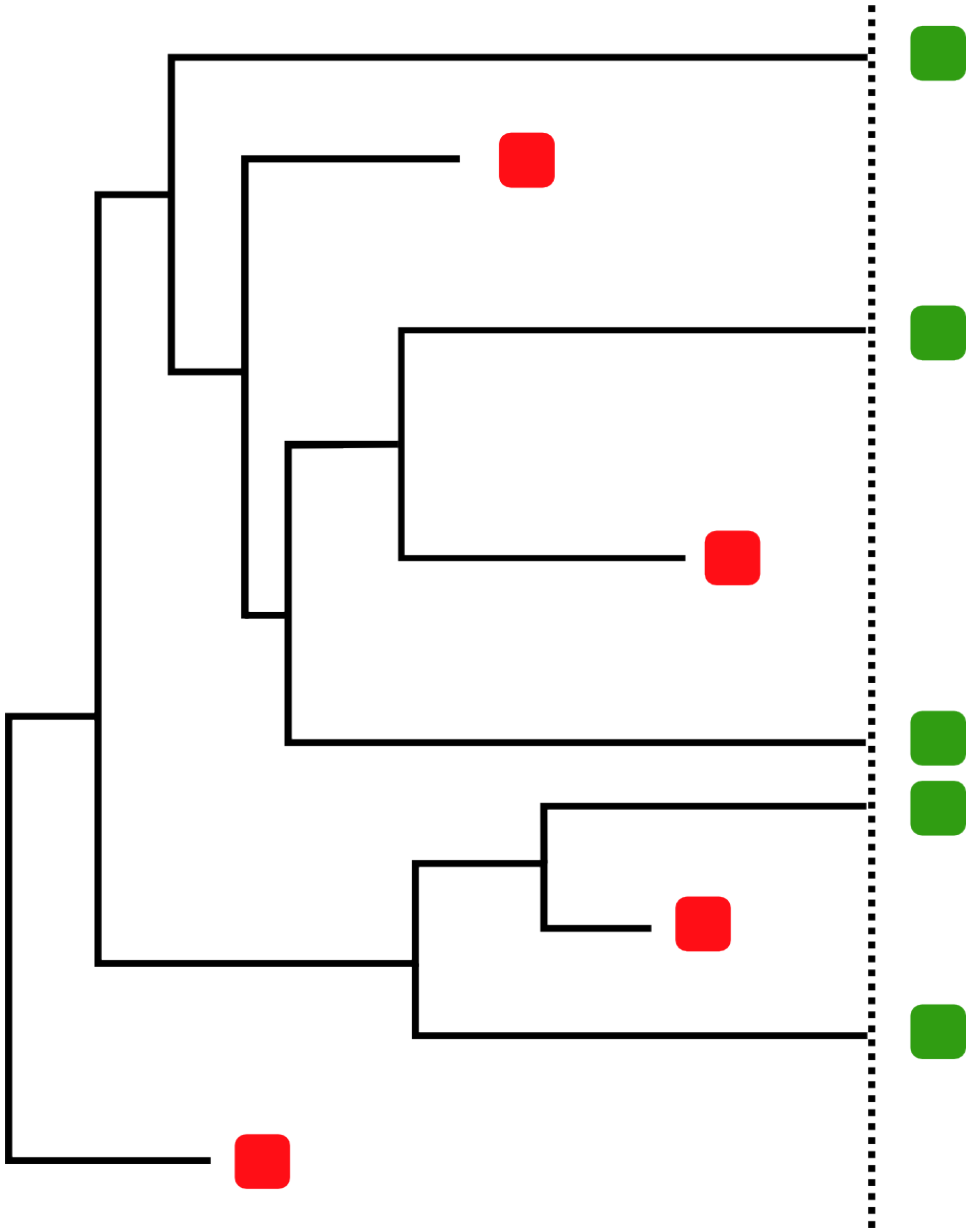
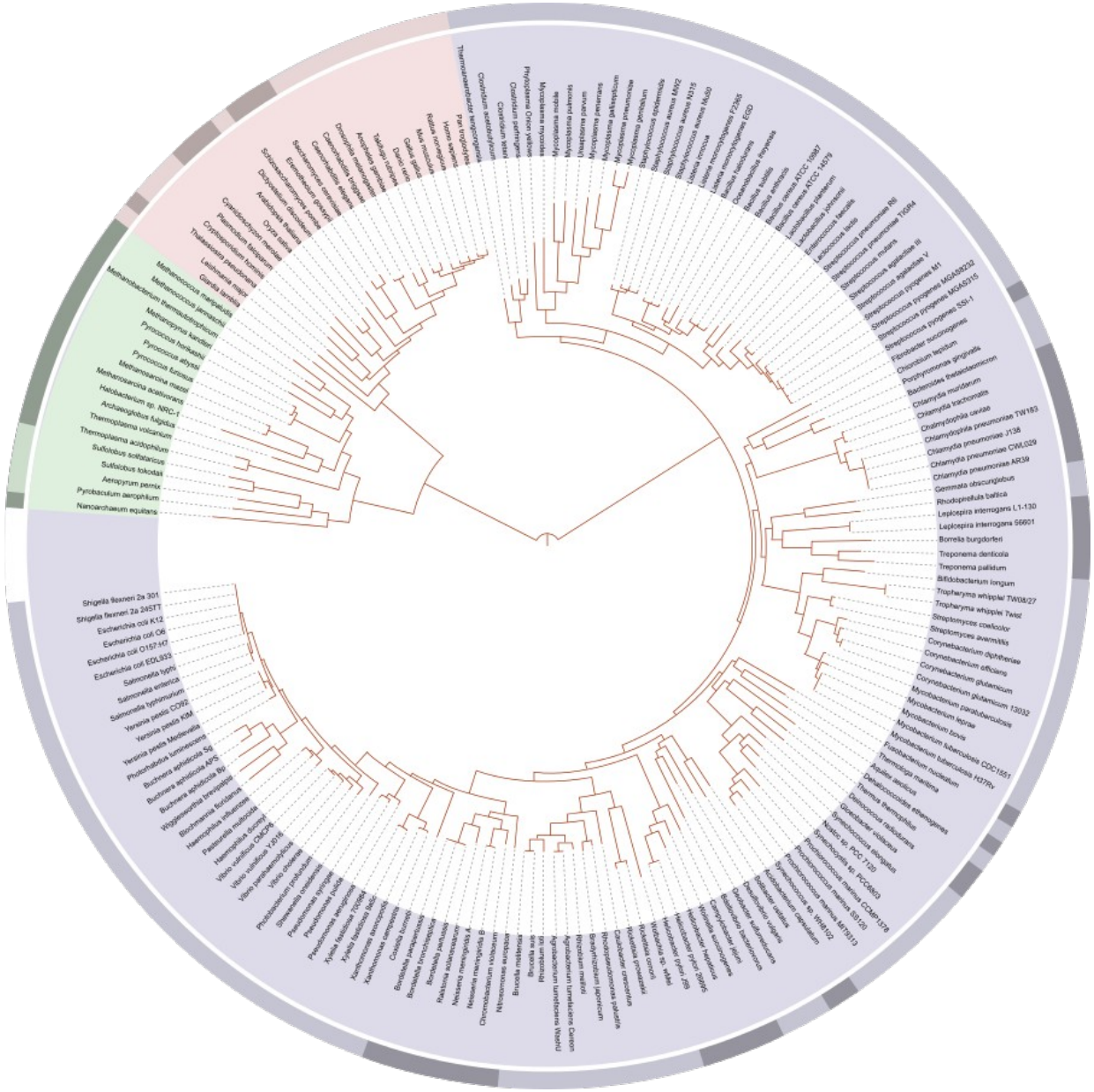


