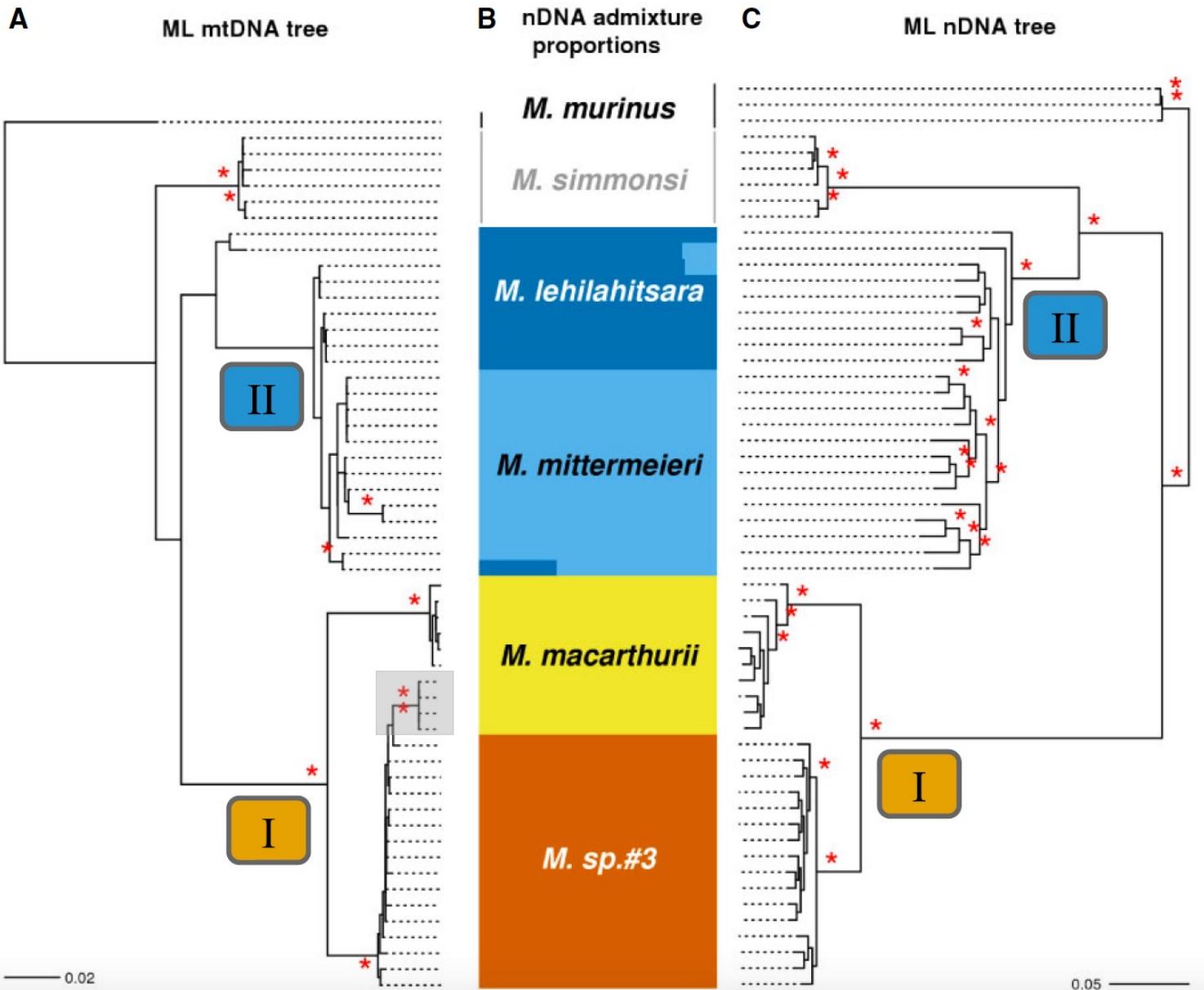


Tree Support and Visualization

George P. Tiley
University of Antananarivo
DBEV Phylogenomics Workshop
9 March 2022

How much do you trust a tree?



Learning Goals

Newick: The tree file format

Bootstrapping

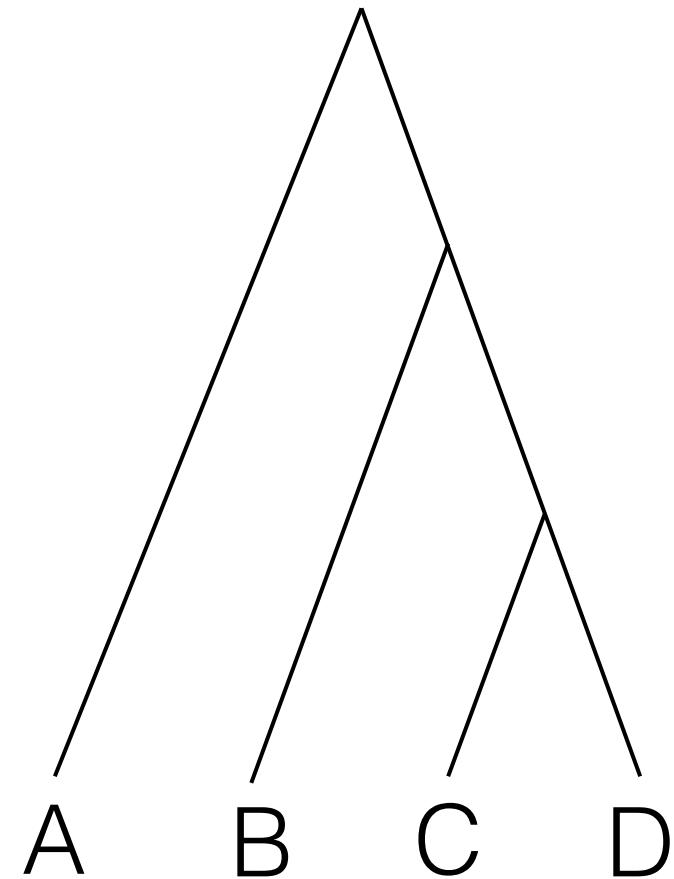
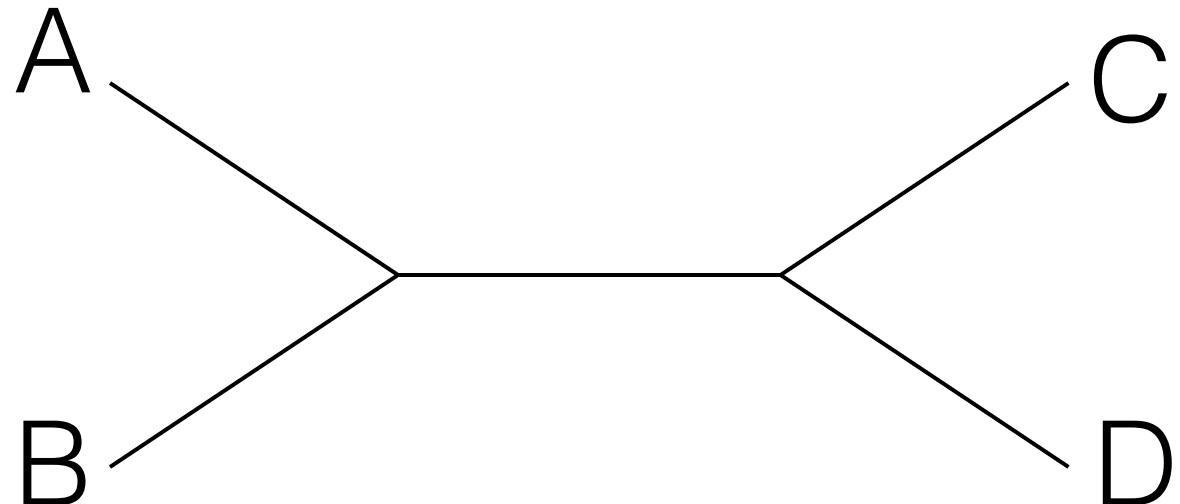
Other support values

Visualization examples

Visualization tools

Newick: The tree file format

These are trees we can interpret



Newick: The tree file format

This is harder to interpret

(A,(B,(C,D)));

Newick: The tree file format

This is harder to interpret

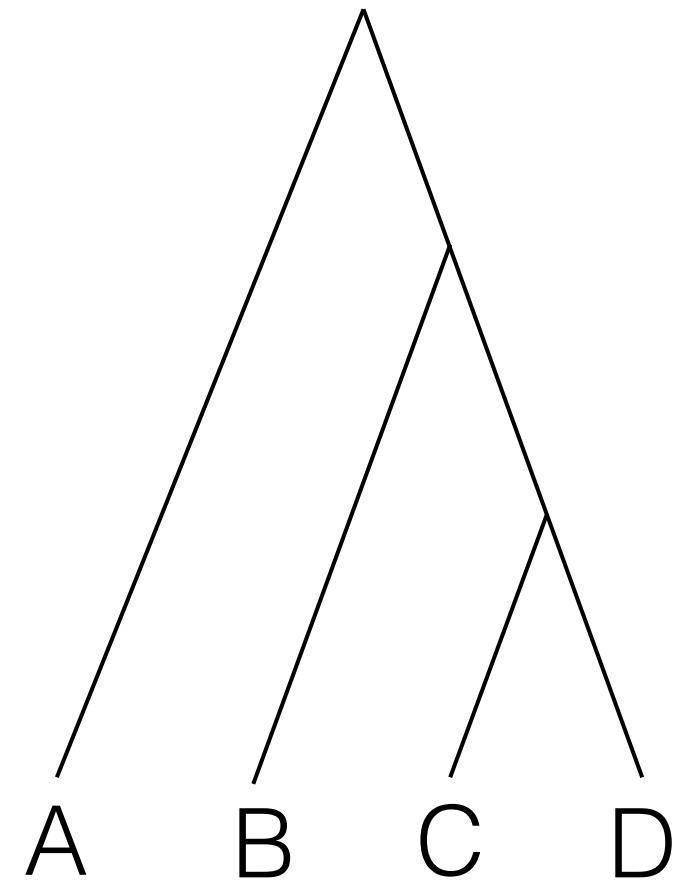
```
(Adi001,((Adi002,Adi003)1:0.12289383895378848,((Age001,(Smi165,(Bce020,Pcr070)1:0.7673411039793523)1:2.2308028194195177)0.89:0.07070841023727185,((Aru001,Aru127)1:1.4910934193376955,((Asu001,(Aga001,Aga002)1:0.9135013297865763)1:0.9912286811765139,(Aza135,((Ape009,Ape001)1:0.5420707776252044,((Ama006,Ama018)1:0.44681098284646115,Aza037)1:0.1744454780692822)1:0.1614514744553304)1:0.9752582079622695)0.72:0.051902544503560834)1:0.34371832066824454)1:2.0008453616154918):0.0);
```

But what we get from analyses

Newick: The tree file format

How to read a newick tree

$(A,(B,(C,D))); =$



Newick: The tree file format



How to read a newick tree

(A,(B,(C,D))); =

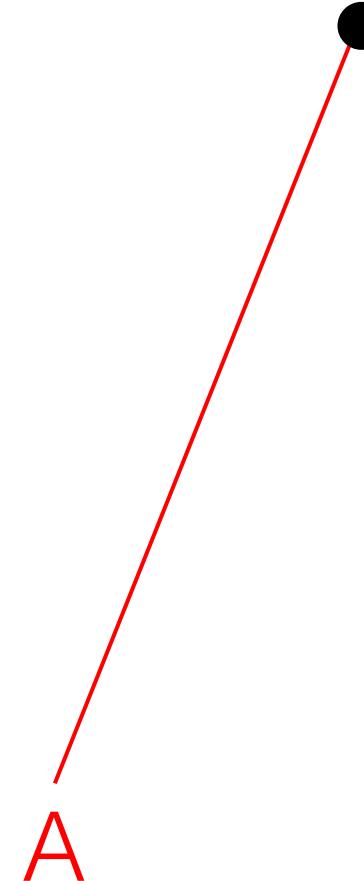
“(” means make a new node

Newick: The tree file format

How to read a newick tree

(A,(B,(C,D))); =

Then go left
This branch ends with species A

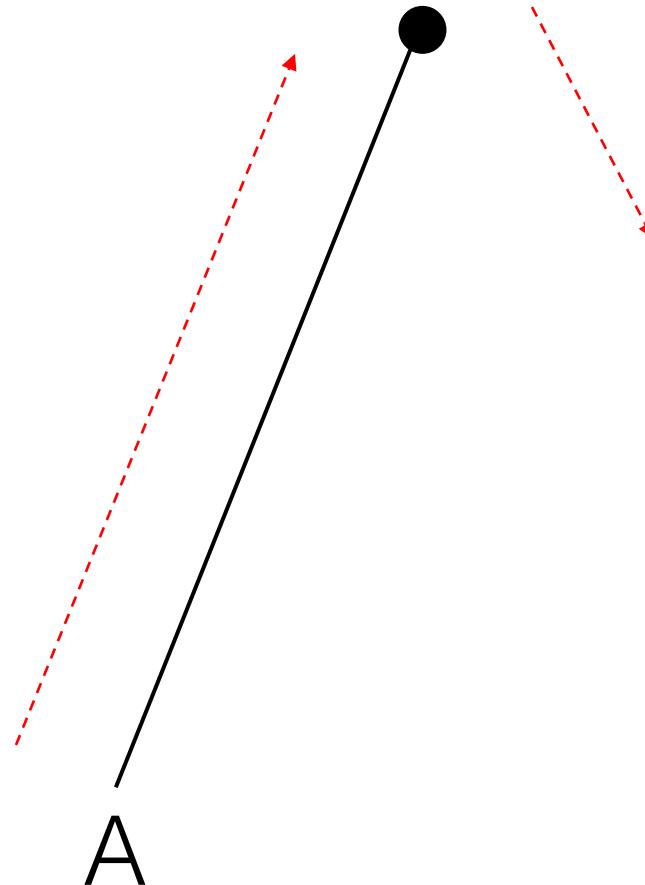


Newick: The tree file format

How to read a newick tree

(A,
,(B,(C,D))); =

“,” means go back to the node
next you go to the right

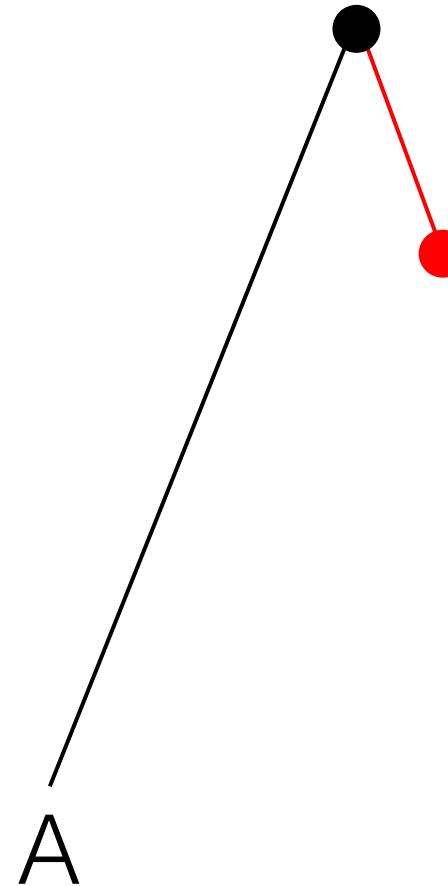


Newick: The tree file format

How to read a newick tree

(A,(B,(C,D))); =

“(”
start a new node

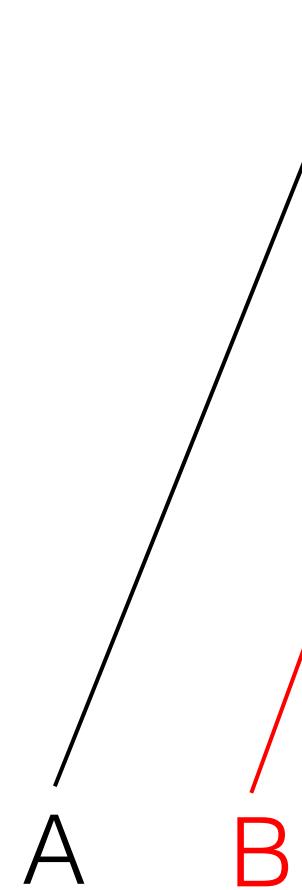


Newick: The tree file format

How to read a newick tree

(A, (B, (C, D))); =

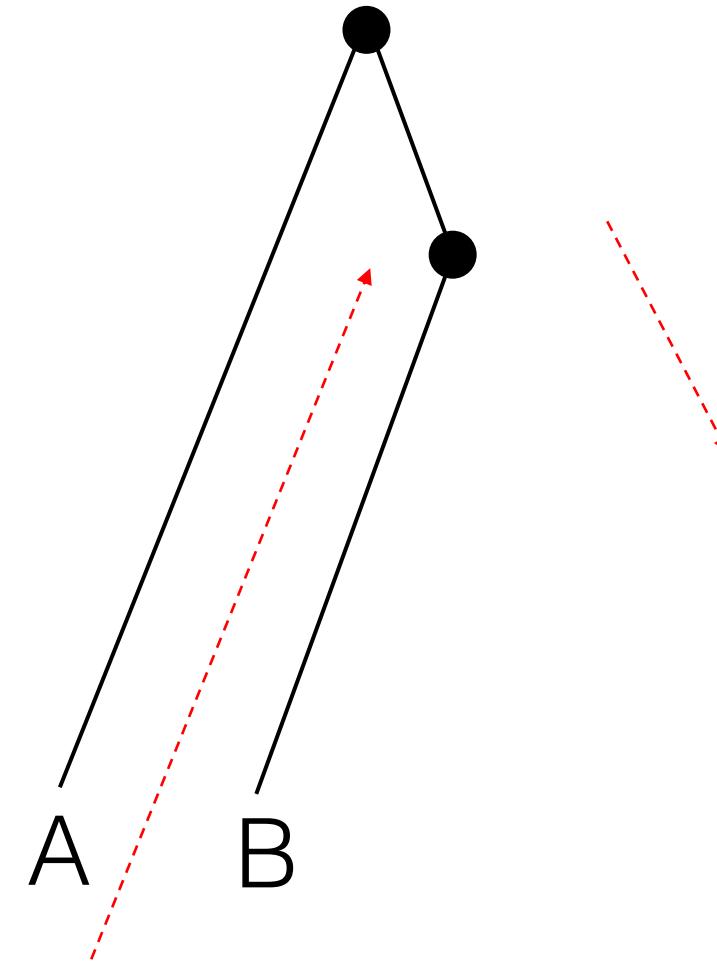
go left and end with “B”