Figure author: Fredrick Leliaert

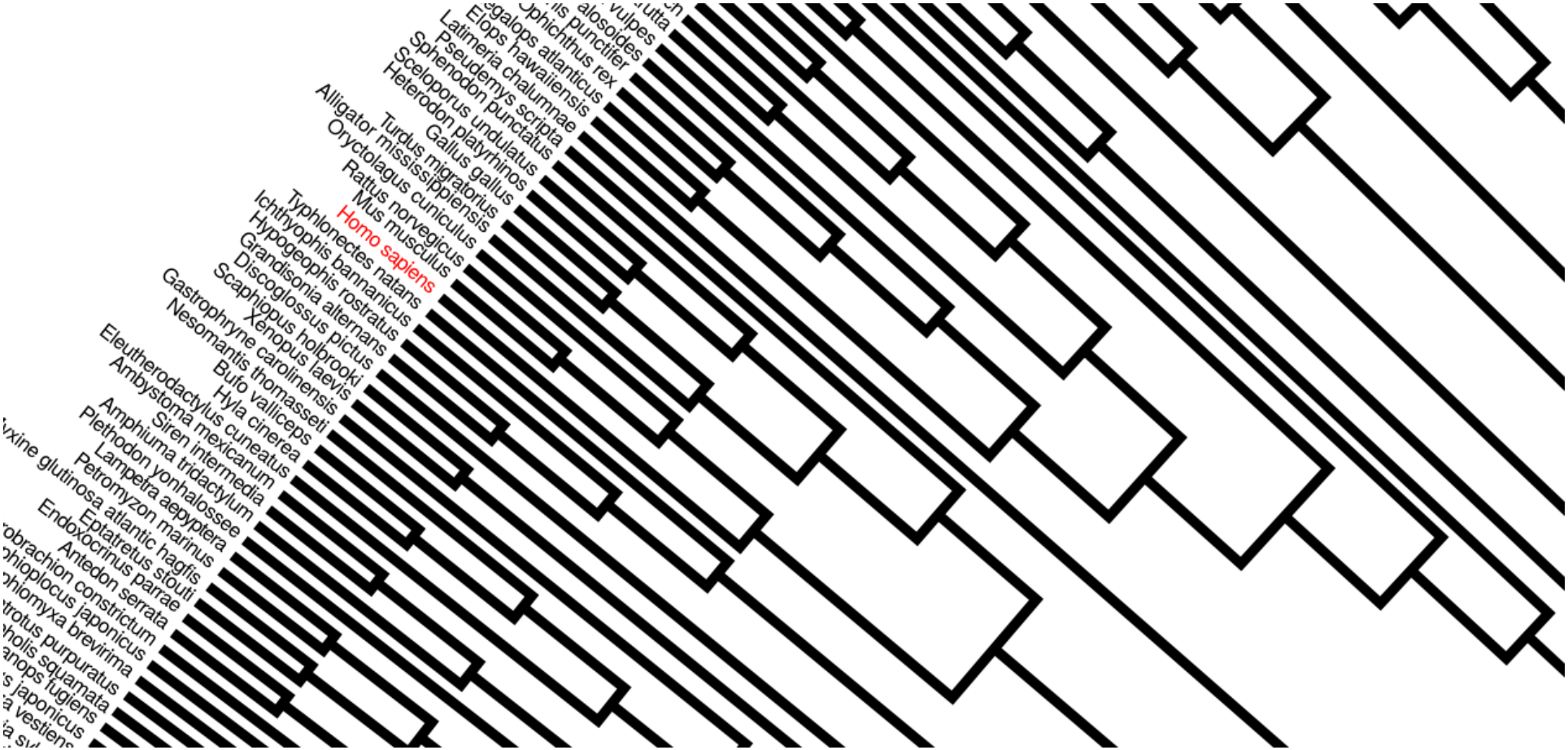
Extinction



Tree of life



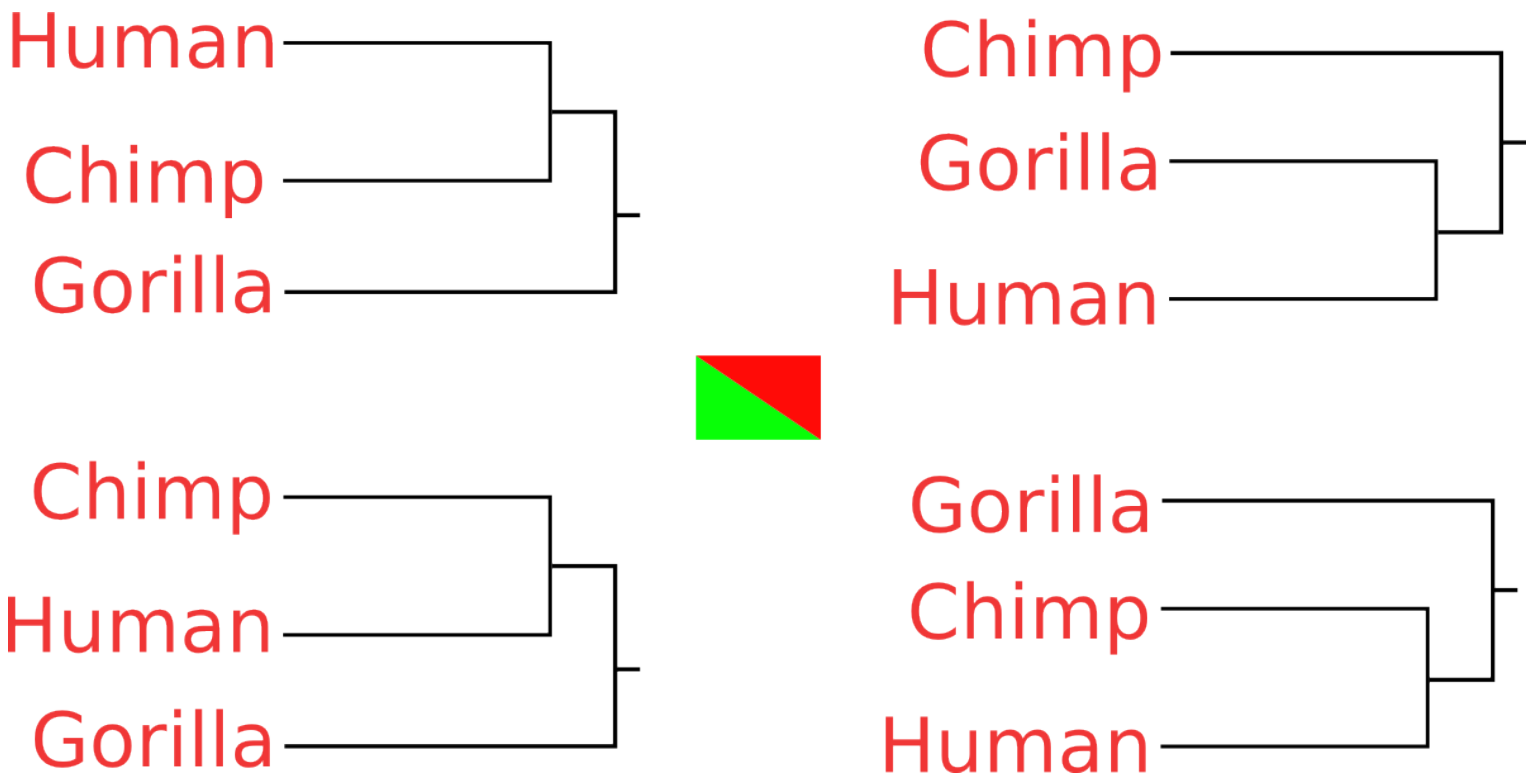
Tree of life



Mobiles

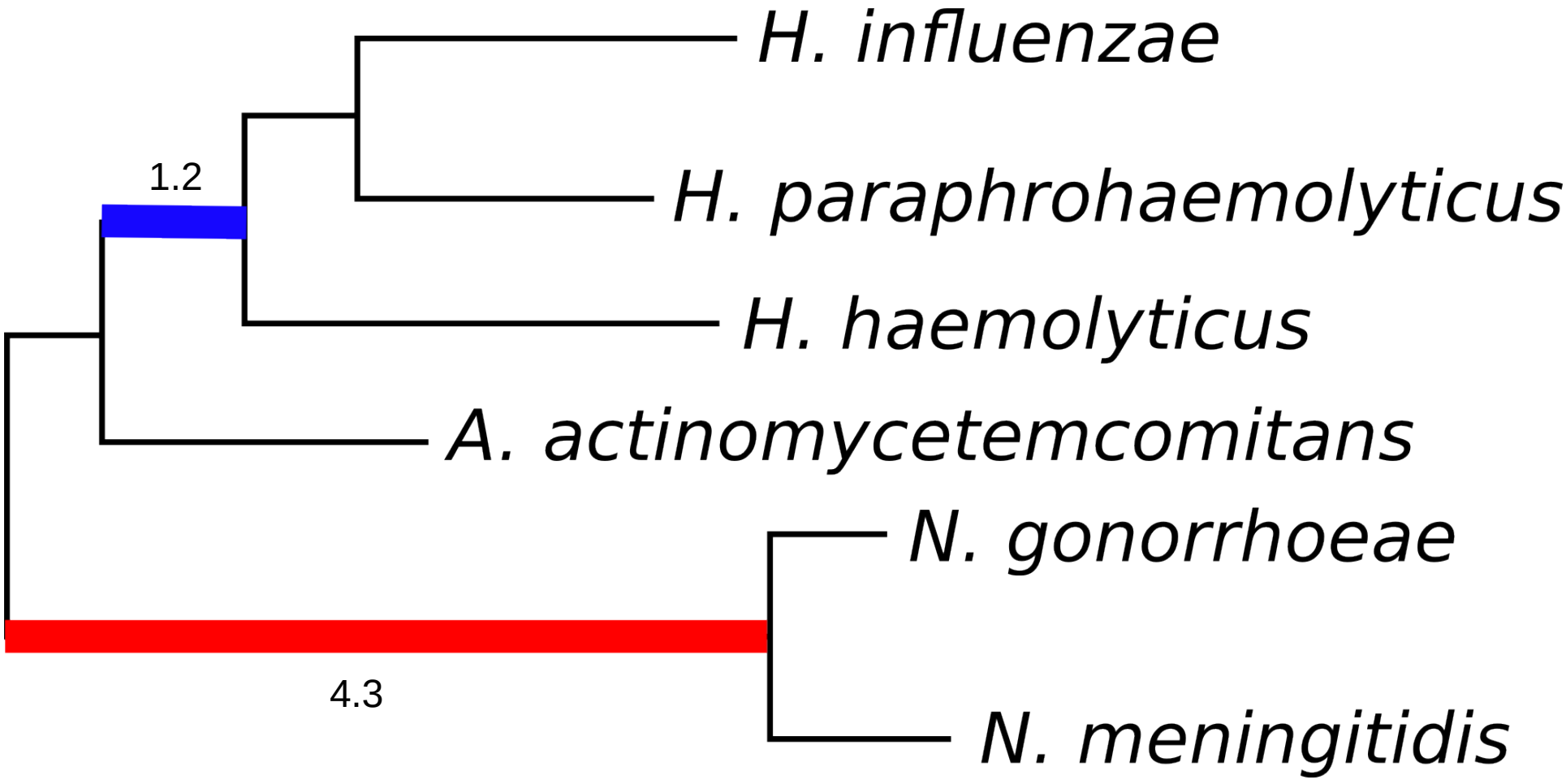
Mobiles

- There is no linear order
- Only branchings matter

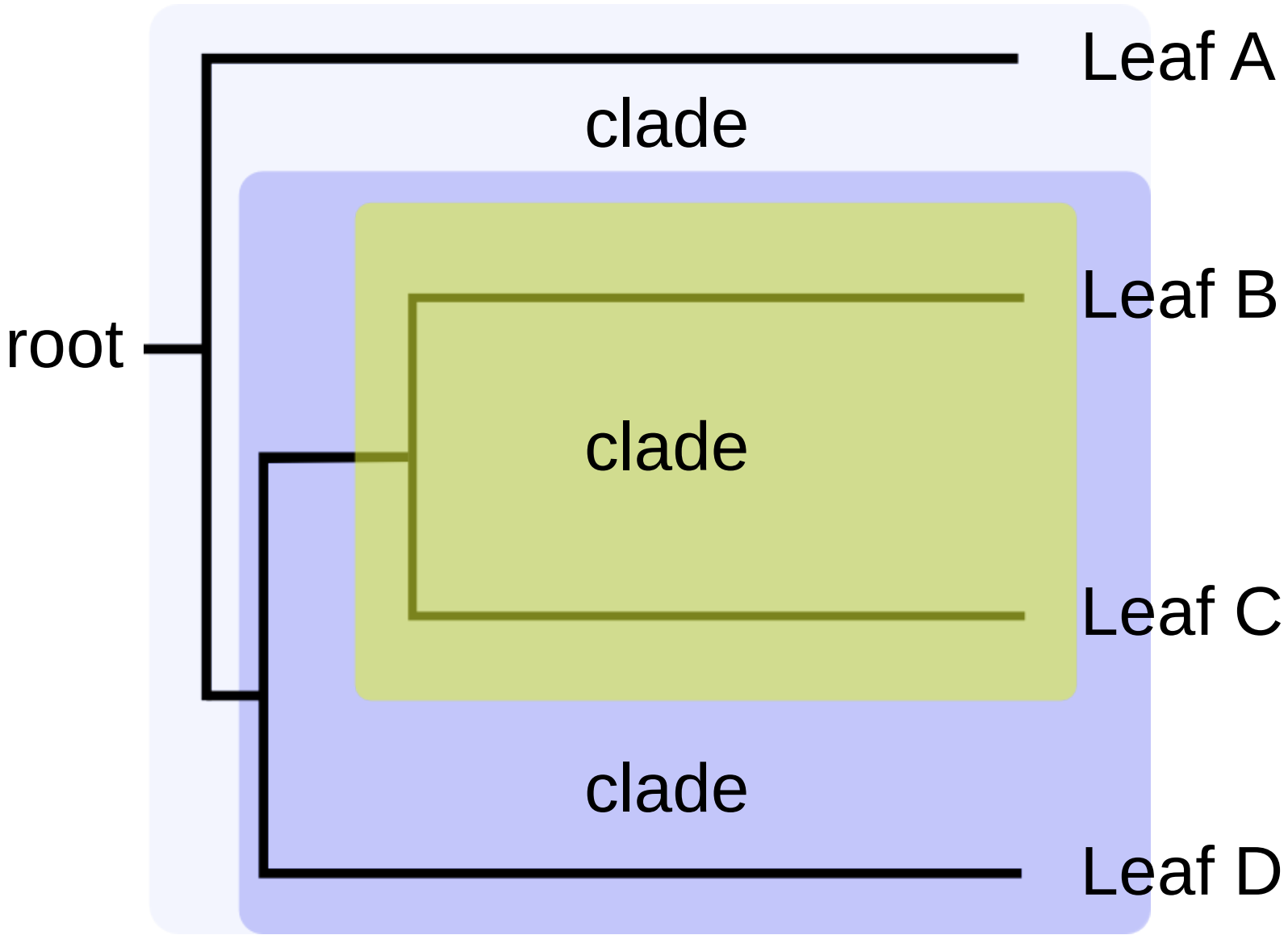


Branch length

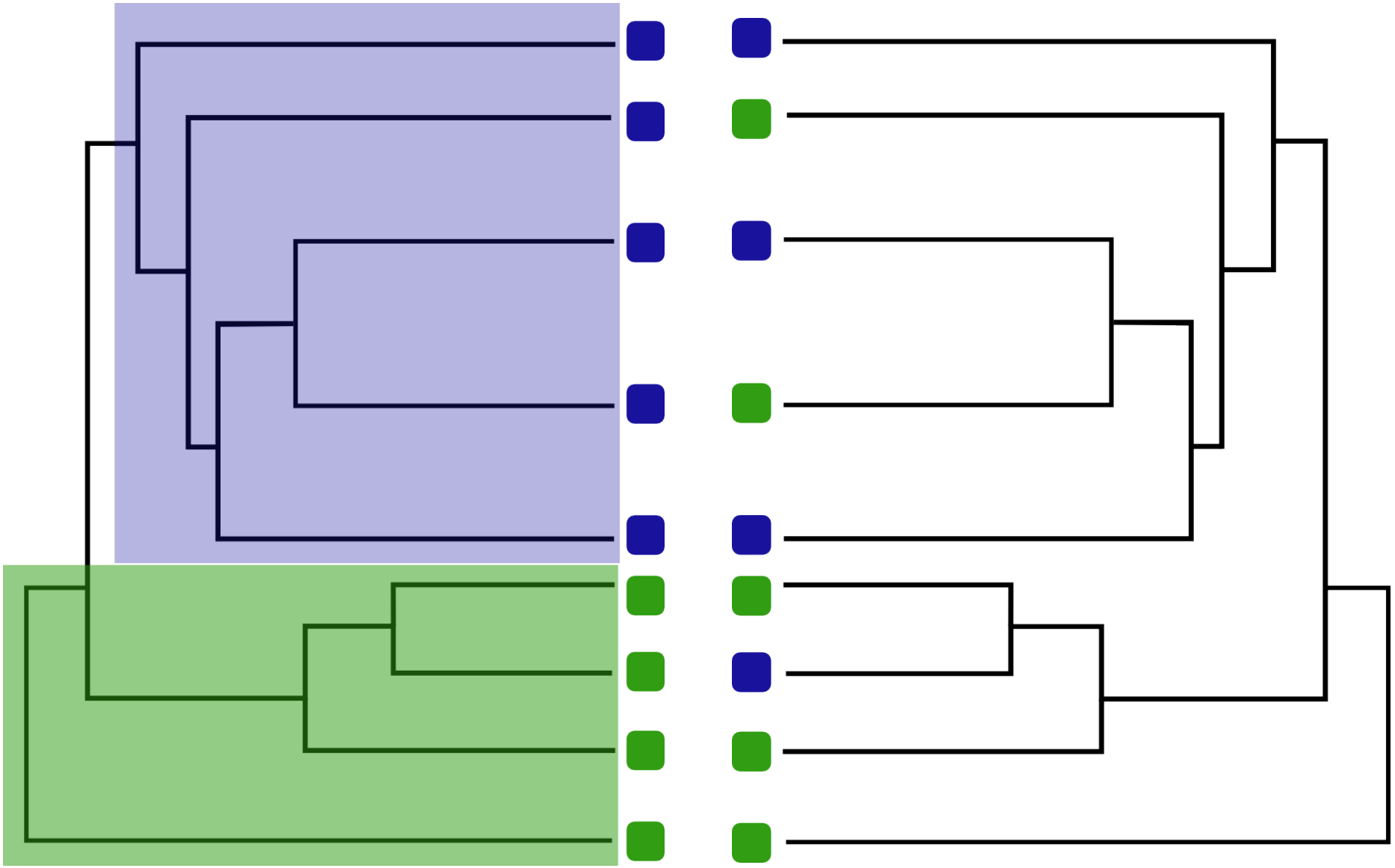
Distance = time * rate of substitutions



Jargon



Grouping



How many trees are there?

$$N_r = \prod_{k=2}^{T} (2k-3)$$

$$N_u = \prod_{k=3}^{T} (2k-5)$$

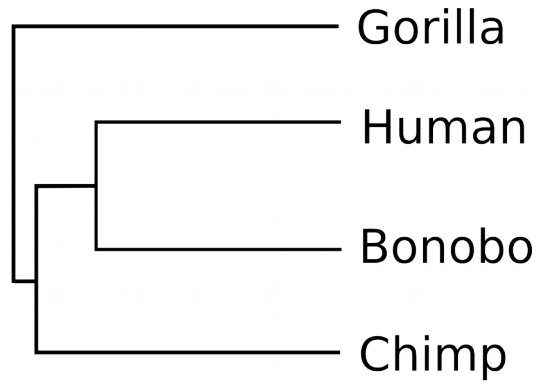