Newick: The tree file format

How to read a newick tree

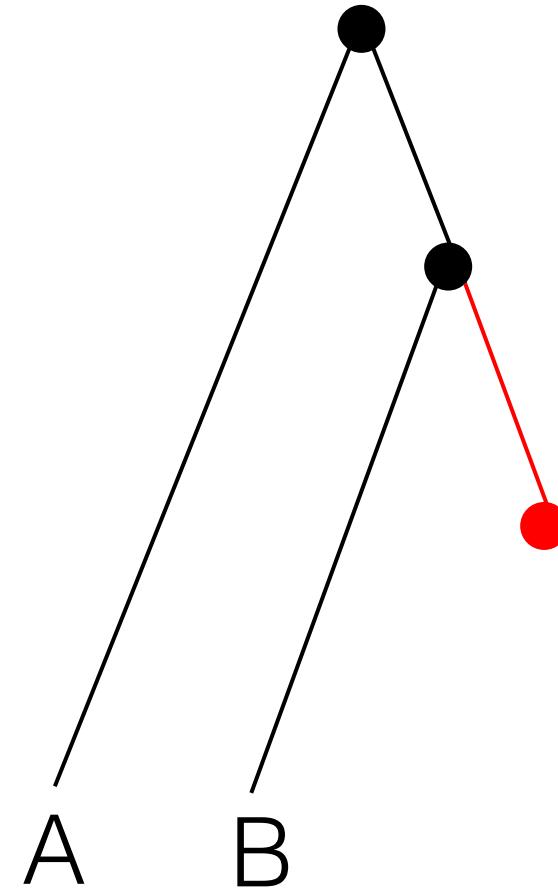
$(A,(B,(C,D))); =$



Newick: The tree file format

How to read a newick tree

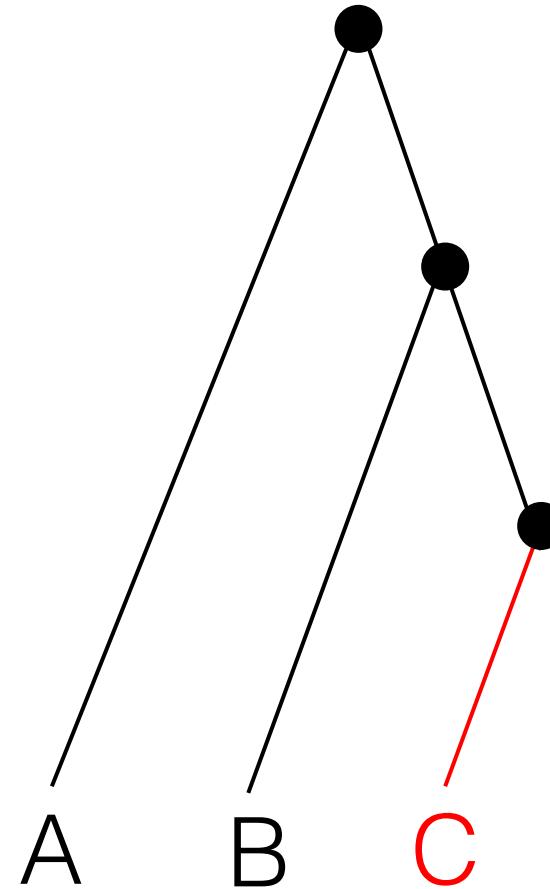
$(A,(B,(C,D))); =$



Newick: The tree file format

How to read a newick tree

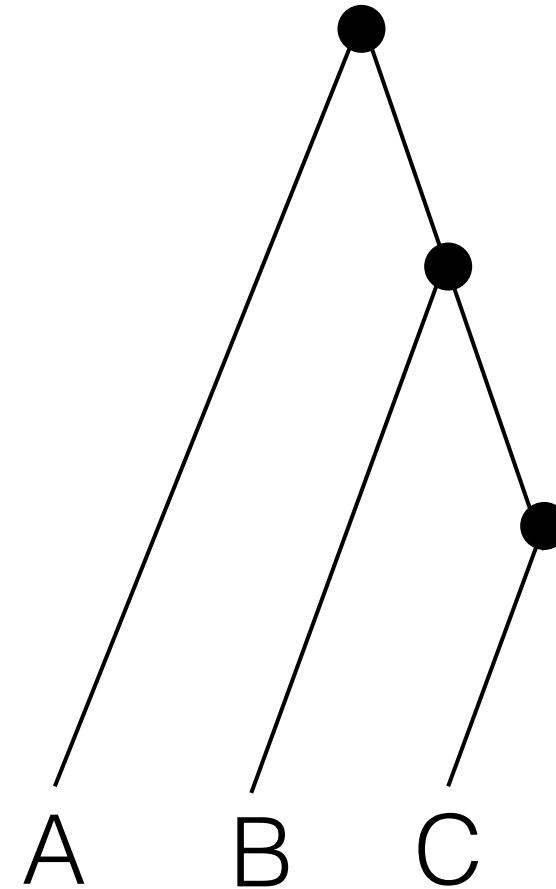
(A,(B,(C,D))); =



Newick: The tree file format

How to read a newick tree

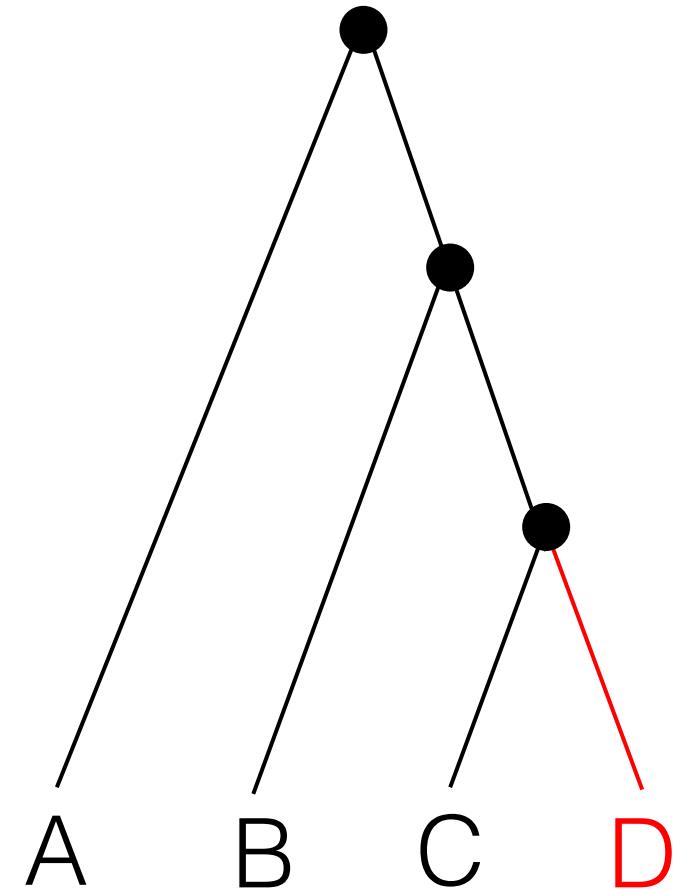
(A,(B,(C,D))); =



Newick: The tree file format

How to read a newick tree

(A,(B,(C,D))); =

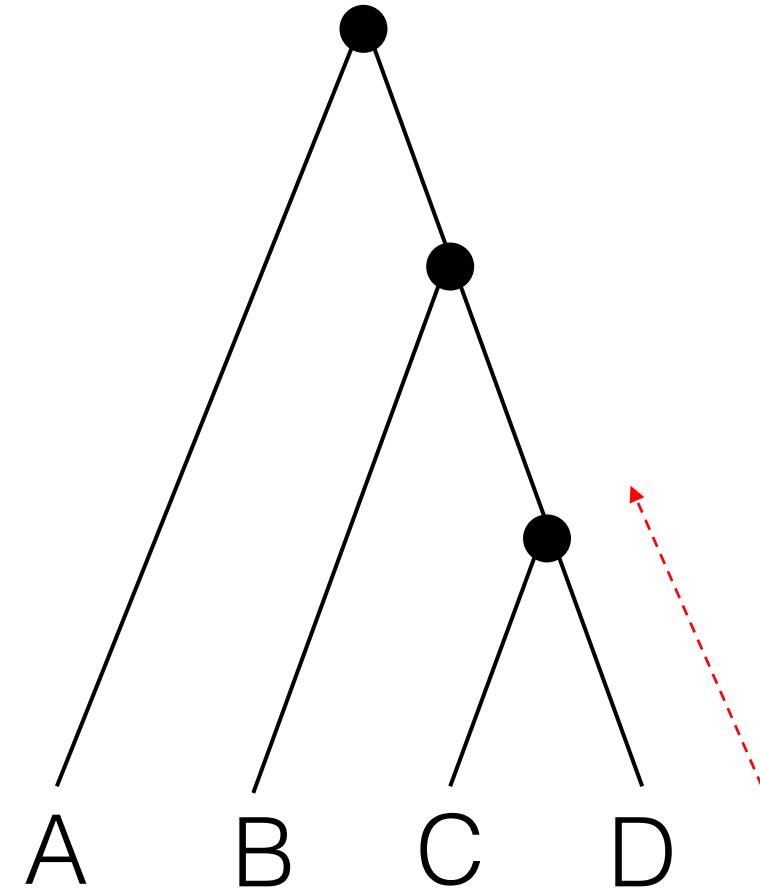


Newick: The tree file format

How to read a newick tree

(A,(B,(C,D))); =

go back and close the node

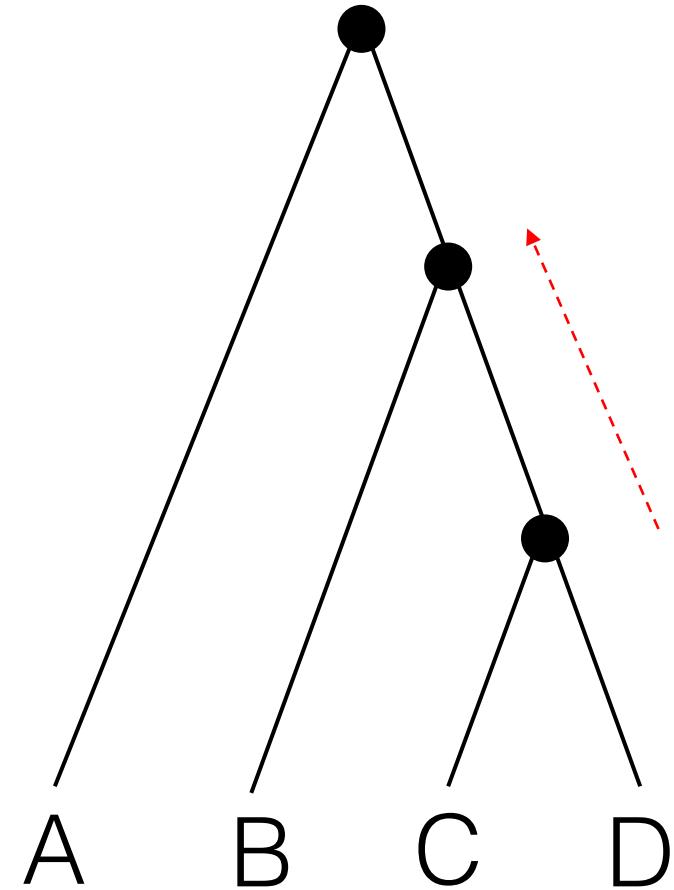


Newick: The tree file format

How to read a newick tree

(A,(B,(C,D))); =

go back and close the node

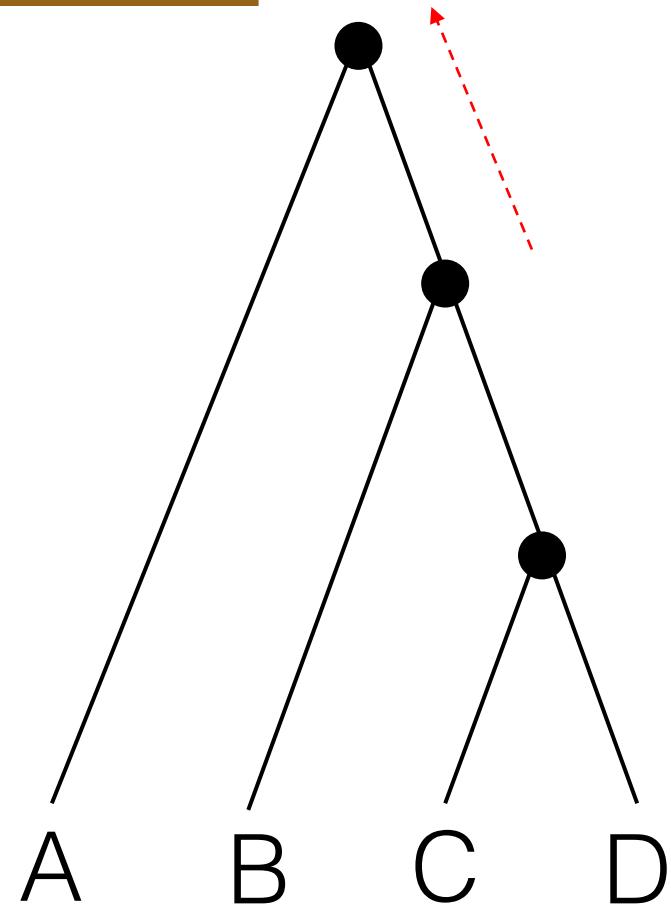


Newick: The tree file format

How to read a newick tree

(A,(B,(C,D))); =

go back and close the node

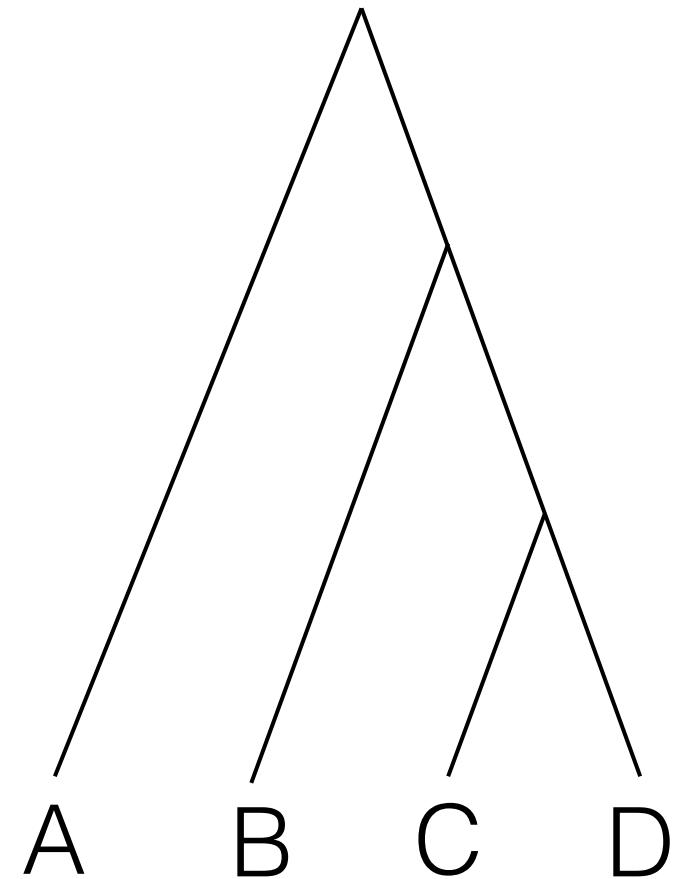


Newick: The tree file format

How to read a newick tree

(A,(B,(C,D))); =

the tree is done!



Newick: The tree file format

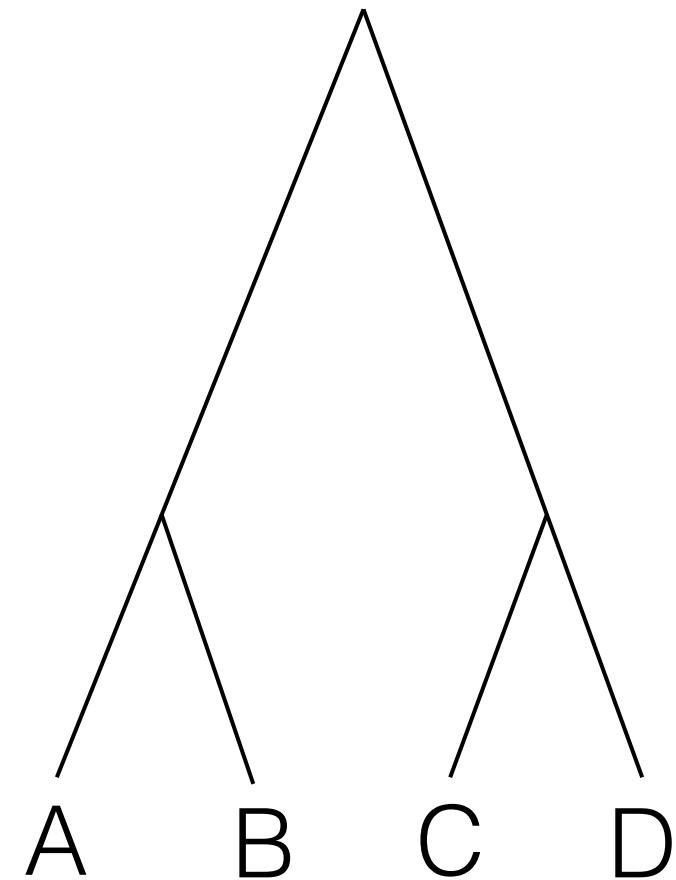
Time to practice

((A,B),(C,D)); = ?

Newick: The tree file format

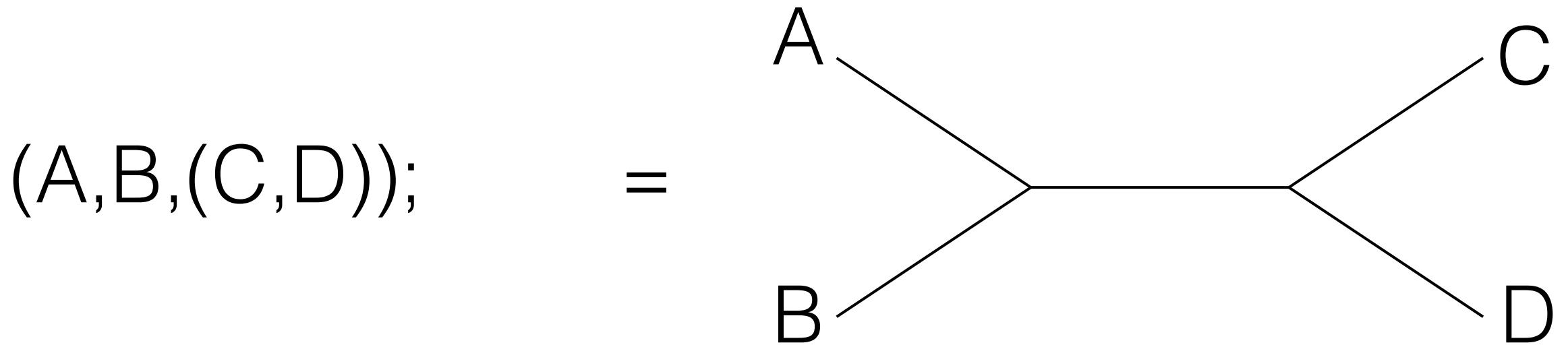
Time to practice

$((A,B),(C,D));$ =



Newick: The tree file format

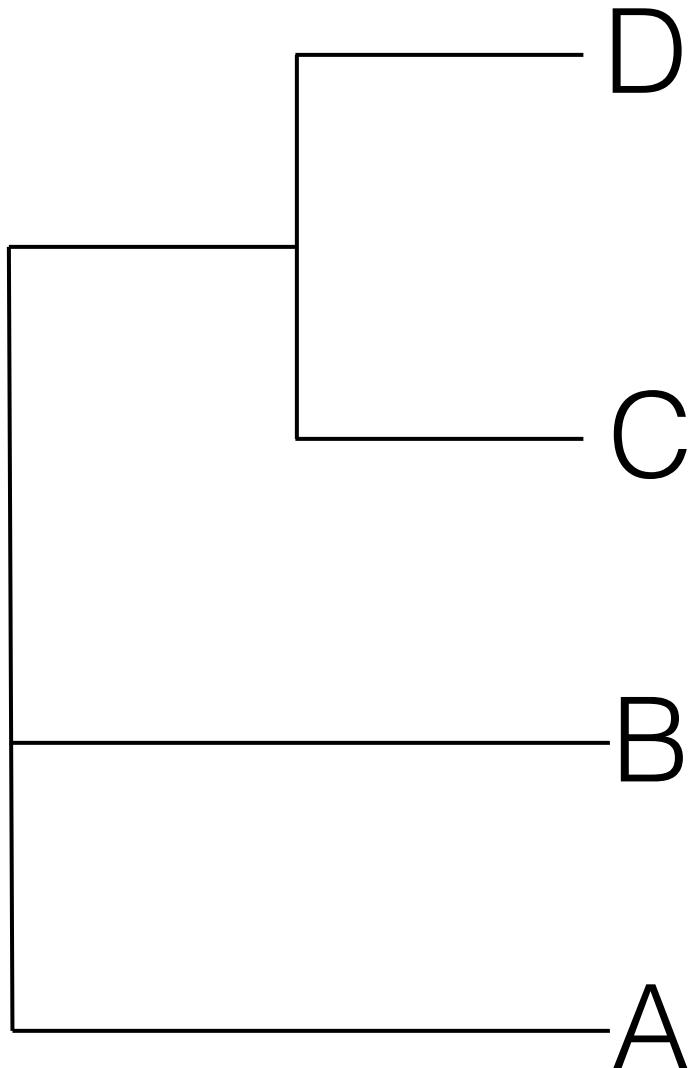
An unrooted example



Newick: The tree file format

An unrooted example

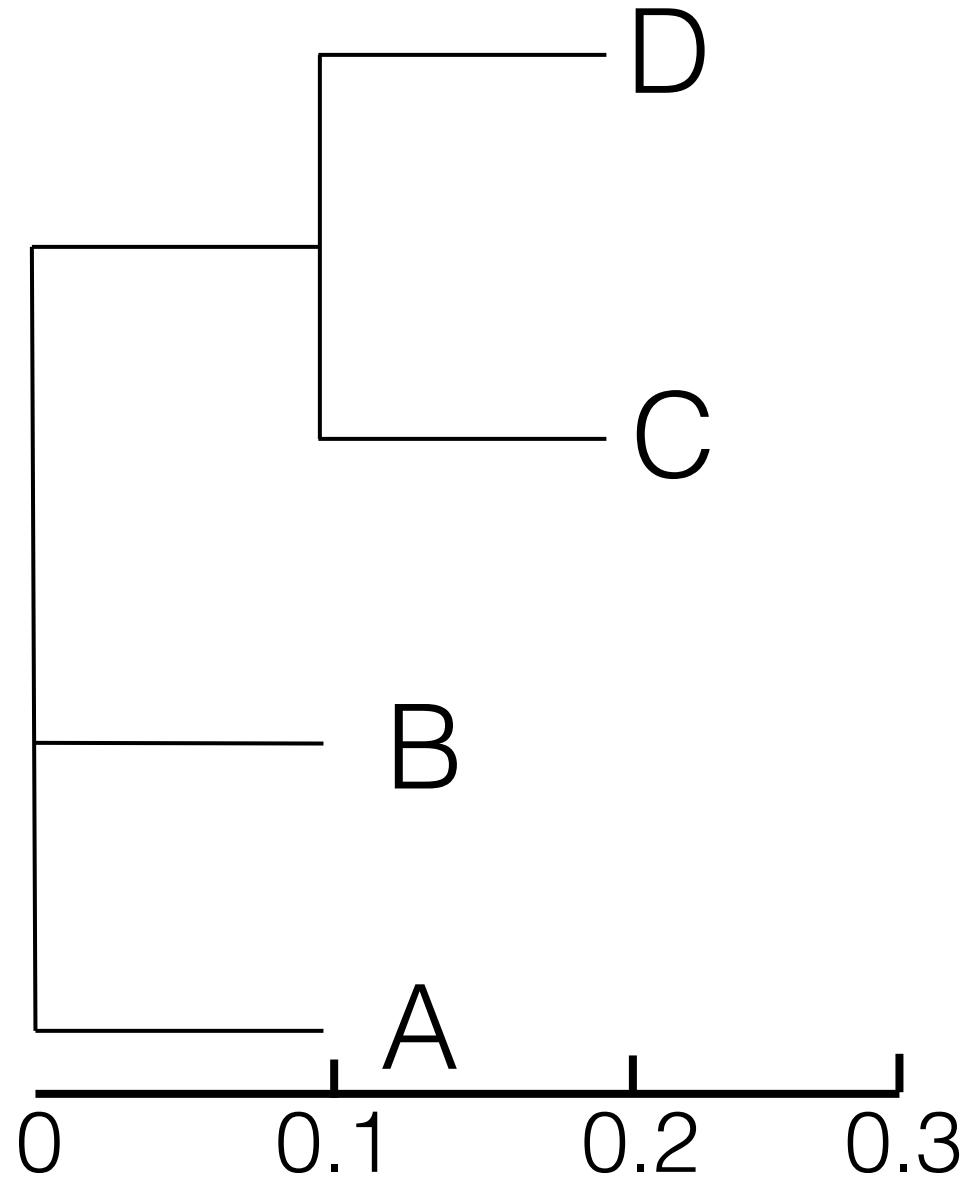
$(A,B,(C,D)); =$



Newick: The tree file format

An unrooted example with
branch lengths

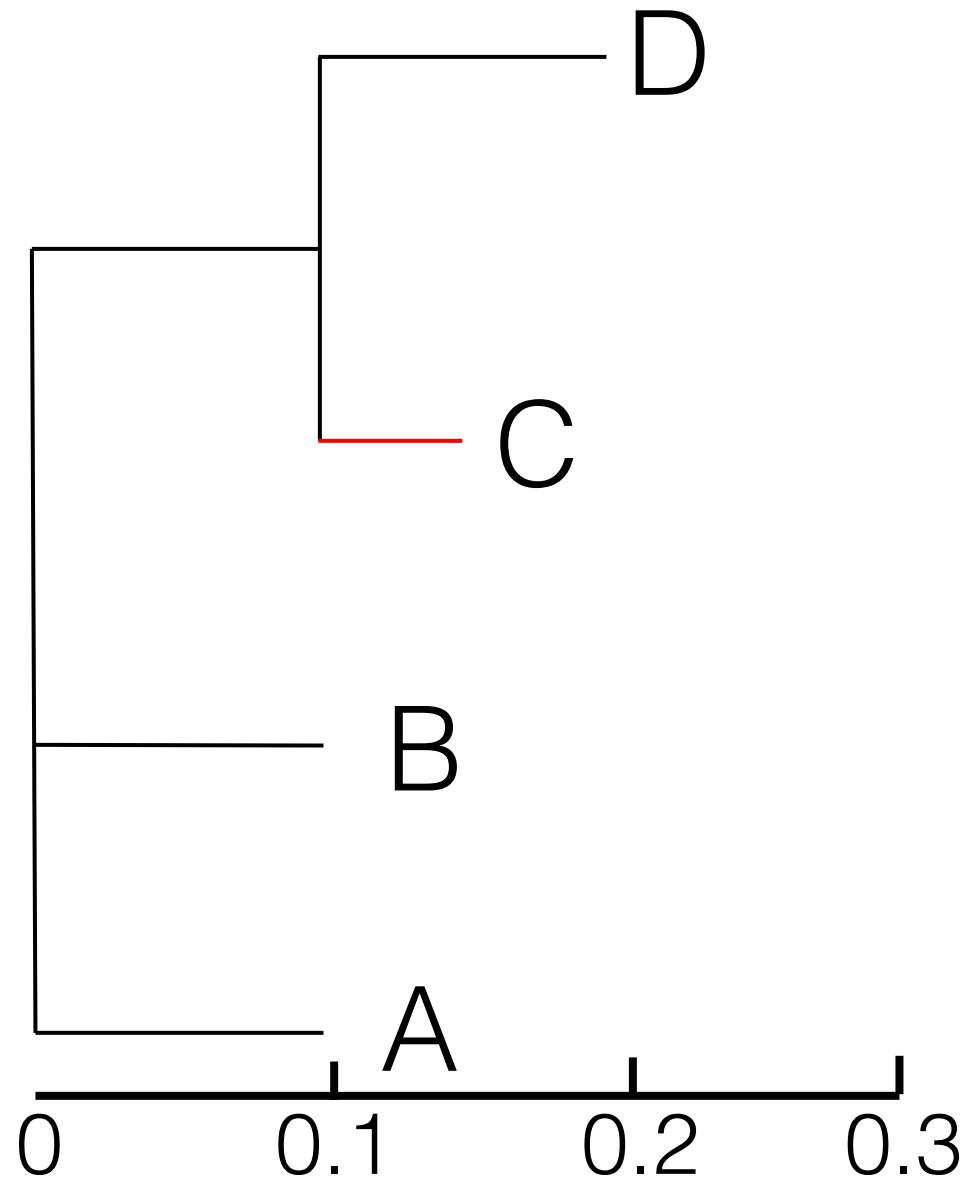
(A:0.1,B:0.1,(C:0.1,D:0.1):0.1); =



Newick: The tree file format

An unrooted example with
branch lengths

(A:0.1,B:0.1,(C:0.05,D:0.1):0.1); =

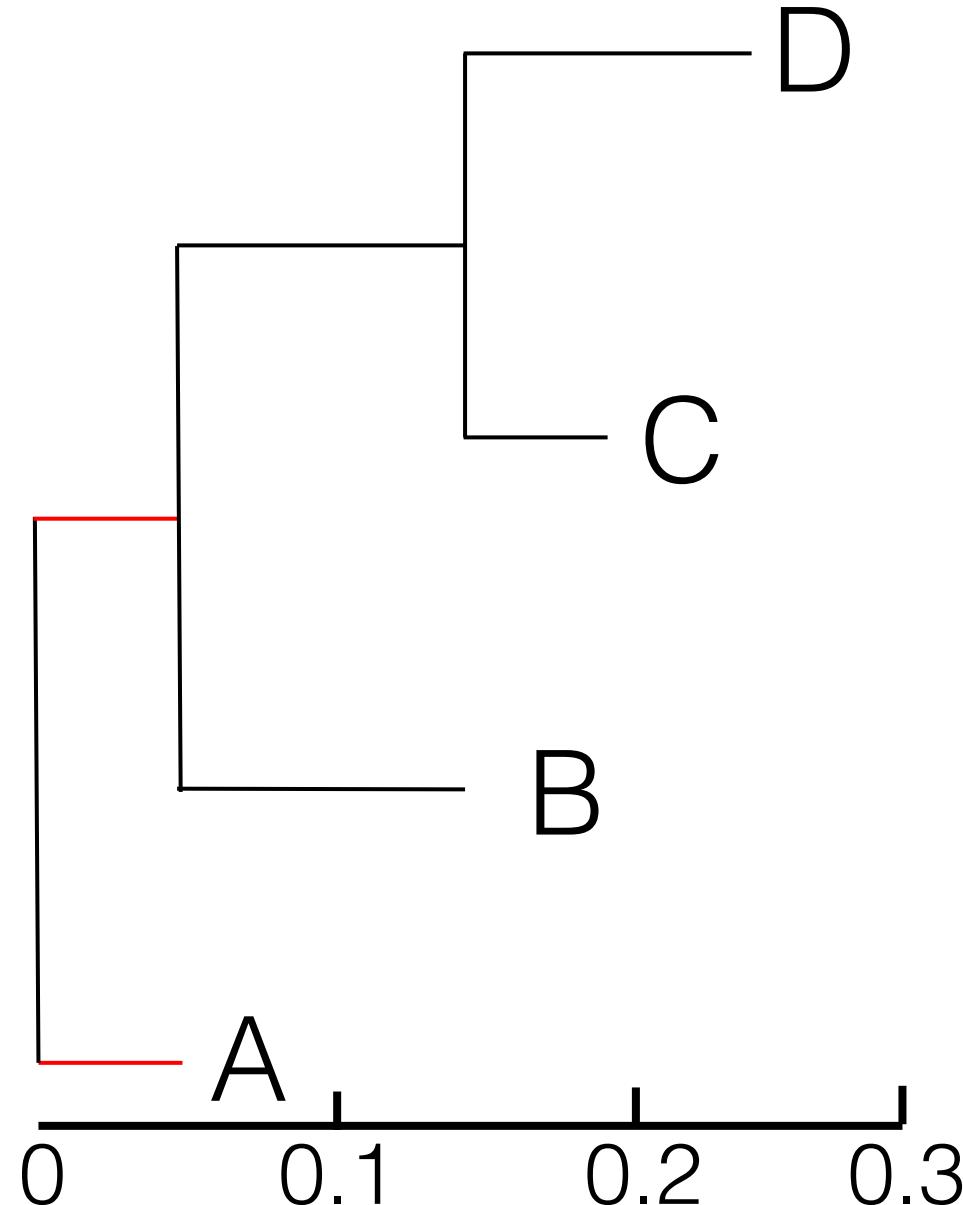


Newick: The tree file format

When we root the tree

(A:0.05,(B:0.1,(C:0.05,D:0.1):0.1):0.05); =

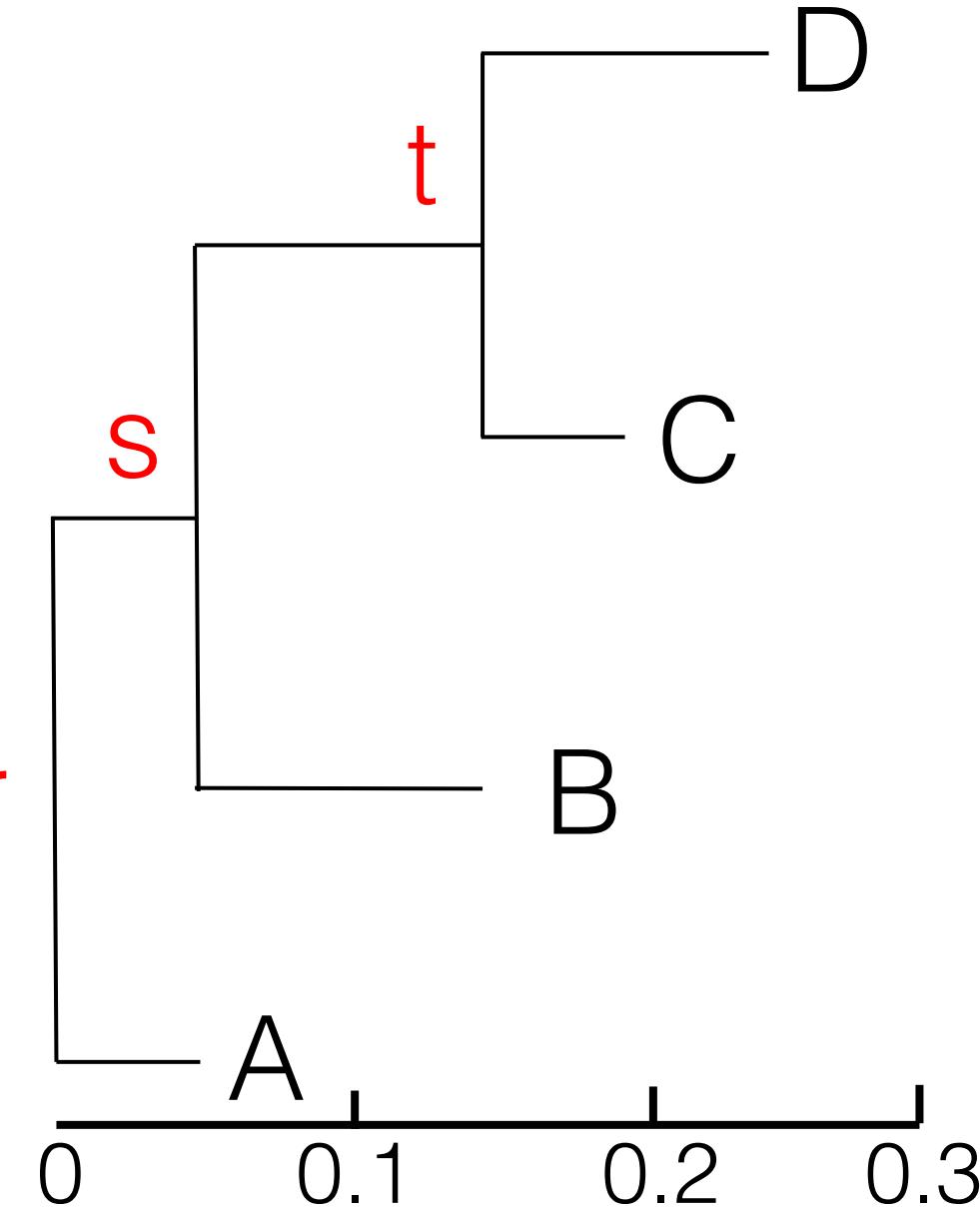
Notice that the total tree length stays the same



Newick: The tree file format

Nodes can have labels too

(A:0.05,(B:0.1,(C:0.05,D:0.1)t:0.1)s:0.05) r; =

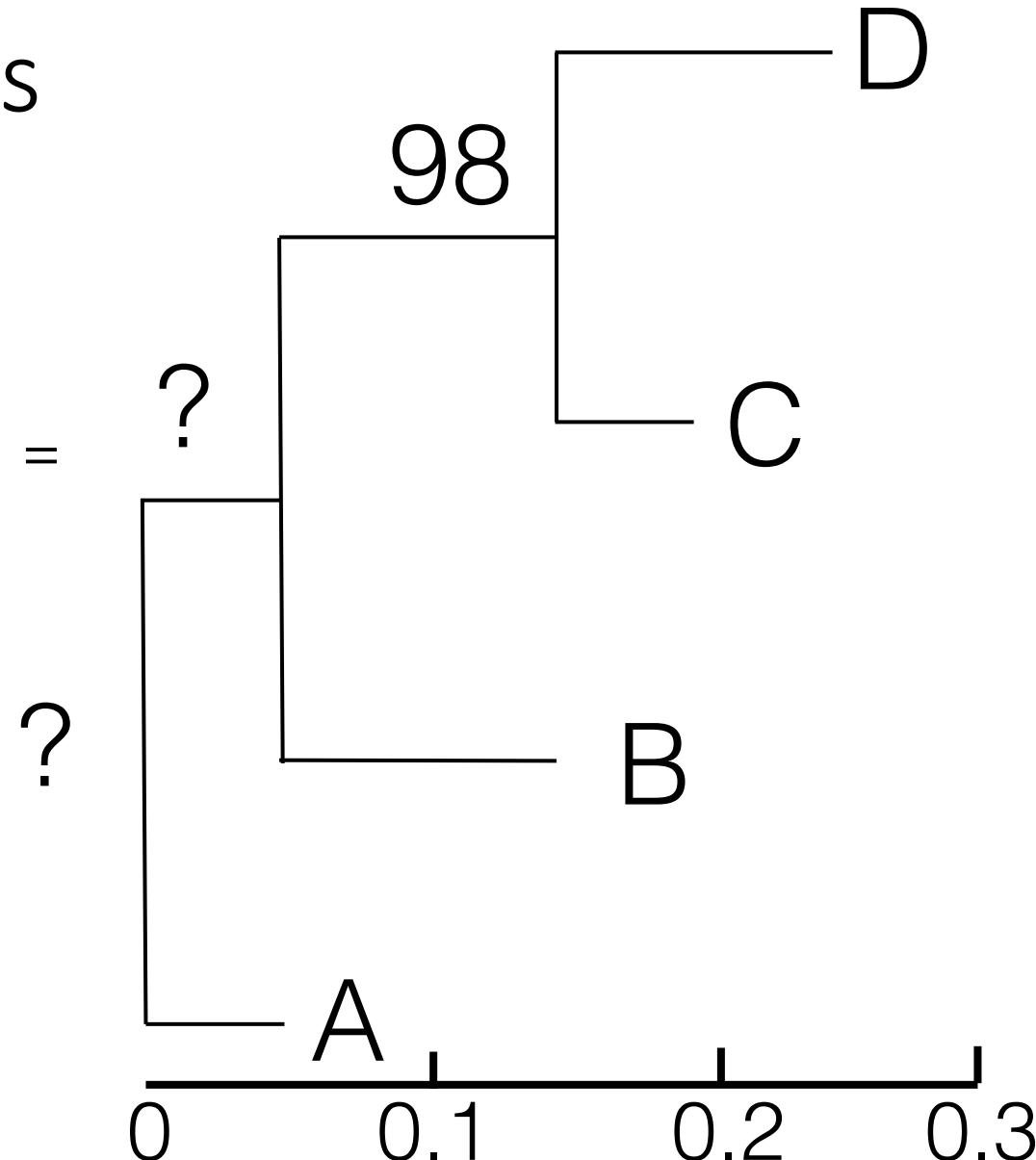


Newick: The tree file format

Note that two nodes in this tree do not have statistical support!