T	Unrooted	Rooted
10	2,027,025	3×10^{7}
15	7×10^{12}	2×10^{8}
20	2×10^{20}	8×10^{21}
100	2×10^{182}	2×10^{184}

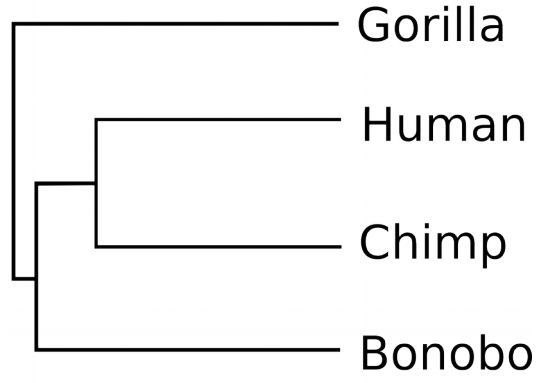
Ways to build a tree

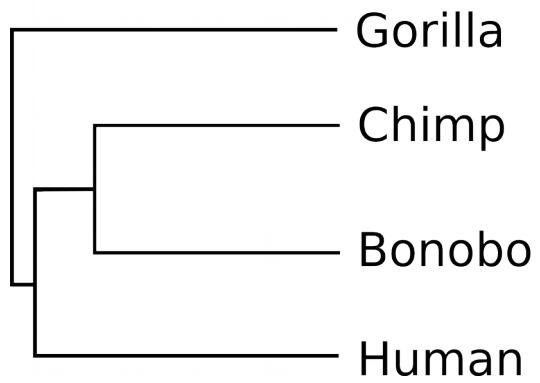
- Parsimony
  - Minimize number of mutations
  - Simple, but inaccurate
- Maximum Likelihood (most common)
  - Maximize probability of observation given model
  - Harder, but accurate
  - Does not provide confidence intervals
- Bayesian inference
  - Choose best model that maximizes probability given data
  - Hardest, very accurate
  - Explores all reasonable hypotheses