(A:0.05,(B:0.1,(C:0.05,D:0.1)98:0.1):0.05); =

This is because these nodes were created by us rooting the tree.
Remember that analyses estimate unrooted trees.



Learning Goals

Newick: The tree file format

Bootstrapping

Other support values

Visualization examples

Visualization tools

Bootstrapping

A statistical approach to measuring uncertainty in parameter estimates due to sampling error.

Often used to generate confidence intervals.

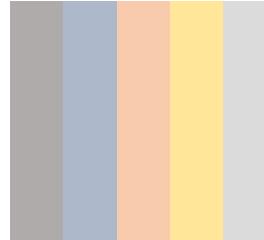
Bootstrapping

Applied to phylogenetics by resampling sites.

Often used to determine clades we trust and clades we do not trust.

Bootstrapping

The Standard Bootstrap

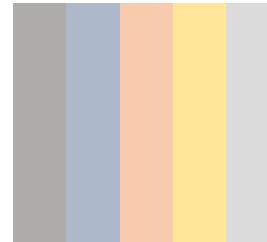


each color is
a different site

Bootstrapping

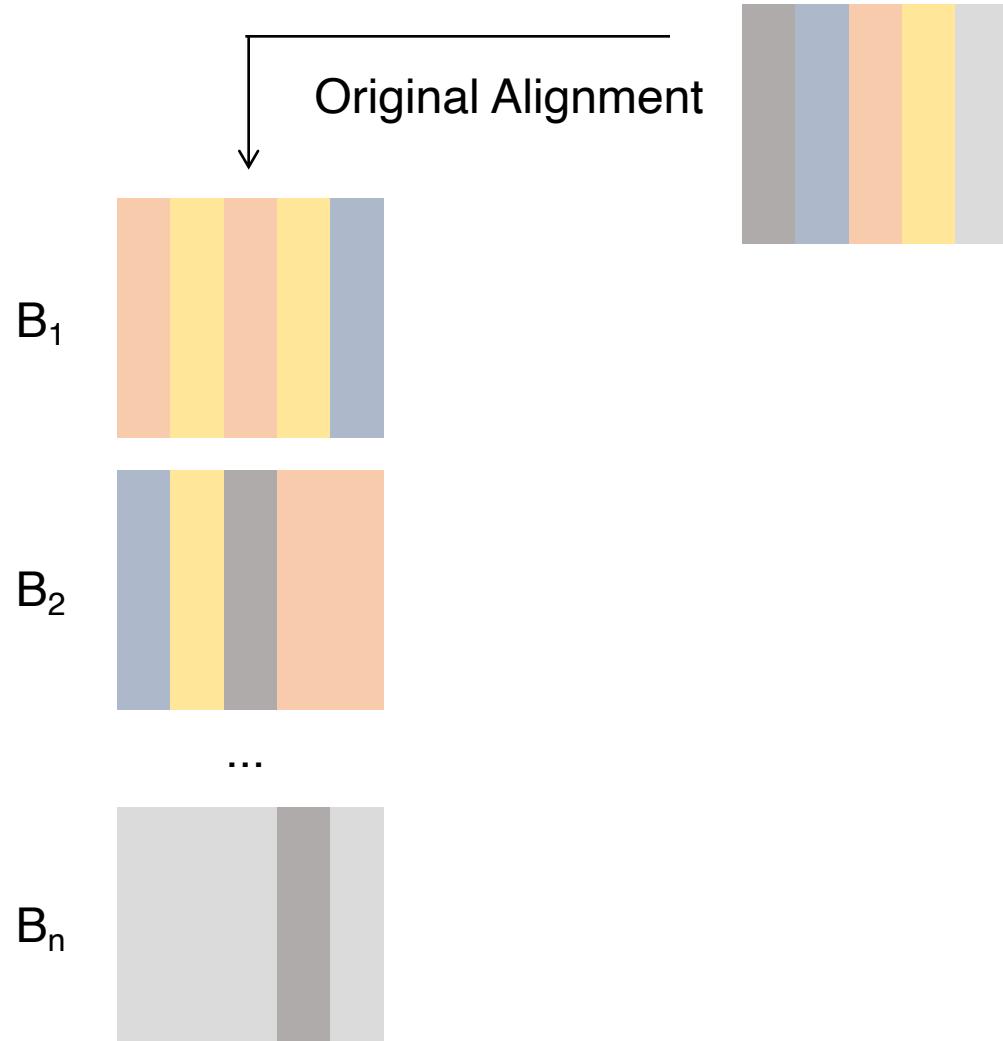
The Standard Bootstrap

Original Alignment



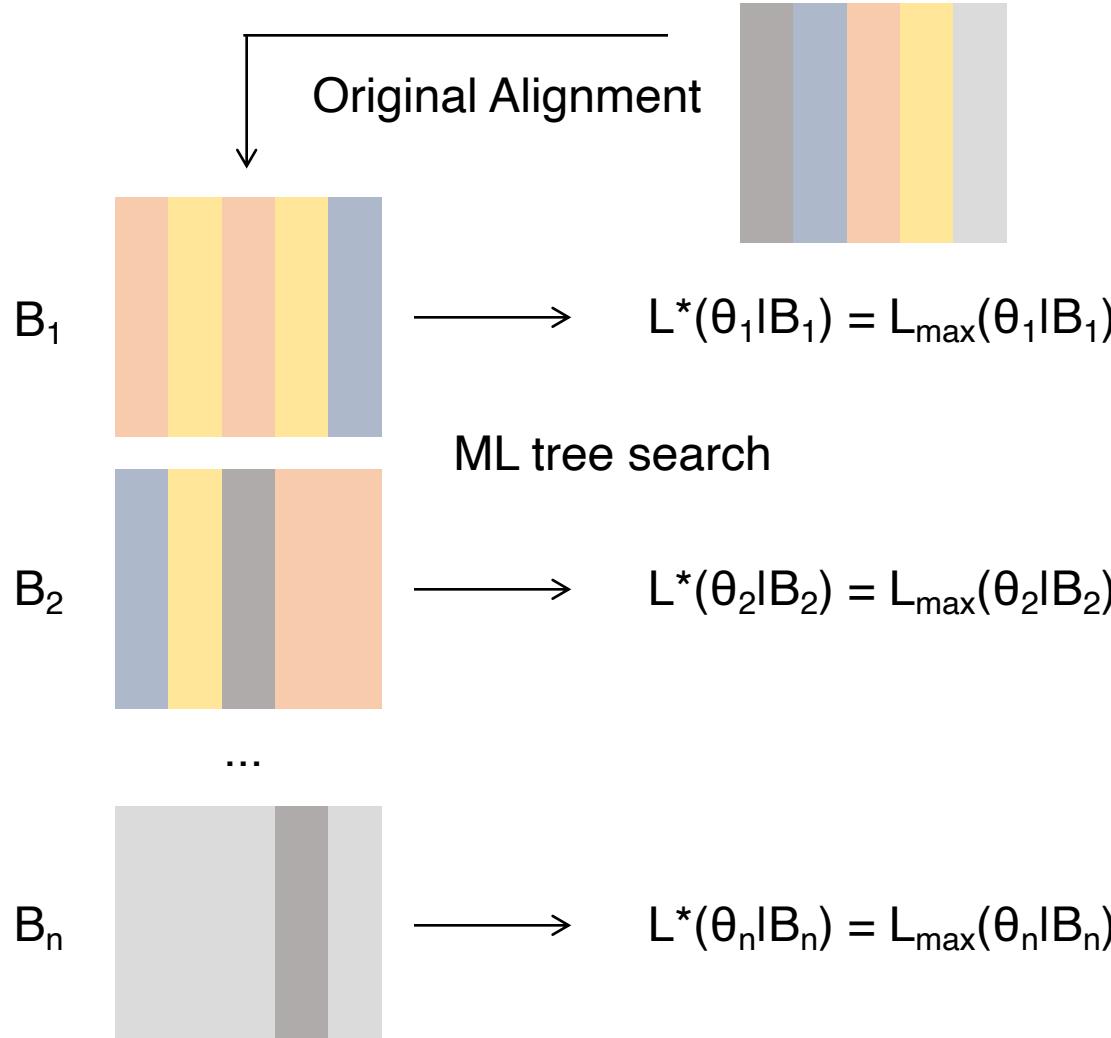
Bootstrapping

The Standard Bootstrap



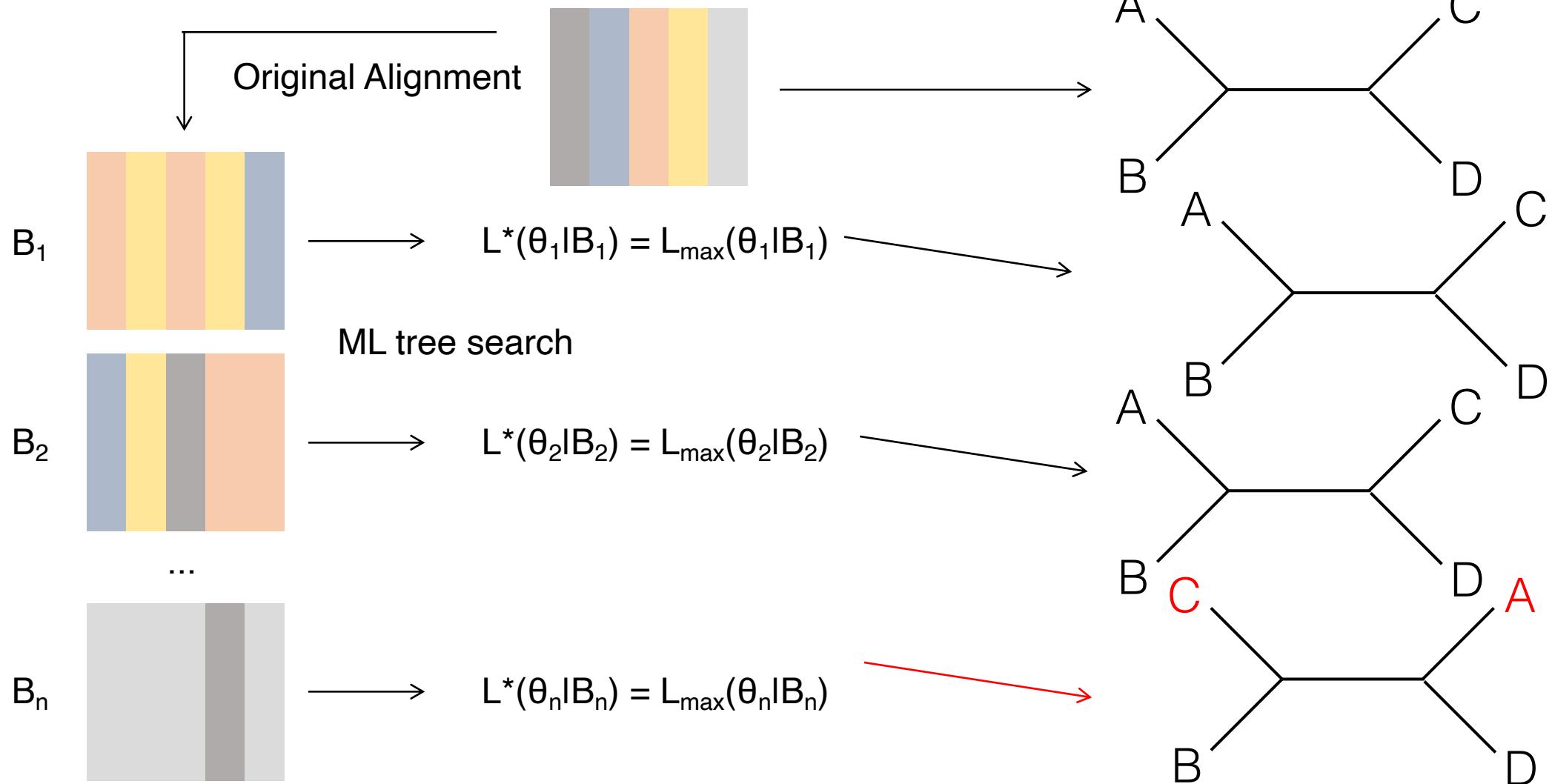
Bootstrapping

The Standard Bootstrap



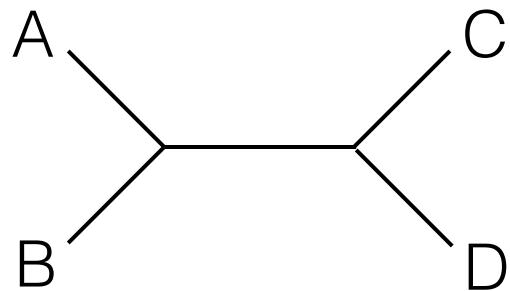
Bootstrapping

The Standard Bootstrap – an example



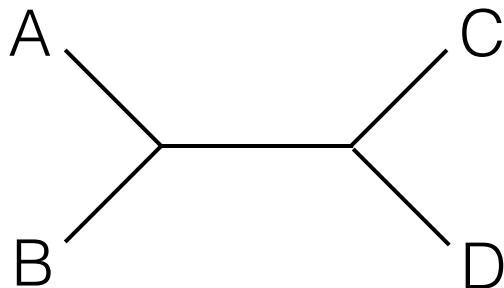
Bootstrapping

our ML tree based on the
original alignment

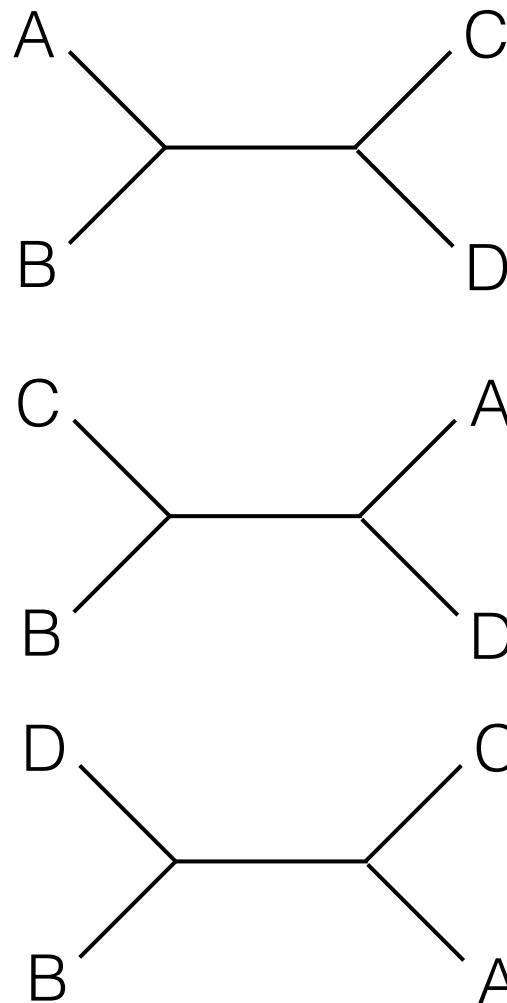


Bootstrapping

our ML tree based on the original alignment



Potential bootstrap trees



Frequency among replicates

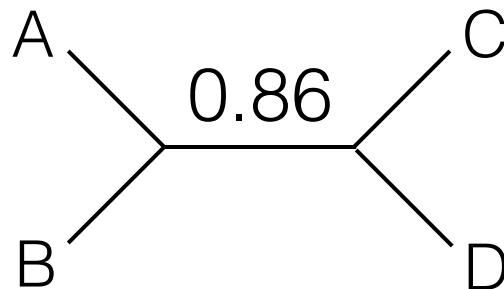