Parsimony

H: ACTGCATG
C: ACGGCCTC
G: CCGGCATG
B: ACTGCCTC

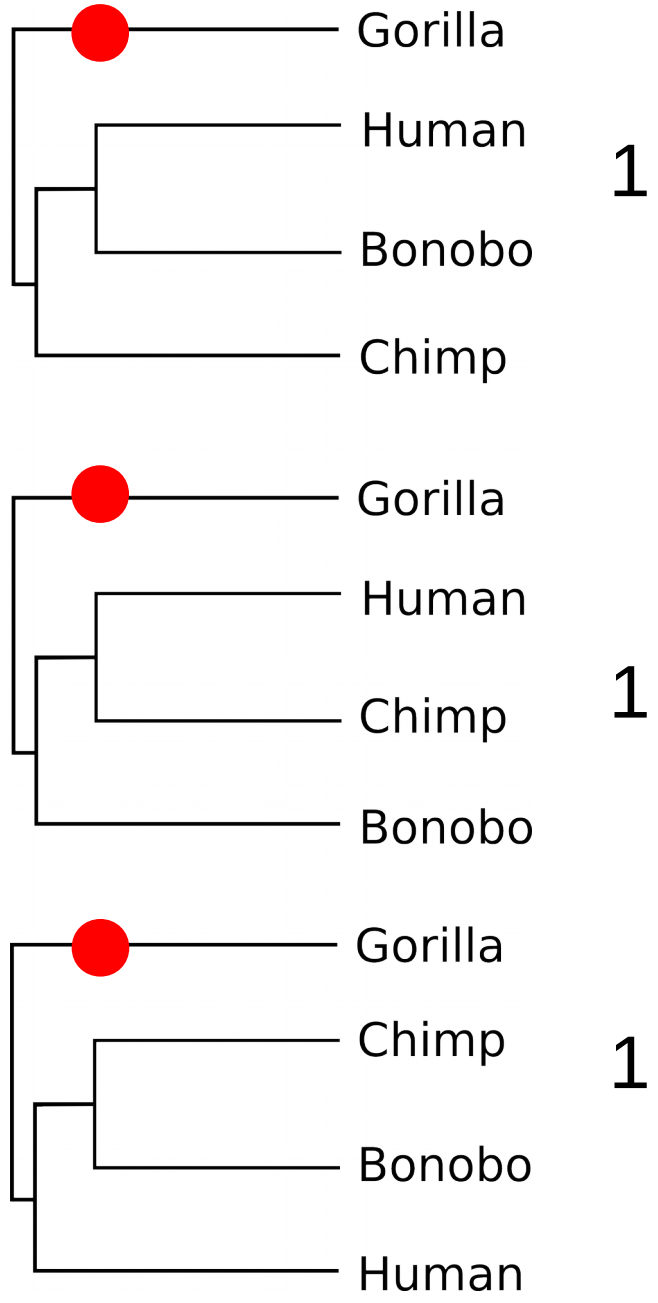






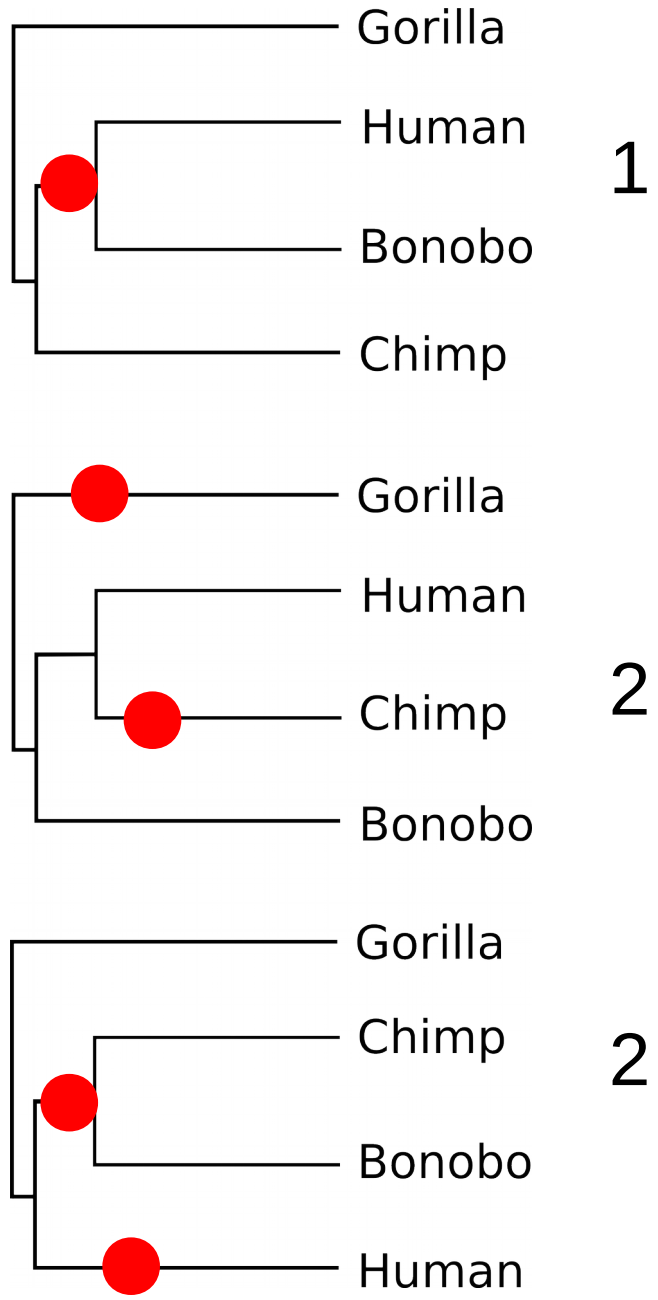
Parsimony

H: ACTGCATG
C: ACGGCCTC
G: CCGGCATG
B: ACTGCCTC



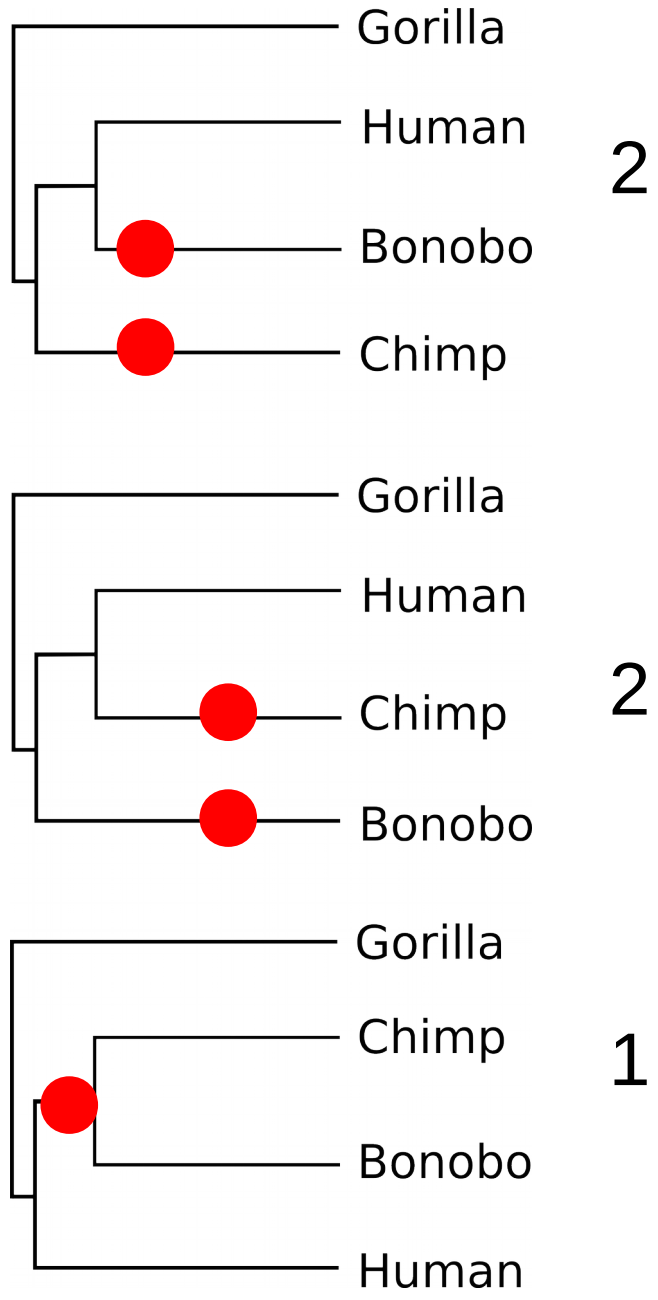
Parsimony

```
H: ACTGCATG
C: ACGGCCTC
G: CCGGCATG
B: ACTGCCTC
```