86

10

4

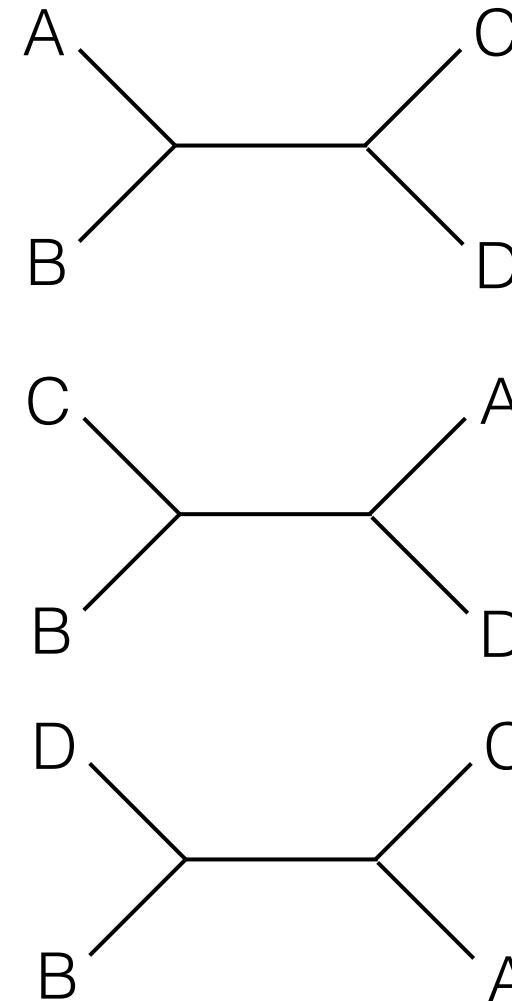
Bootstrapping

our ML tree based on the original alignment



Resulting bootstrap value on the ML tree

Potential bootstrap trees



Frequency among replicates

86

10

4

Bootstrapping

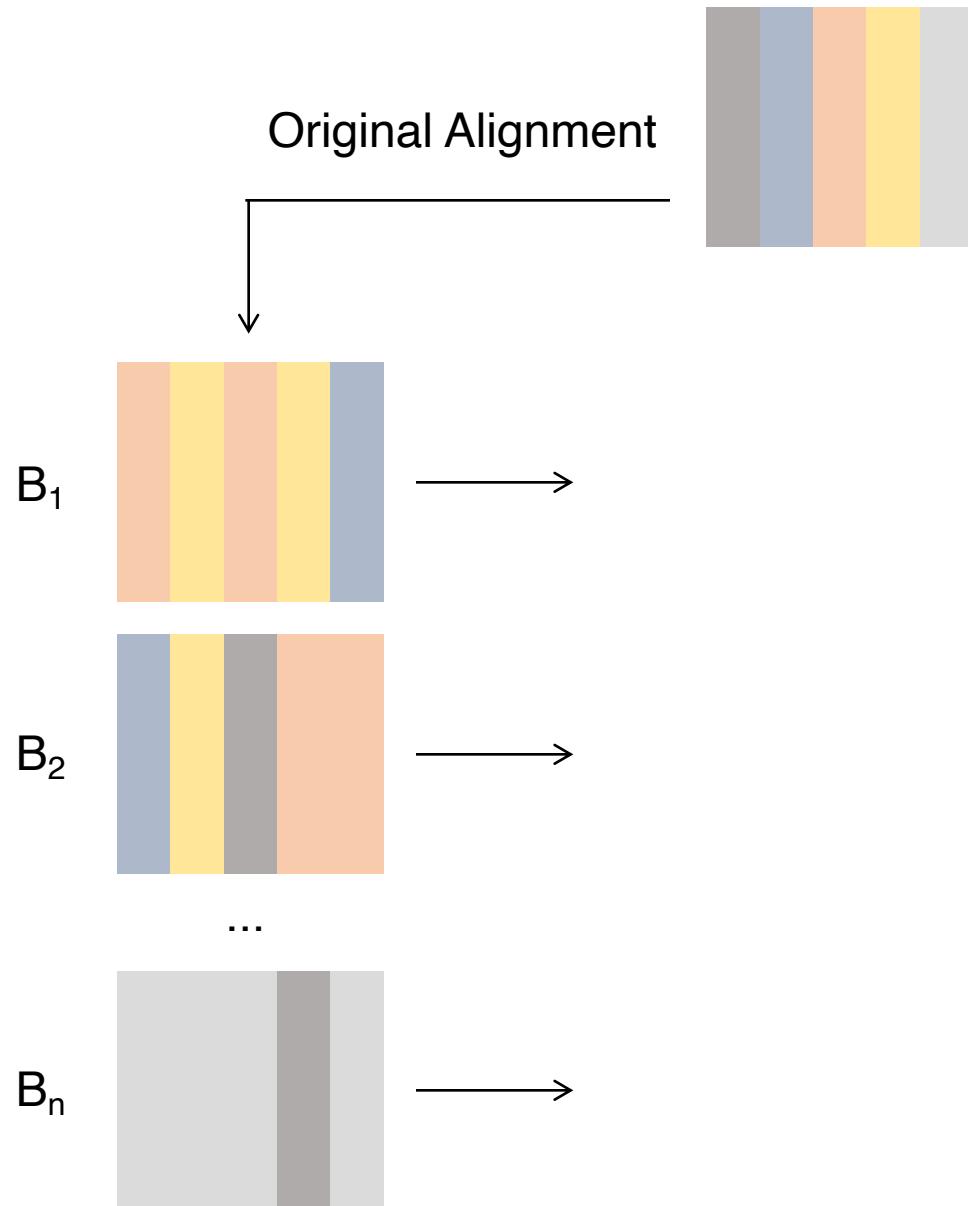
The standard bootstrap is frequently considered to be conservative

Values greater than 80% or greater than 70% are interpreted as a well-supported node.

If we rejected nodes with less than 95% support, we would miss a lot of true relationships.

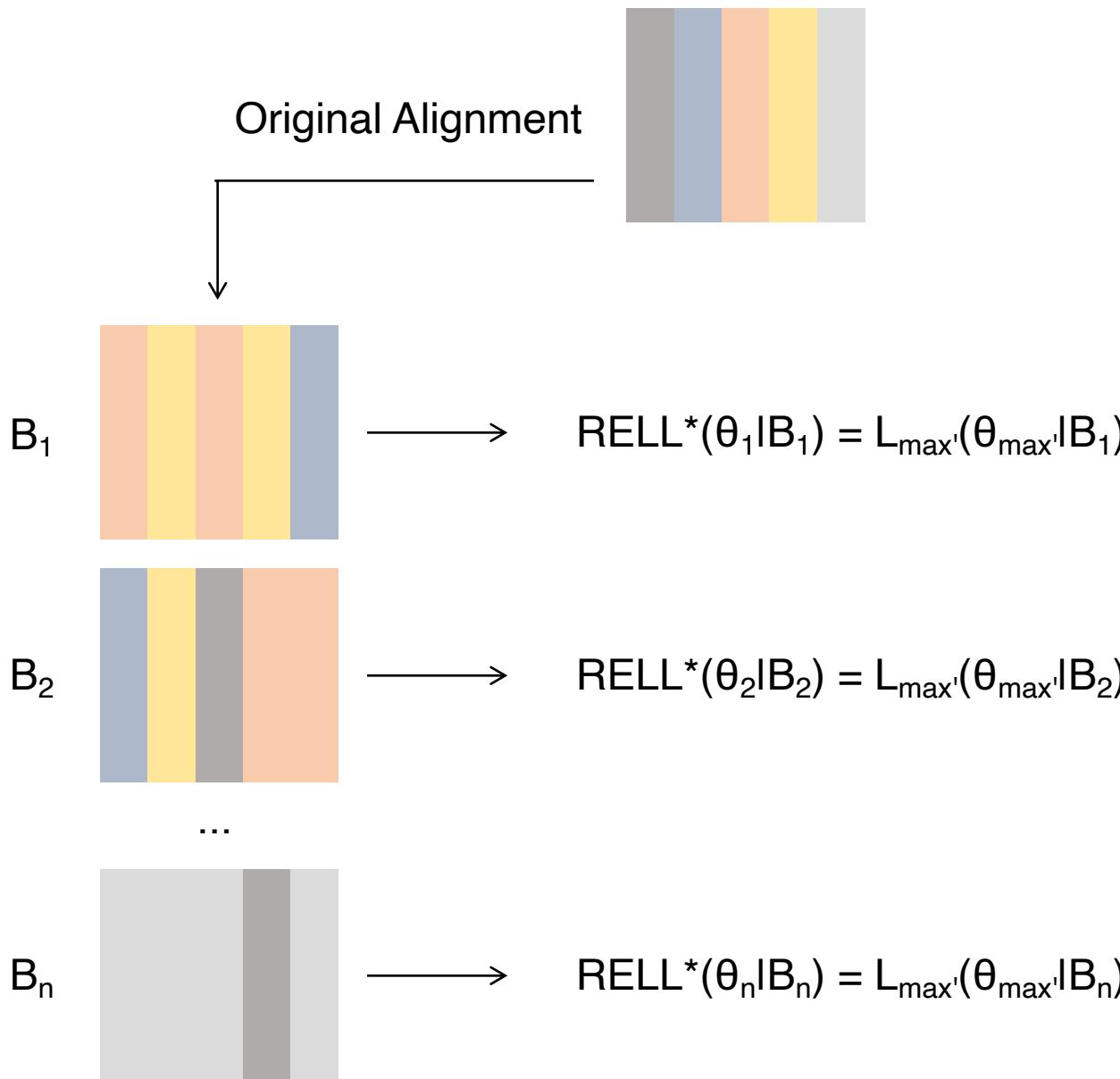
Zharkikh and Li 1992;
Hillis and Bull 1993

The Ultra Fast Bootstrap



Minh et al. 2013

The Ultra Fast Bootstrap



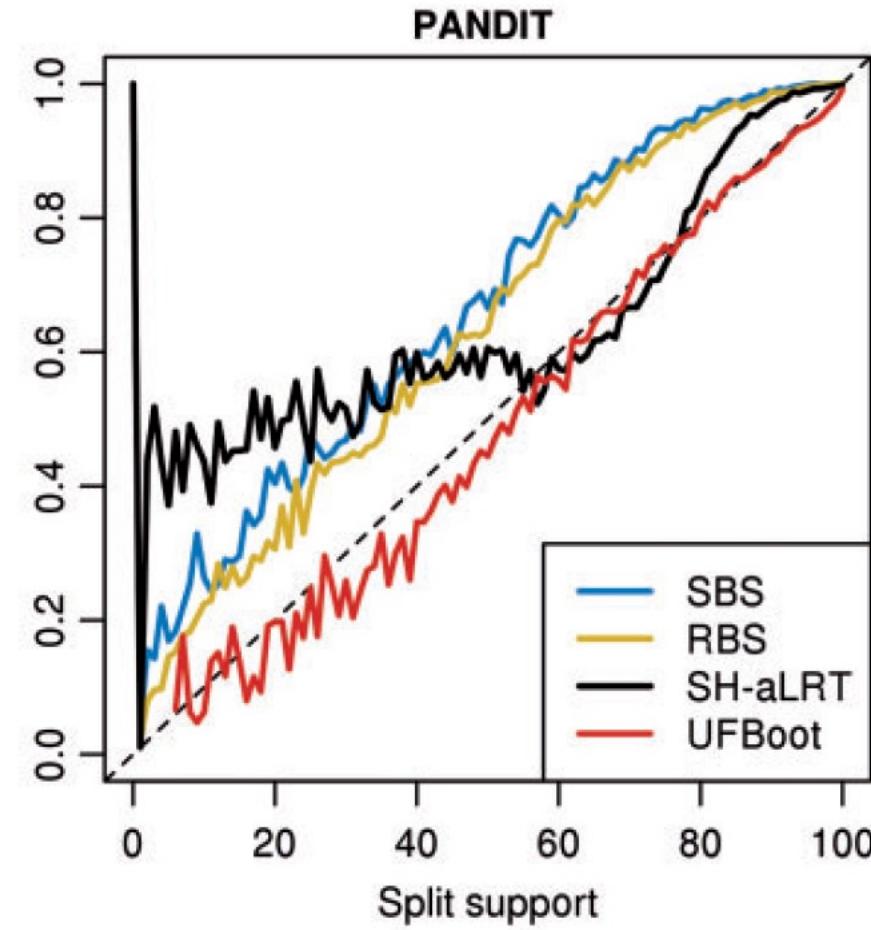
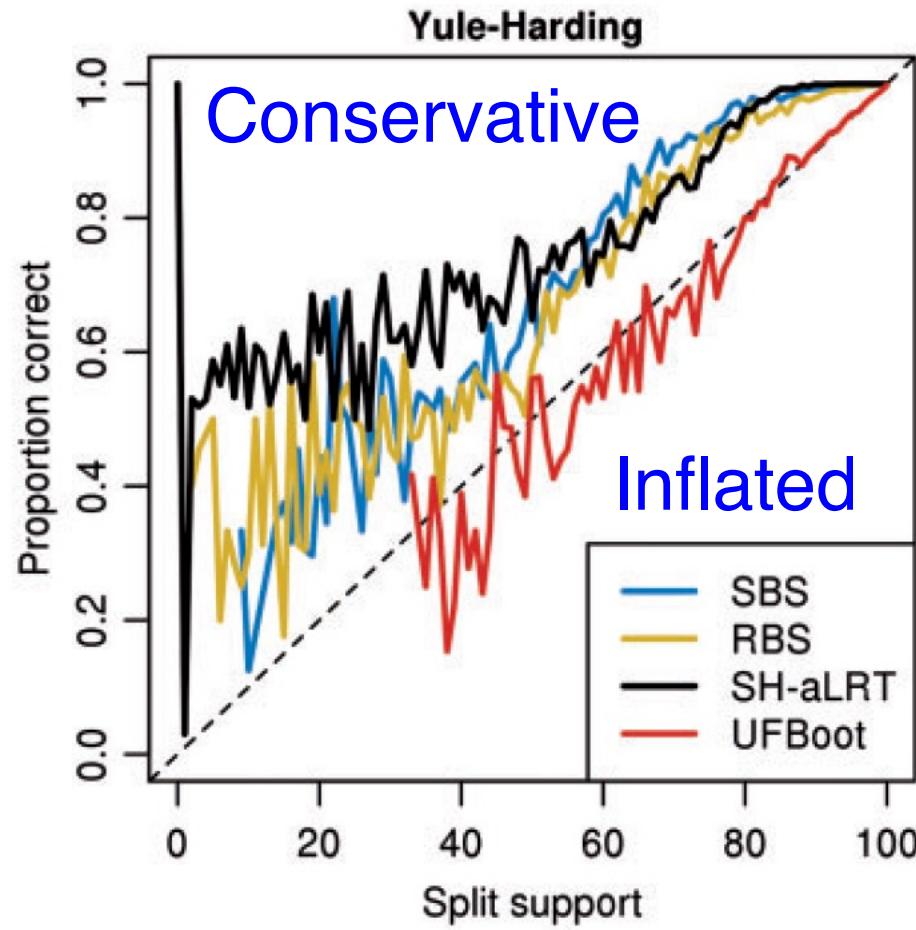
Minh et al. 2013

Advantages of Ultra Fast Bootstrap

- 1) No tree search for each replicate
- 2) No recomputing of likelihoods for each replicate

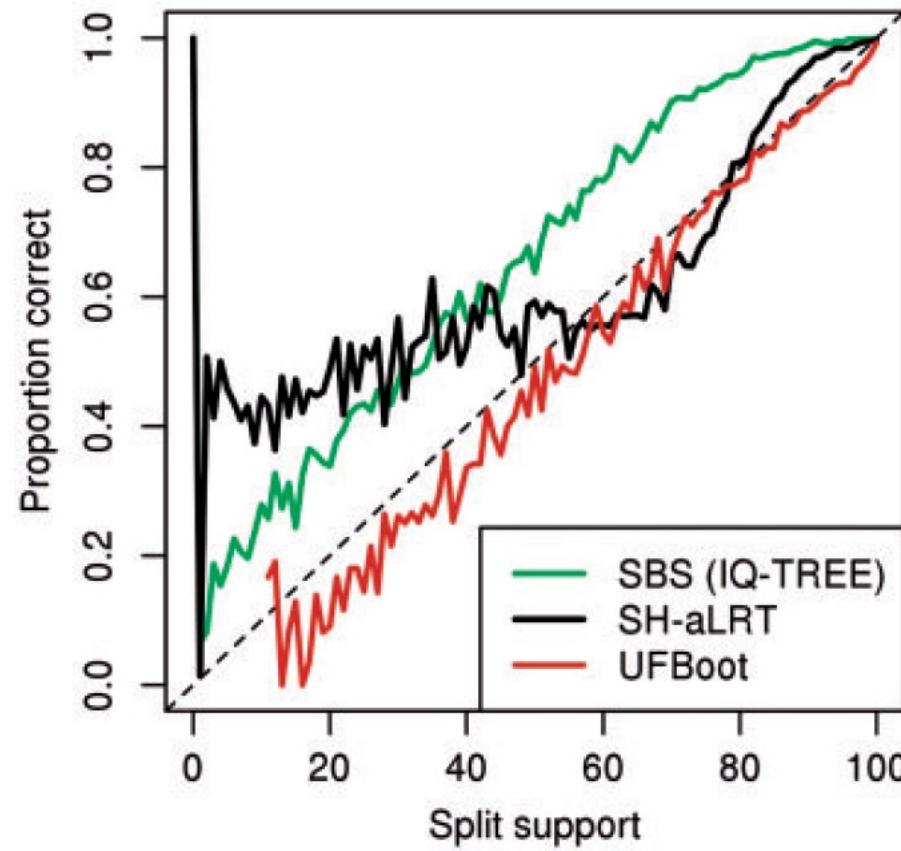
Importance Quartet Sampling – Getting around the tree search

RELL – Resampling Estimated Log Likelihoods

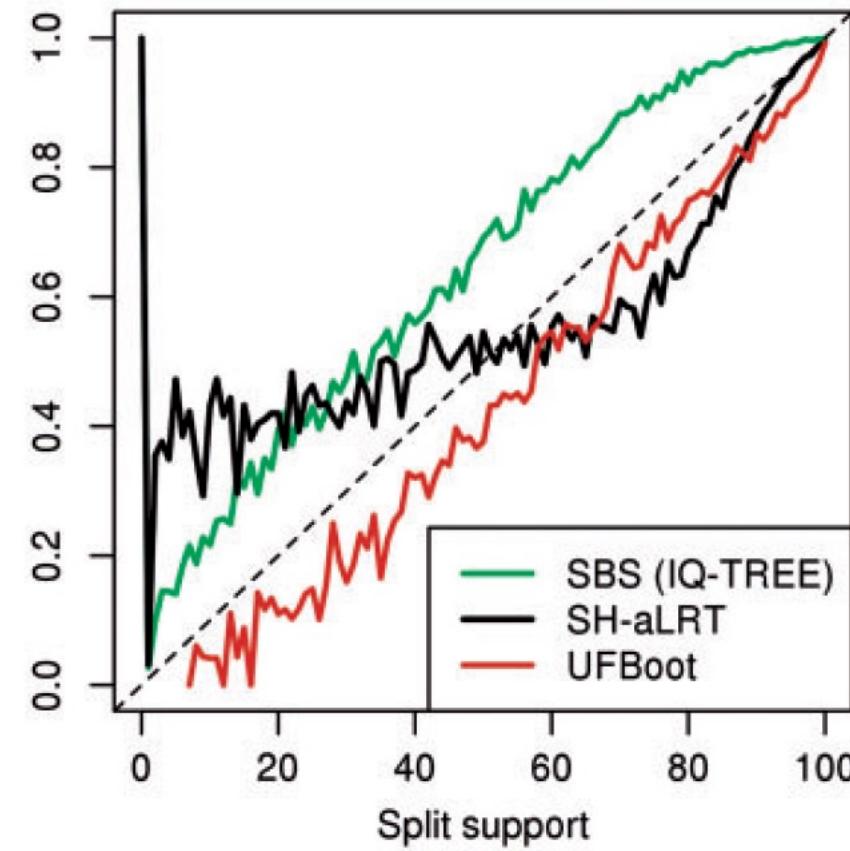


Proportion Correct = $\frac{\text{\# of Correct splits at X\%}}{\text{total \# of splits with X\%}}$

Minh et al. 2013

PANDIT (JC+ Γ)

PANDIT (JC)



Additional support value is the SH-aLRT
(Shimodaira-Hasegawa-like approximate Likelihood Ratio Test)

Minh et al. 2013

APPENDIX 3

SH-like interpretation of aLRT statistic

Algorithm: SH-like interpretation of aLRT statistic

Input: lists of site log-likelihood values sLNL1, sLNL2, sLNL3

Output: branch-support SH

$aLRT \leftarrow 2(LNL1 - LNL2);$

$SH \leftarrow 0;$

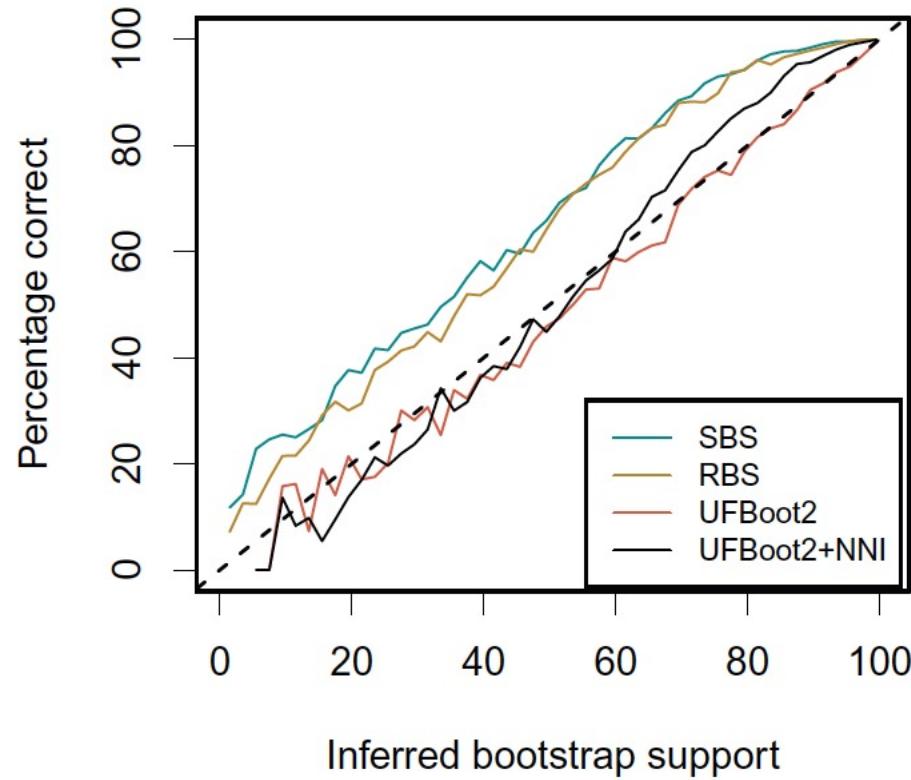
For $i = 1$ to ReplicateNumber (typically 1000 or 10,000):

{

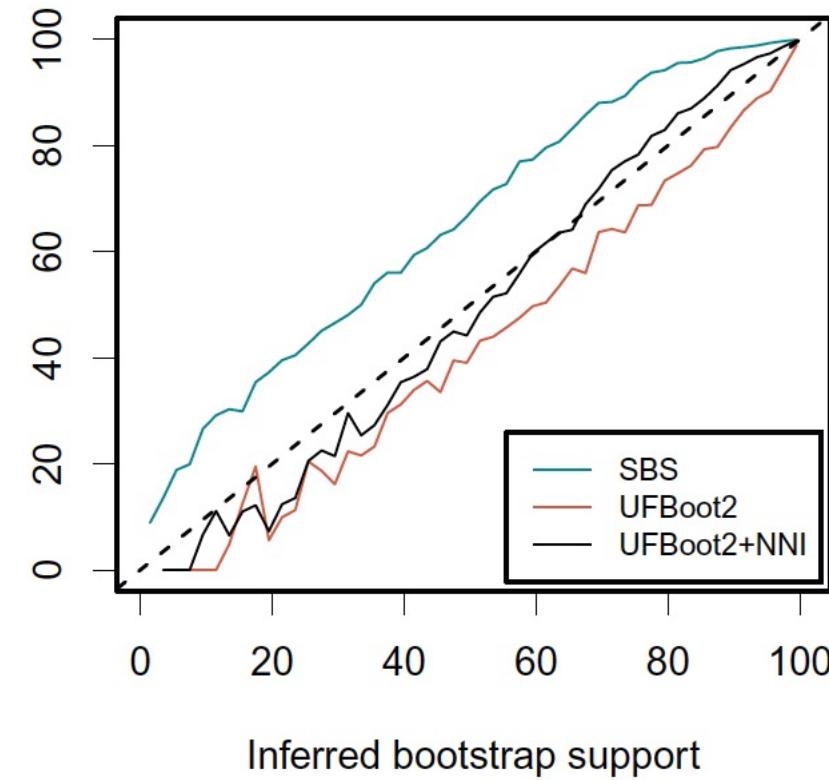
- A. Draw $sLNL1^*$, $sLNL2^*$ and $sLNL3^*$;
- B. Compute $LNL1^*$, $LNL2^*$ and $LNL3^*$;
- C. Compute centered sums $CS1^*$, $CS2^*$ and $CS3^*$, using $CSX^* \leftarrow LNLX^* - LNLX$;
- D. Let CS_First and CS_Second be the highest and second highest CSX^* values, respectively;
- E. If $aLRT > 2(CS_First - CS_Second) + \varepsilon$, then
 $SH \leftarrow SH+1;$ }

Return $SH/ReplicateNumber$

A **Correctly specified model**

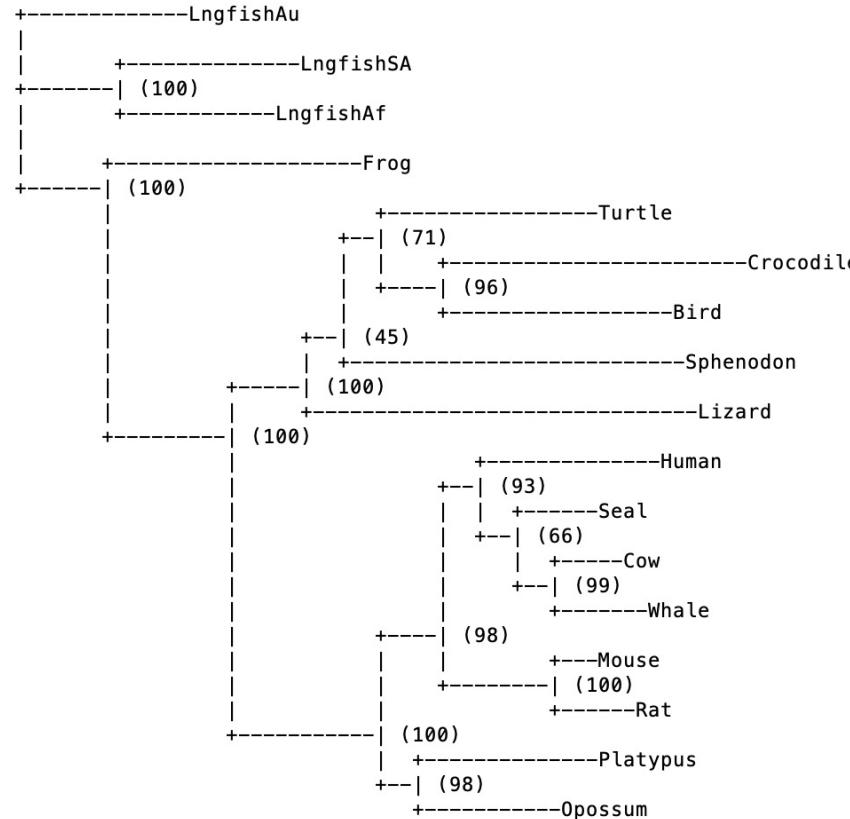


B **Severe violation**



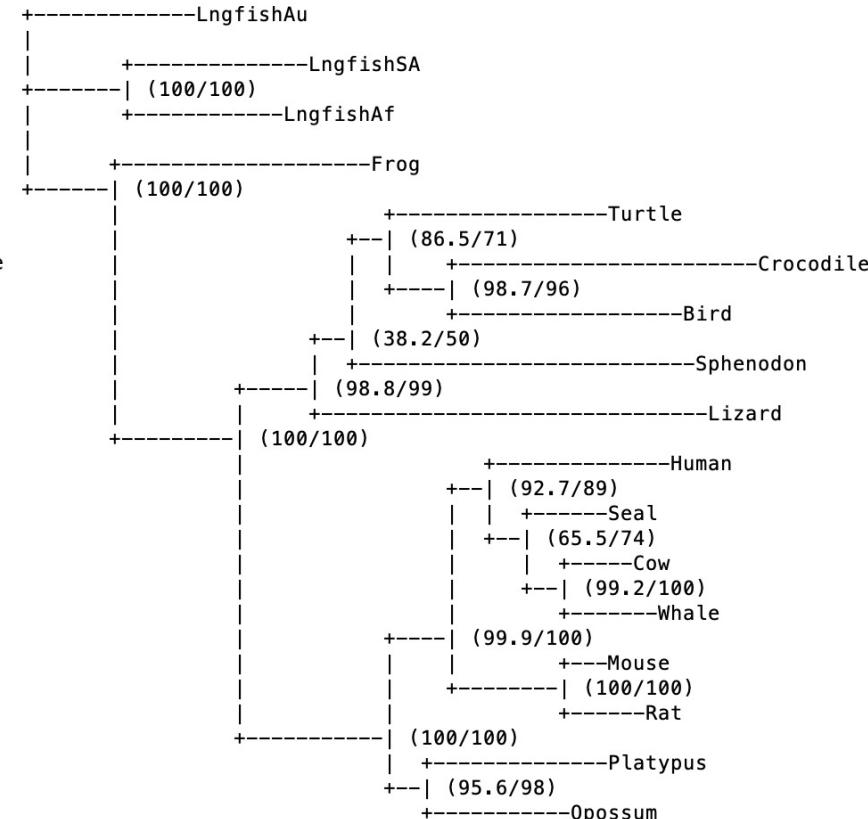
SBS result

NOTE: Tree is UNROOTED although outgroup taxon 'LngfishAu' is drawn at root
Numbers in parentheses are standard bootstrap support (%)



UFBOOT result

NOTE: Tree is UNROOTED although outgroup taxon 'LngfishAu' is drawn at root
Numbers in parentheses are SH-aLRT support (%) / ultrafast bootstrap support (%)

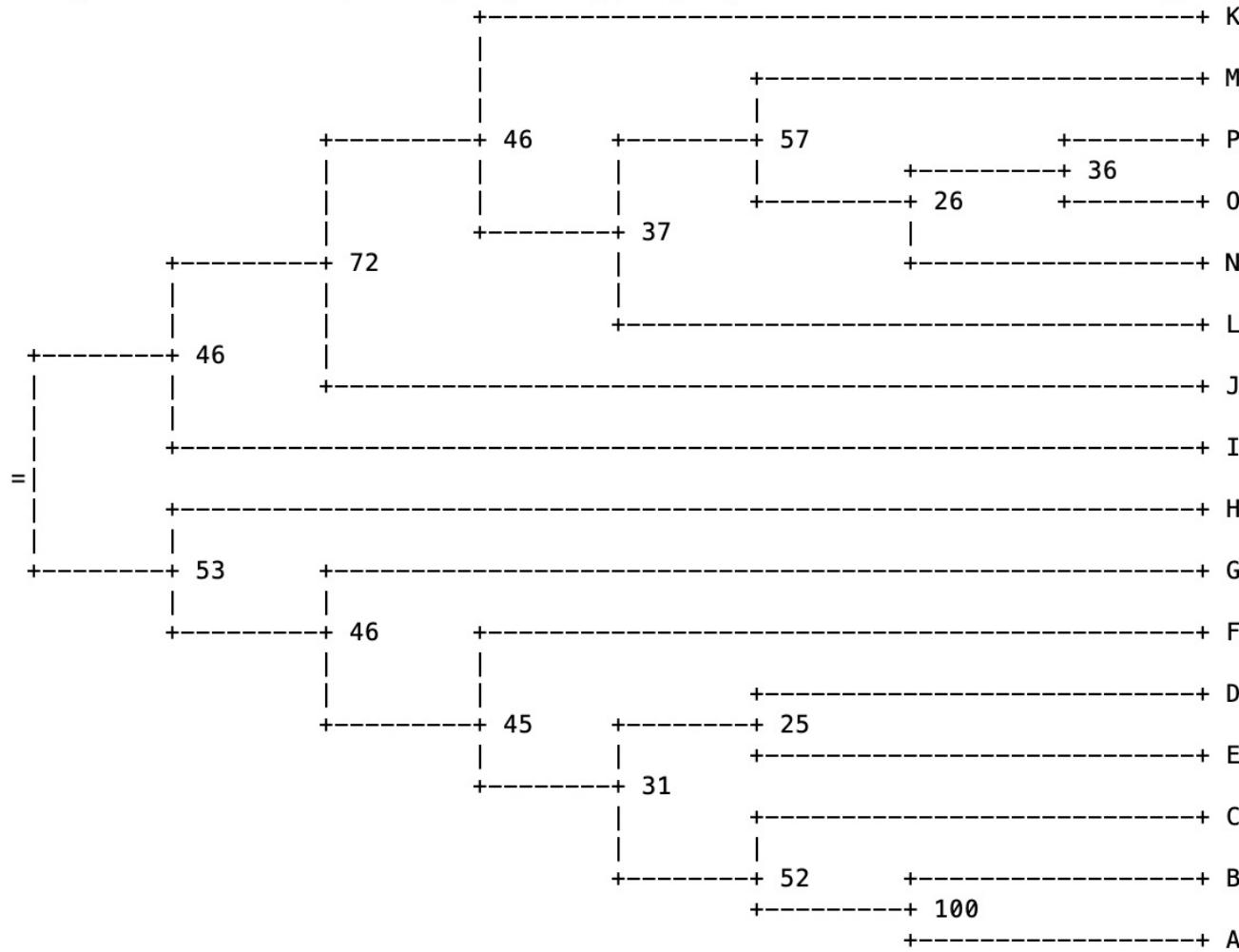


12 seconds

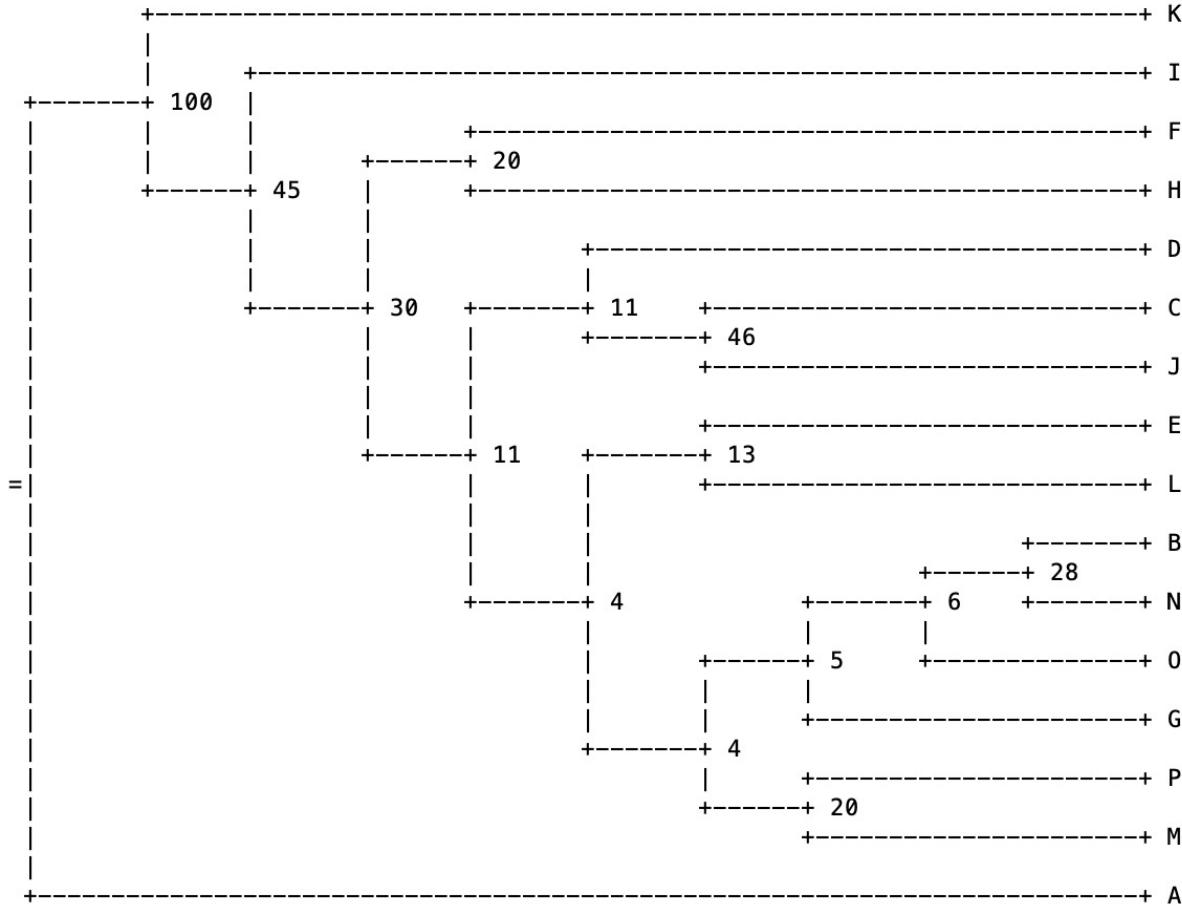
13 seconds

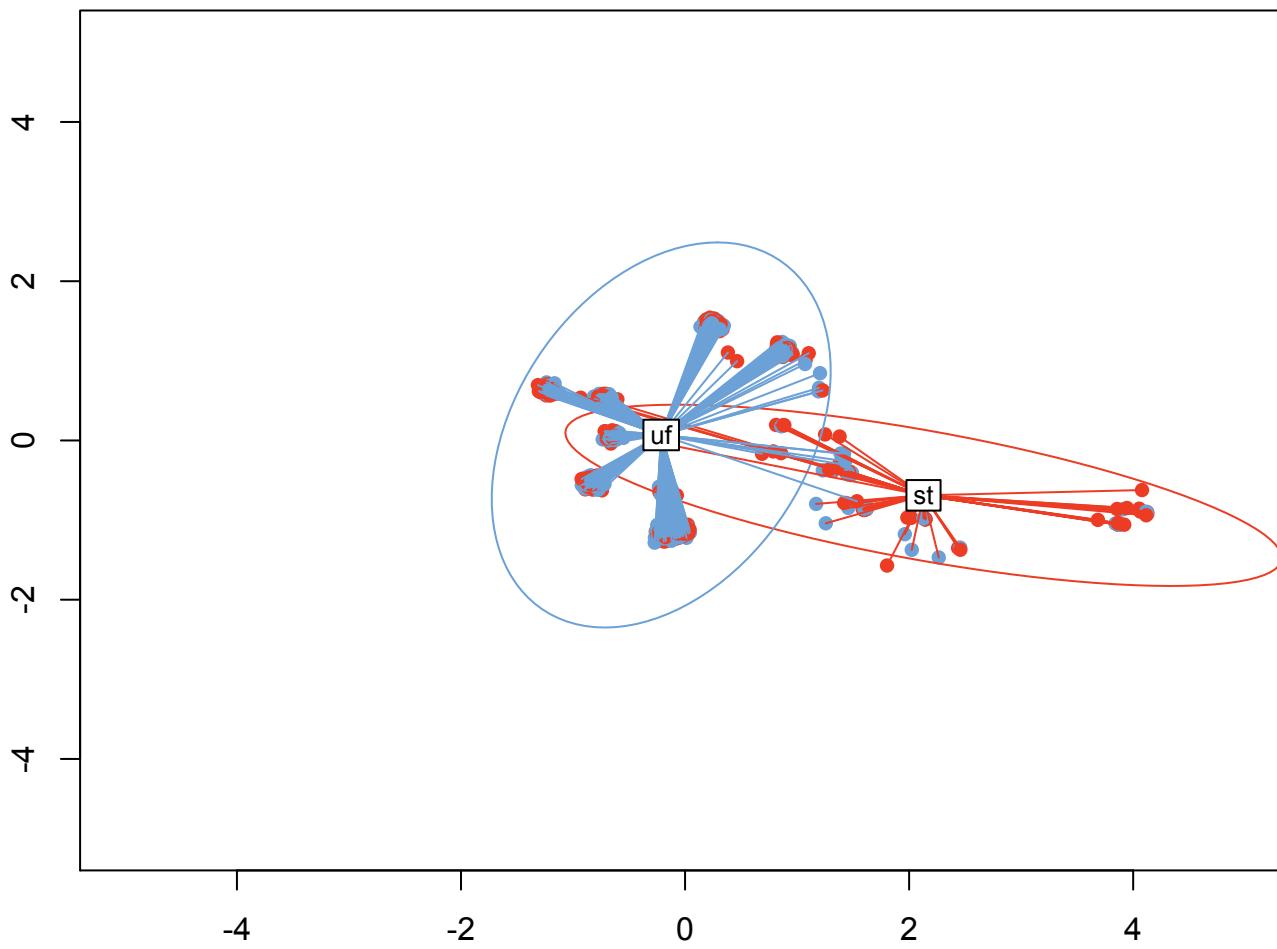
Potential topological biases
from trees obtained with
the ufboot algorithm

```
[Georges-MacBook-Pro:iqstuff georgetiley$ ~/programs/newick-utils-1.6/src/nw_display stboot.astral.tre.contree.reroot
```



```
[Georges-MacBook-Pro:iqstuff georgetiley$ ~/programs/newick-utils-1.6/src/nw_display ufboot.astral.tre.contree.reroot
```





Learning Goals

Newick: The tree file format

Bootstrapping

Other support values

Visualization examples

Visualization tools

Other support values

ASTRAL does not use a bootstrap. It uses a value called “local posterior probability.” We often accept 0.95 or greater as a supported node in the ASTRAL species tree.