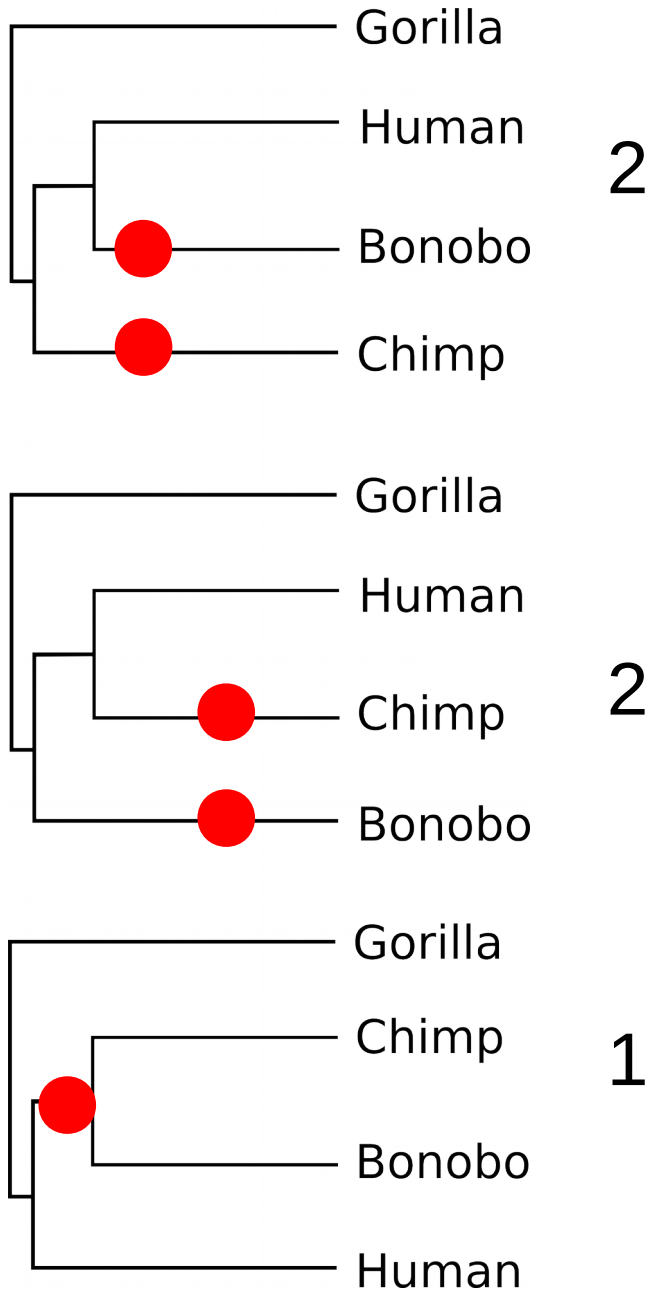
Parsimony

H: ACTGCATG
C: ACGGCCTC
G: CCGGCATG
B: ACTGCCTC



Parsimony

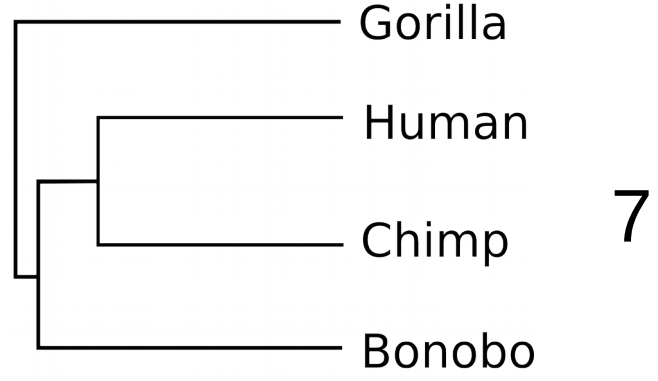
H: ACTGCATG
C: ACGGCCTC
G: CCGGCATG
B: ACTGCCTC