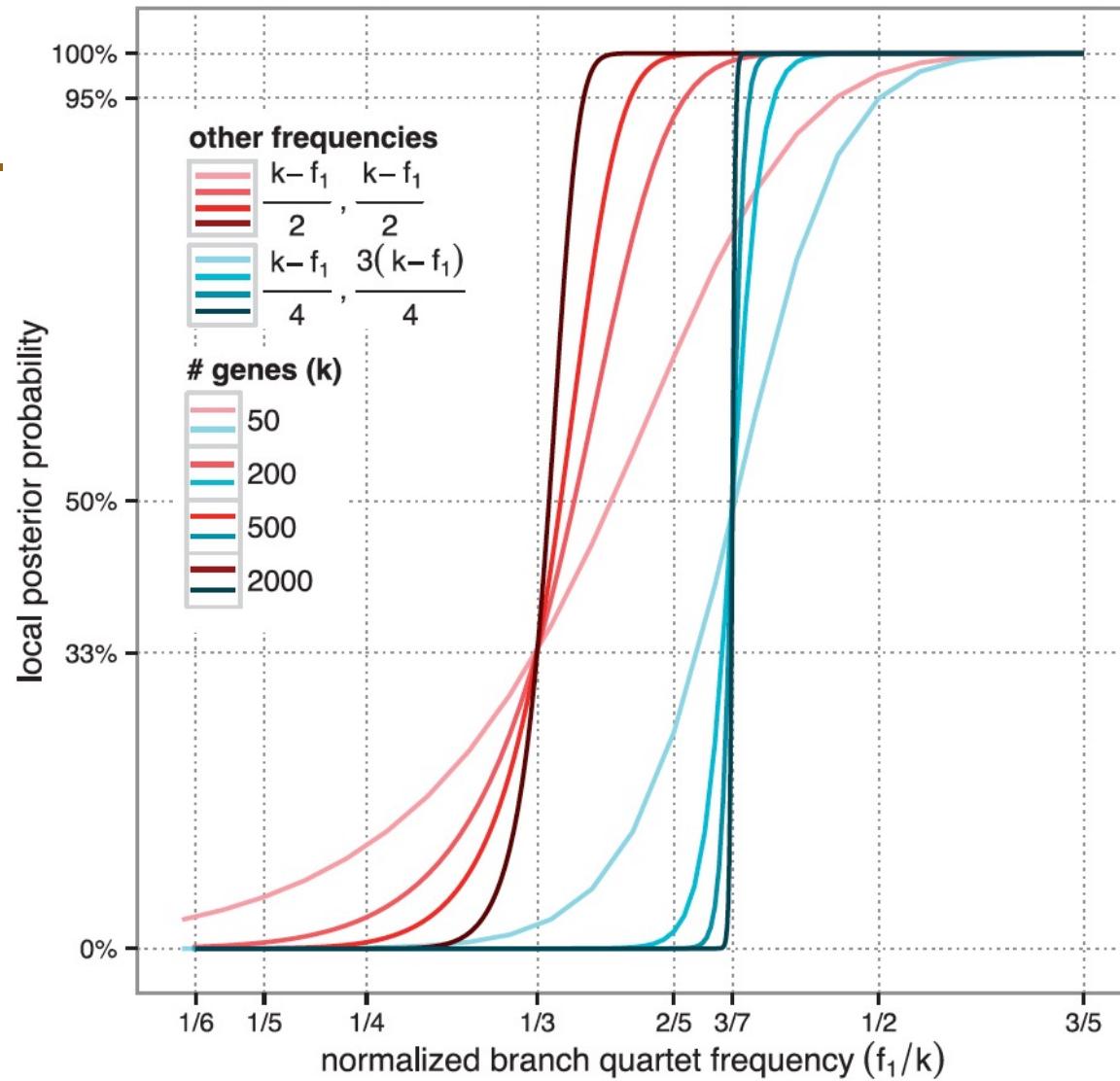
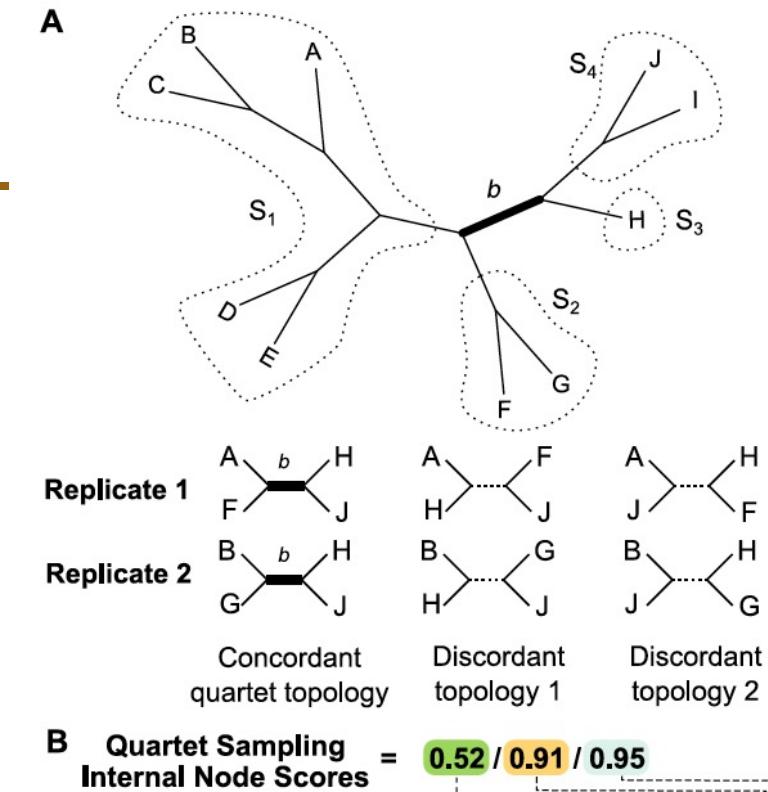


FIG. 2. The local PP of a branch as a function of its normalized quartet support for varying numbers of genes. Red lines: alternative topologies have equal frequencies (thus, conform to properties of the MSC for $x > 1/3$); Blue lines: alternative topologies do not have equal frequencies (contrary to the MSC).

Other support values

Quartet sampling

- samples 4-taxon subsets of the alignment across genes
- estimates likelihood scores for the 3 possible *quartets*
- used to estimate 3 support values that can inform us about gene tree variation



Quartet Concordance (QC) How often is the concordant quartet inferred over both discordant quartets?	QC=1 → all concordant QC=0 → equivocal conc./disc. QC<0 → discordant > conc.
Quartet Differential (QD) Are discordant #1 and #2 frequencies equal or skewed?	QD=1 → equal #1 and #2 QD=0.3 → skewed QD=0 → all #1 or #2
Quartet Informativeness (QI) What proportion of replicates were informative?	QI=1 → all informative QI=0.3 → 30% informative (exceeded likelihood differential) QI=0 → none informative
Quartet Sampling Terminal Node Scores = (0.52)	
Quartet Fidelity (QF) When this taxon is sampled, how often does it produce a concordant topology?	Examples: QF=1 → all concordant QF=0.1 → 10% concordant QF=0 → none concordant

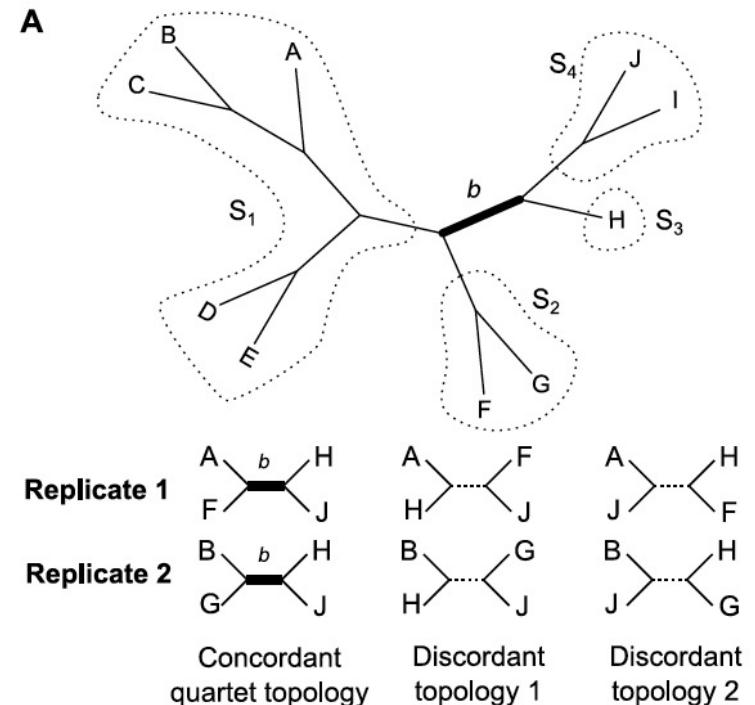
Other support values

Quartet sampling

TABLE 1. Quartet Sampling (QS) score interpretation.

Example QS score (QC/QD/QI)	Interpretation
1.0/-1.0	Full support: All sampled quartet replicates support the focal branch (QC = 1) with all trees informative when likelihood cutoffs are used (QI = 1).
0.5/0.98/0.97	Strong support: A strong majority of quartets support the focal branch (QC = 0.5), and the low skew in discordant frequencies (QD ≈ 1) indicate no alternative history is favored.
0.7/0.1/0.97	Strong support with discordant skew: A strong majority of quartets support the focal branch (QC = 0.7), but the skew in discordance (QD = 0.1) indicates the possible presence of a supported secondary evolutionary history.
0.05/0.96/0.97	Weak support: Only a weak majority of quartets support the focal branch (QC = 0.05), and the frequency of all three possible topologies is similar (QD ≈ 1).
0.1/0.1/0.97	Weak support with discordant skew: Only a weak majority of quartets support the focal branch (QC = 0.1), and the skew in discordance (QD = 0.1) indicates the possible presence of a supported secondary evolutionary history.
-0.5/0.1/0.93	Counter-support: A strong majority of quartets support one of the alternative discordant quartet arrangement history (QC < 0; QD expected to be low).
1/0.97/0.05	Poorly informed: Despite supportive QC/QD values, only 5% of quartets passed the likelihood cutoff (QI = 0.05), likely indicating few informative sites.
0.0/0.0/1.0	Perfectly conflicted: The (unlikely) case where the frequencies of all three possible trees are equal and all trees are informative, which indicates a rapid radiation or highly complex conflict.

Notes: QC = Quartet Concordance; QD = Quartet Differential; QI = Quartet Informativeness.



B Quartet Sampling Internal Node Scores = 0.52 / 0.91 / 0.95

Quartet Concordance (QC)

How often is the concordant quartet inferred over both discordant quartets?

- QC=1 → all concordant
- QC=0 → equivocal conc./disc.
- QC<0 → discordant > conc.

Quartet Differential (QD)

Are discordant #1 and #2 frequencies equal or skewed?

- QD=1 → equal #1 and #2
- QD=0.3 → skewed
- QD=0 → all #1 or #2

Quartet Informativeness (QI)

What proportion of replicates were informative?

- QI=1 → all informative
- QI=0.3 → 30% informative (exceeded likelihood differential)
- QI=0 → none informative

Quartet Sampling Terminal Node Scores = (0.52)

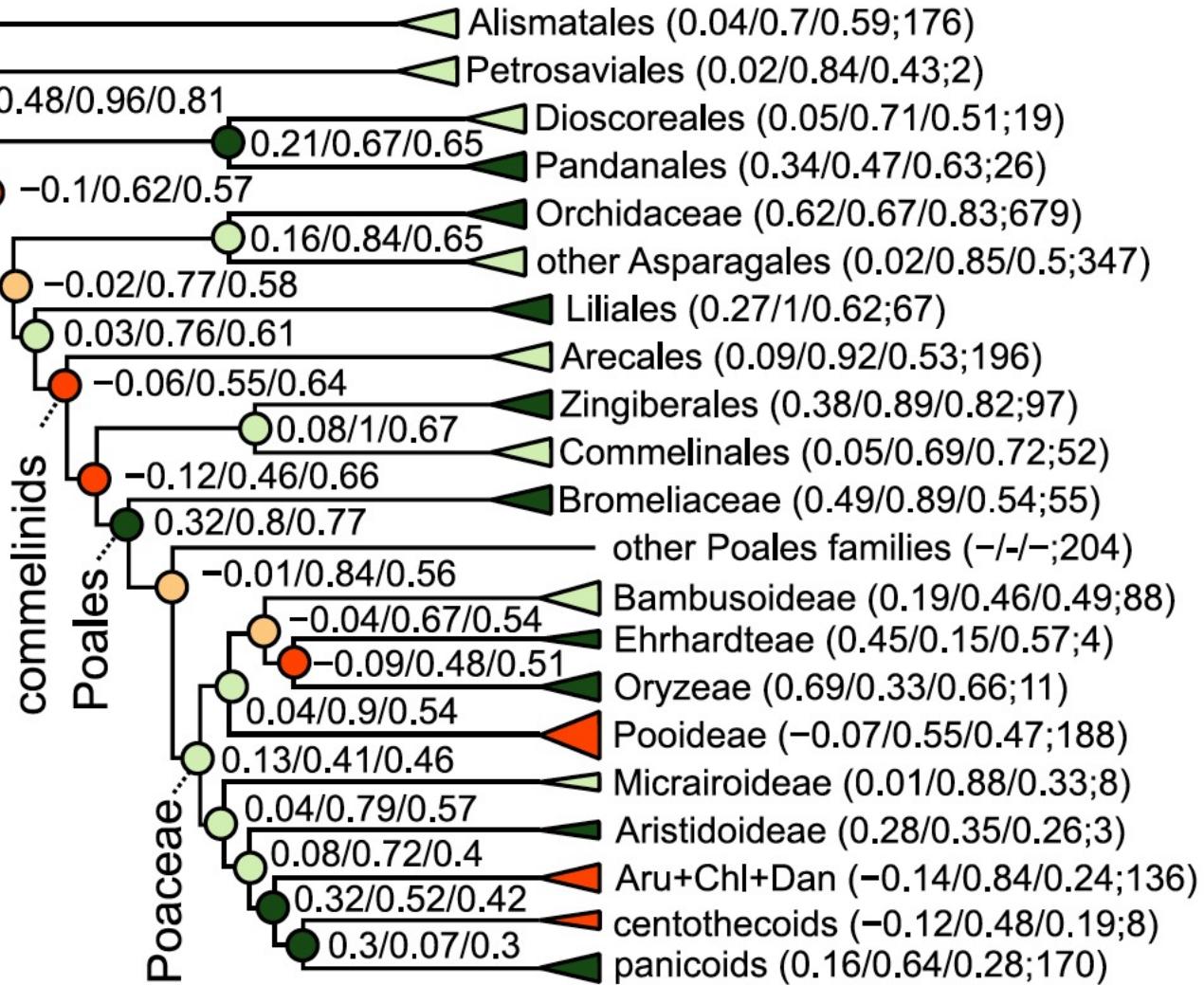
Quartet Fidelity (QF)

When this taxon is sampled, how often does it produce a concordant topology?

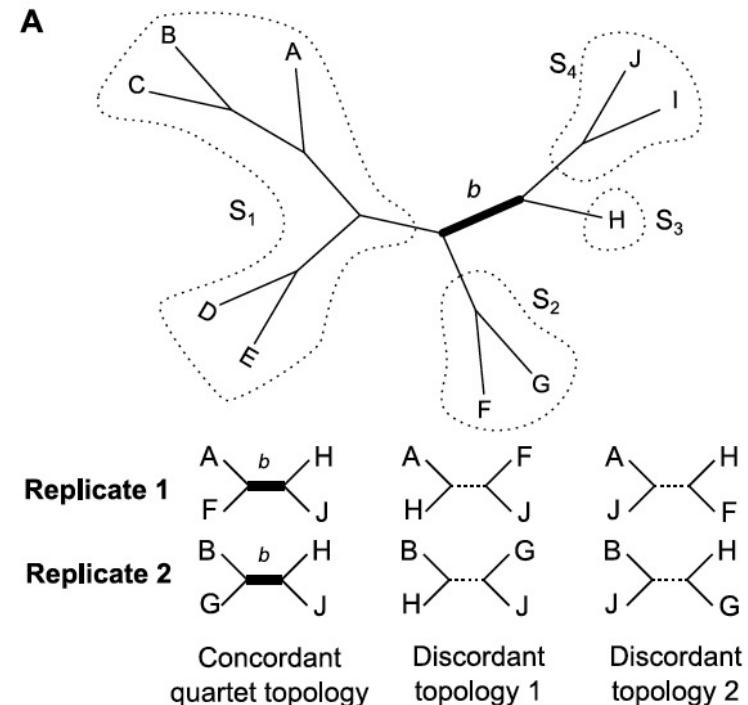
- QF=1 → all concordant
- QF=0.1 → 10% concordant
- QF=0 → none concordant

Other support values

C. Monocots



Pease et al. 2018



B Quartet Sampling Internal Node Scores = $0.52 / 0.91 / 0.95$

Quartet Concordance (QC)

How often is the concordant quartet inferred over both discordant quartets?

- QC=1 → all concordant
- QC=0 → equivocal conc./disc.
- QC<0 → discordant > conc.

Quartet Differential (QD)

Are discordant #1 and #2 frequencies equal or skewed?

- QD=1 → equal #1 and #2
- QD=0.3 → skewed
- QD=0 → all #1 or #2

Quartet Informativeness (QI)

What proportion of replicates were informative?