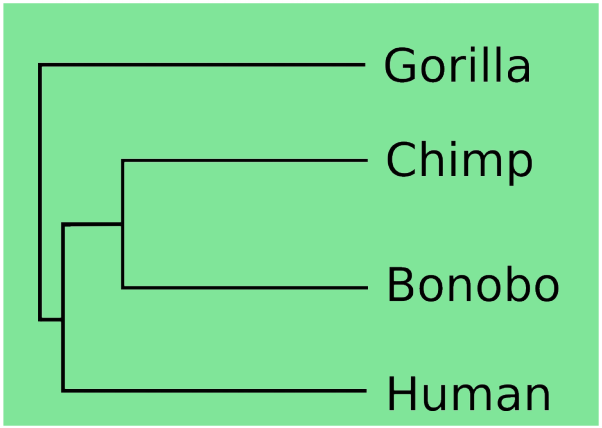
Parsimony

H: ACTGCATG
C: ACGGCCTC
G: CCGGCATG
B: ACTGCCTC

6



7



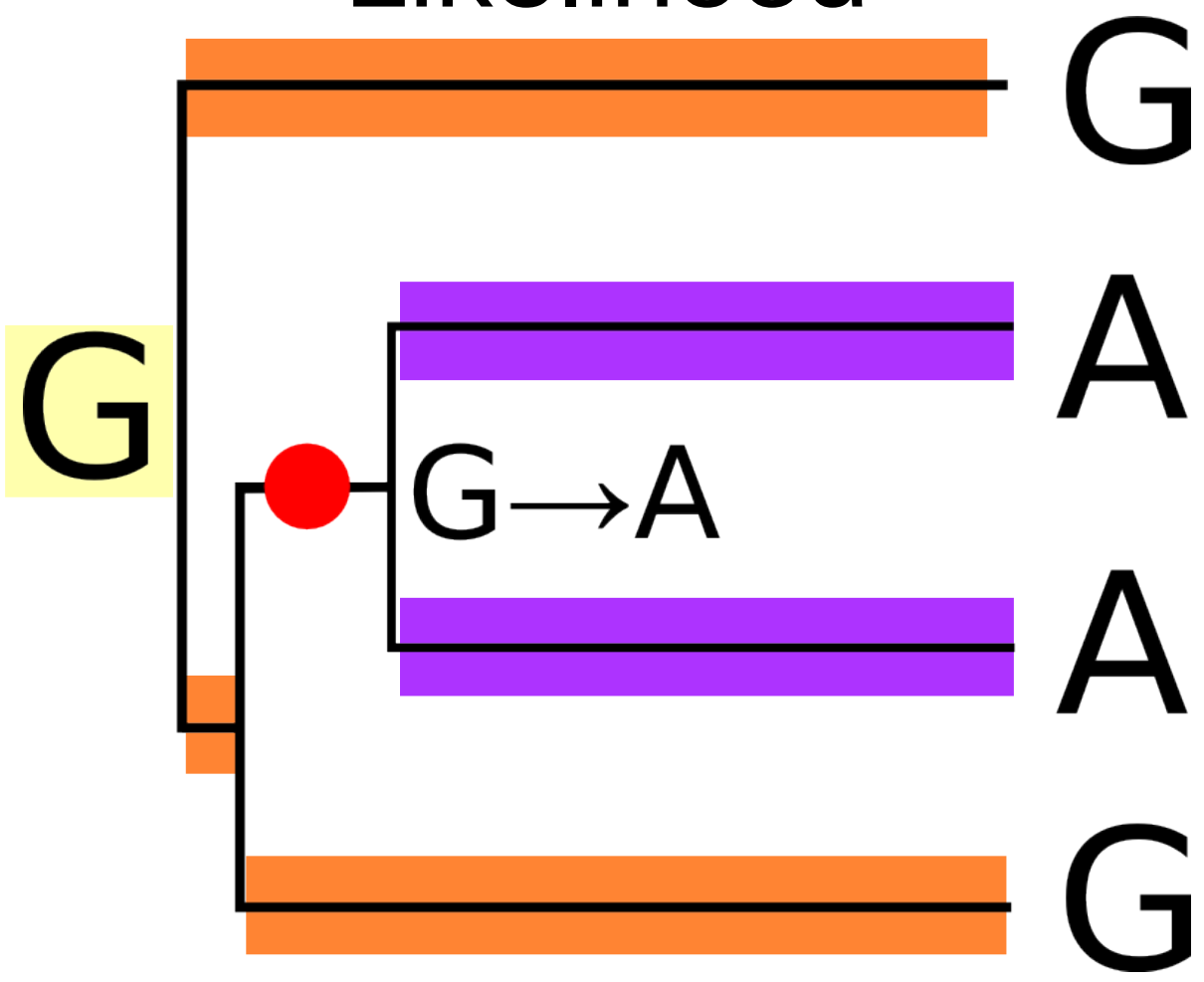
5

Likelihood

- Calculate the probability of each event according to a model
- Need to do this for all substitution histories
- Kimura 2 parameter model
  - Transition rate *a*
  - Transversion rate *b*

	A	C	T	G
A		b	a	b
C	b		b	a
T	a	b		b
G	b	a	b	

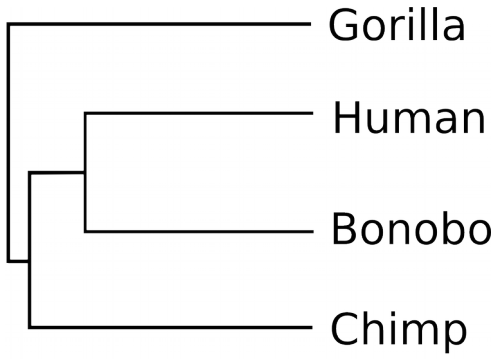
Likelihood

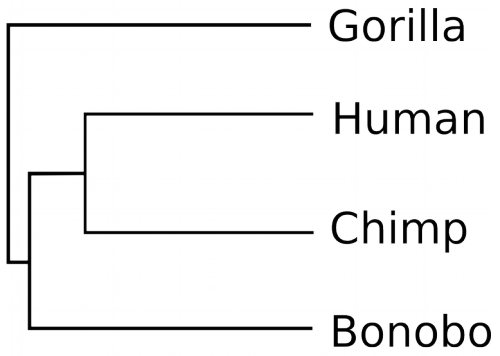


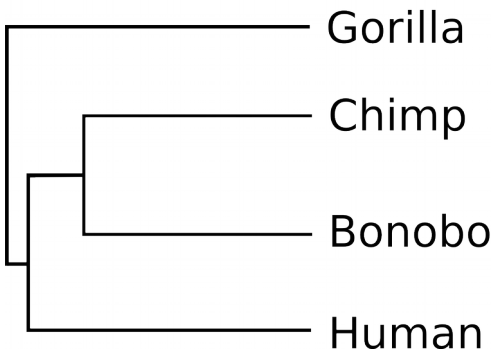
$$P(T) = P(G) \times P(G \to G) \times P(G \to G) \times P(G \to G) \times P(G \to A) \times P(A \to A) \times P(A \to A)$$

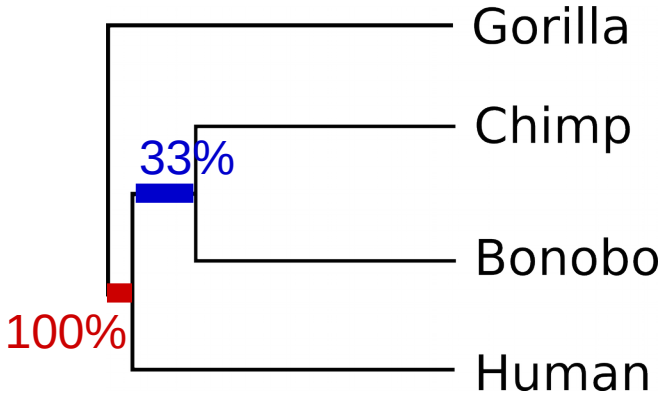
Bootstrap

```
H: ACGCGTAT
C: ACGCCGCT
G: CCGCGGAT
B: ACGCCTCT

H: ATGGCTCA
C: AGCGCTCC
G: CGGGCTCA
B: ATCGCTCC

H: ACTGCATG
C: ACGGCCTC
G: CCGGCATG
B: ACTGCCTC
```









References

- March of progress http://en.wikipedia.org/wiki/March_of_Progress
- Genealogy of Ludwig Herzog von Wurttemberg http://commons.wikimedia.org/wiki/File:900-158_Ahnentafel_Herzog_Ludwig.jpg#mediaviewer/File:900-158_Ahnentafel_Herzog_Ludwig.jpg
- Mobile ad: http://weedecor.com/wp-content/uploads/2013/05/safari_baby_mobile.jpg
- Phylogenetic trees on wikipedia: http://en.wikipedia.org/wiki/Phylogenetic_tree
- Fredrick Leliaert on species trees: http://frederikleliaert.wordpress.com/green-algae/dna-based-species-delimitation-in-algae/
- MAFFT: http://mbe.oxfordjournals.org/content/30/4/772
- Phylogeny.fr: http://www.ncbi.nlm.nih.gov/pubmed/18424797