- QI=1 → all informative
- QI=0.3 → 30% informative (exceeded likelihood differential)
- QI=0 → none informative

Quartet Sampling Terminal Node Scores = (0.52)

Quartet Fidelity (QF)

When this taxon is sampled, how often does it produce a concordant topology?

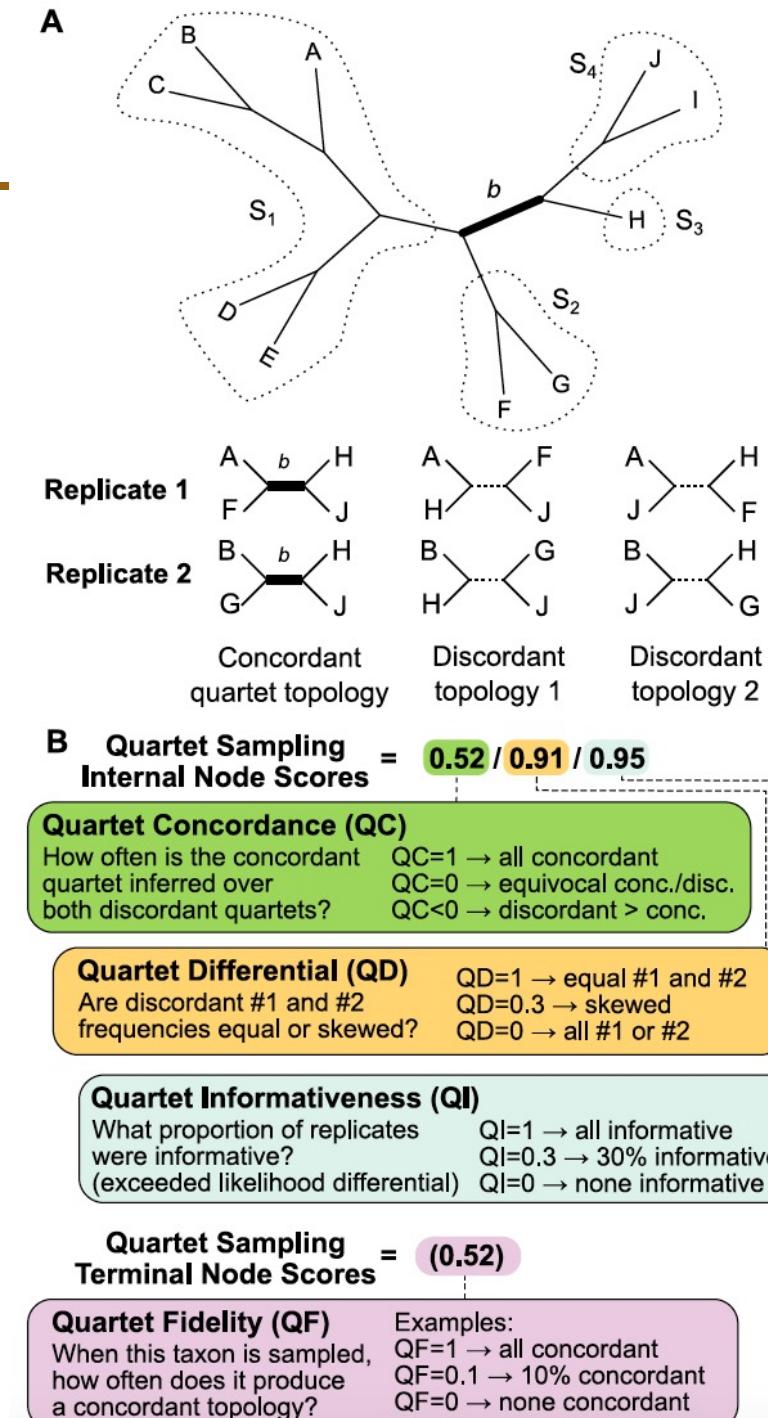
- QF=1 → all concordant
- QF=0.1 → 10% concordant
- QF=0 → none concordant

Other support values

Quartet sampling

The quartet informative scores can be calculated with respect to any tree

This means you could compare scores between your ML and MSC trees to check if the gene trees agree with one hypothesis for some specific nodes better than the other



Learning Goals

Newick: The tree file format

Bootstrapping

Other support values

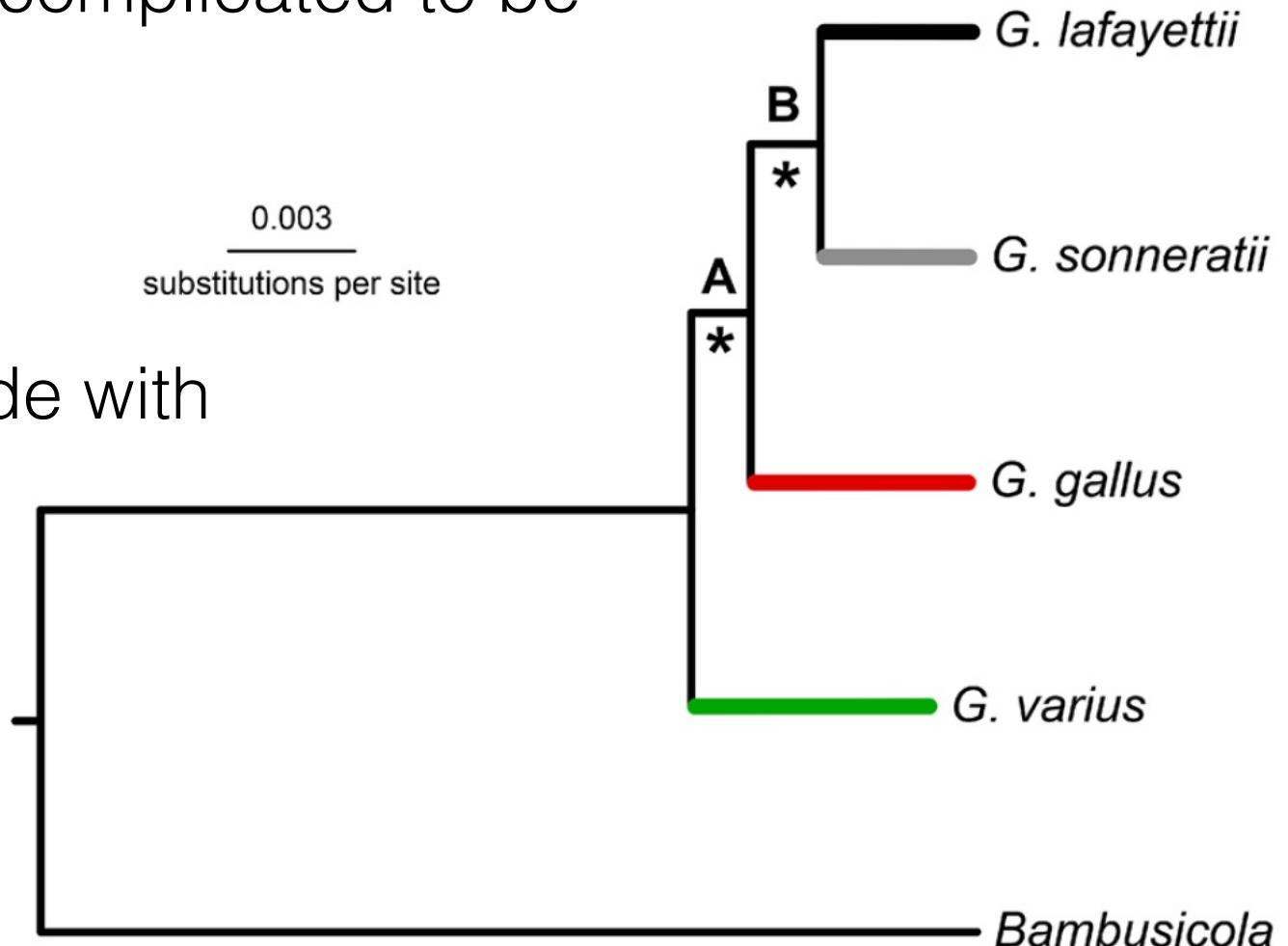
Visualization examples

Visualization tools

Visualization examples

Figures do not have to be complicated to be effective

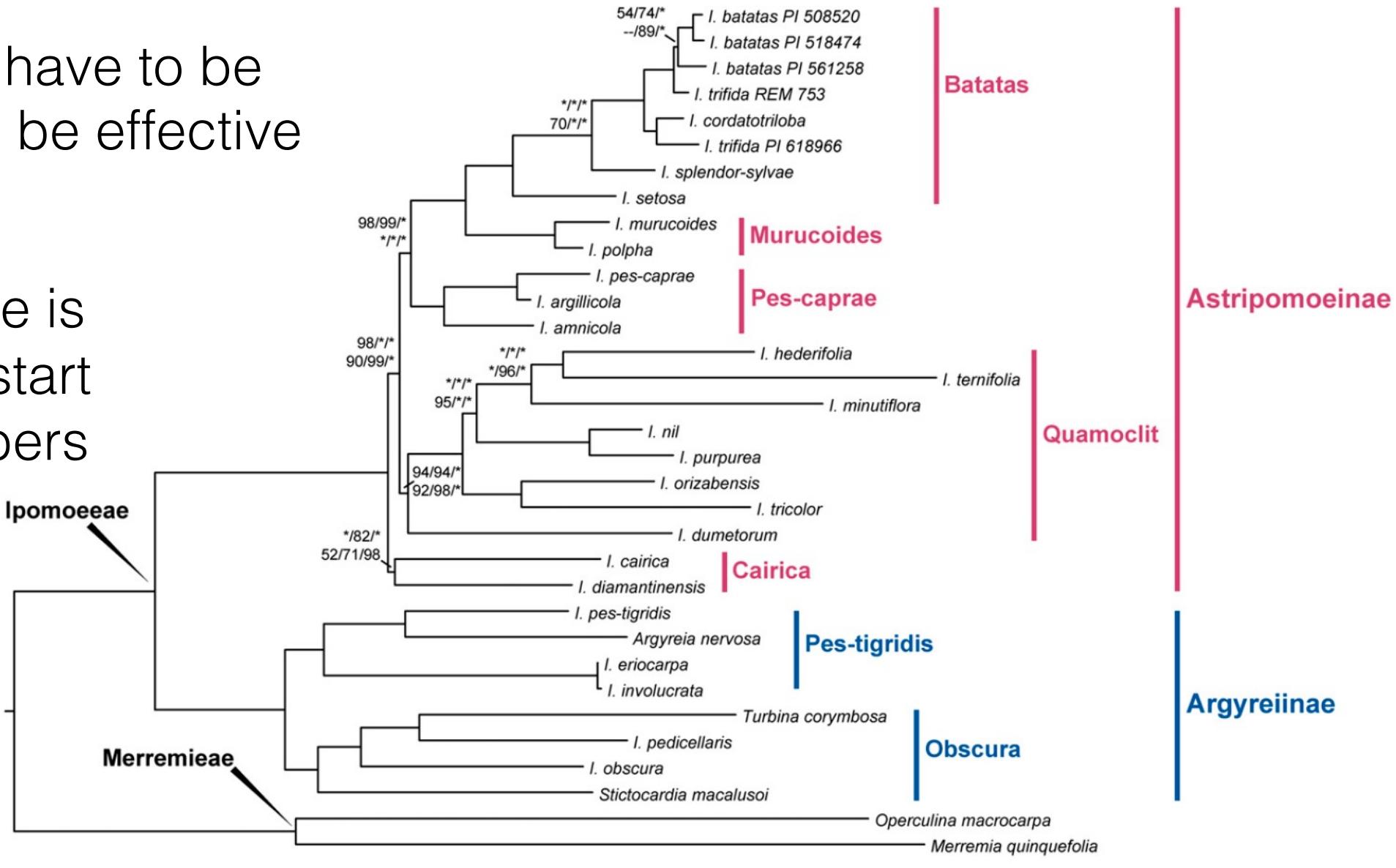
A “*” usually indicates a node with 100% bootstrap support



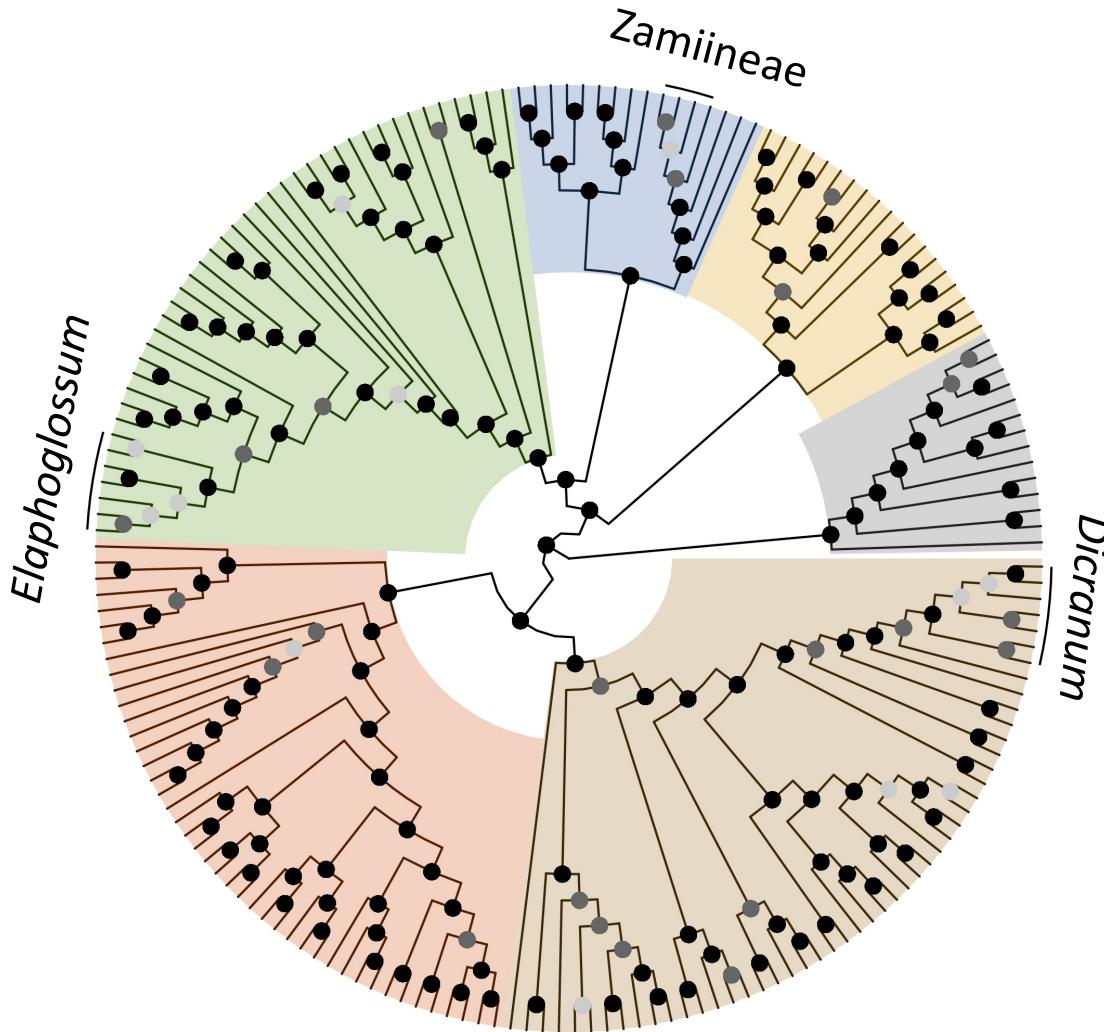
Visualization examples

Figures do not have to be complicated to be effective

If a support value is not perfect, we start writing the numbers



Visualization examples

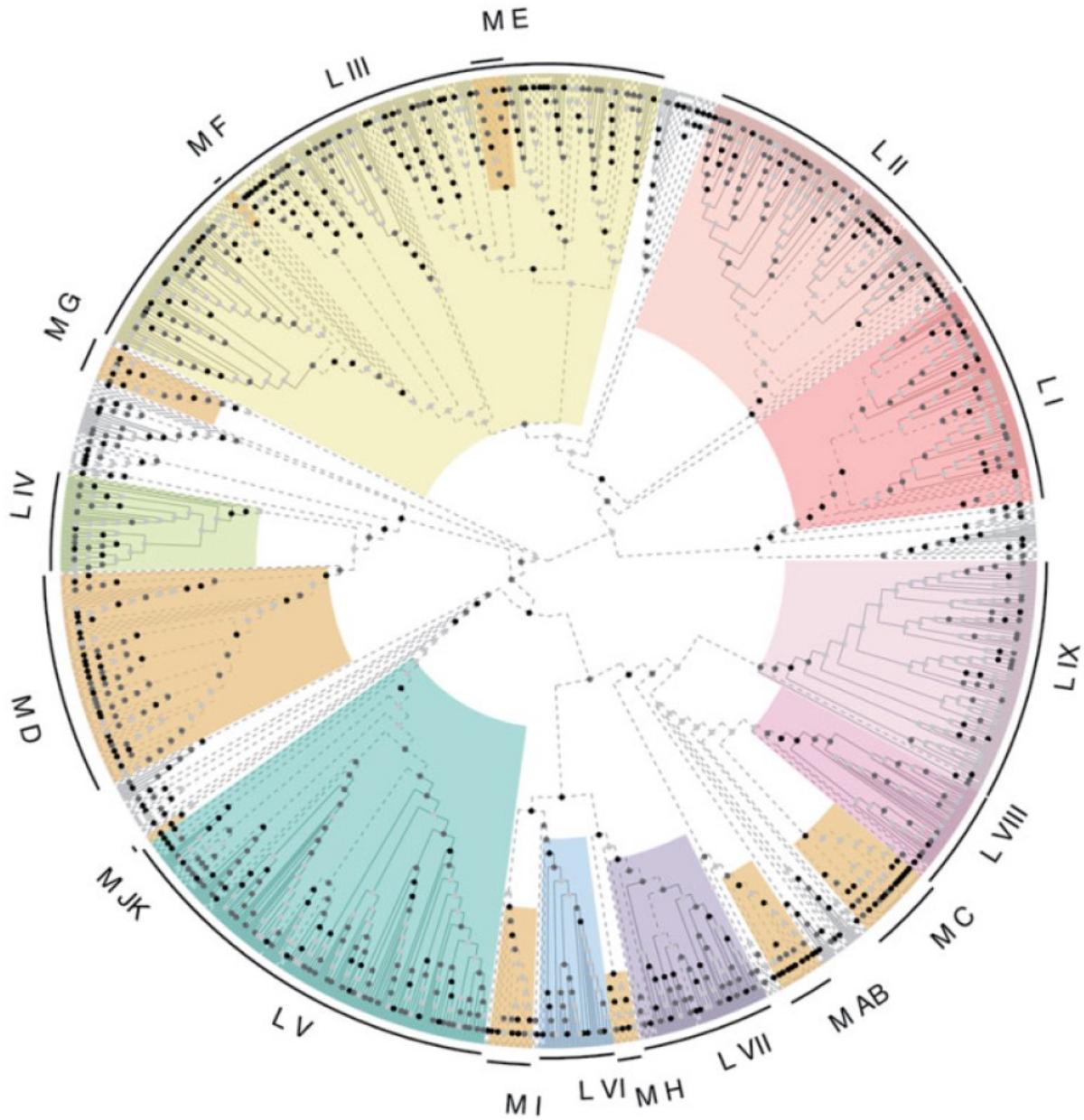


As trees get large, you might have to get more creative with how information is displayed

Bootstrap Support

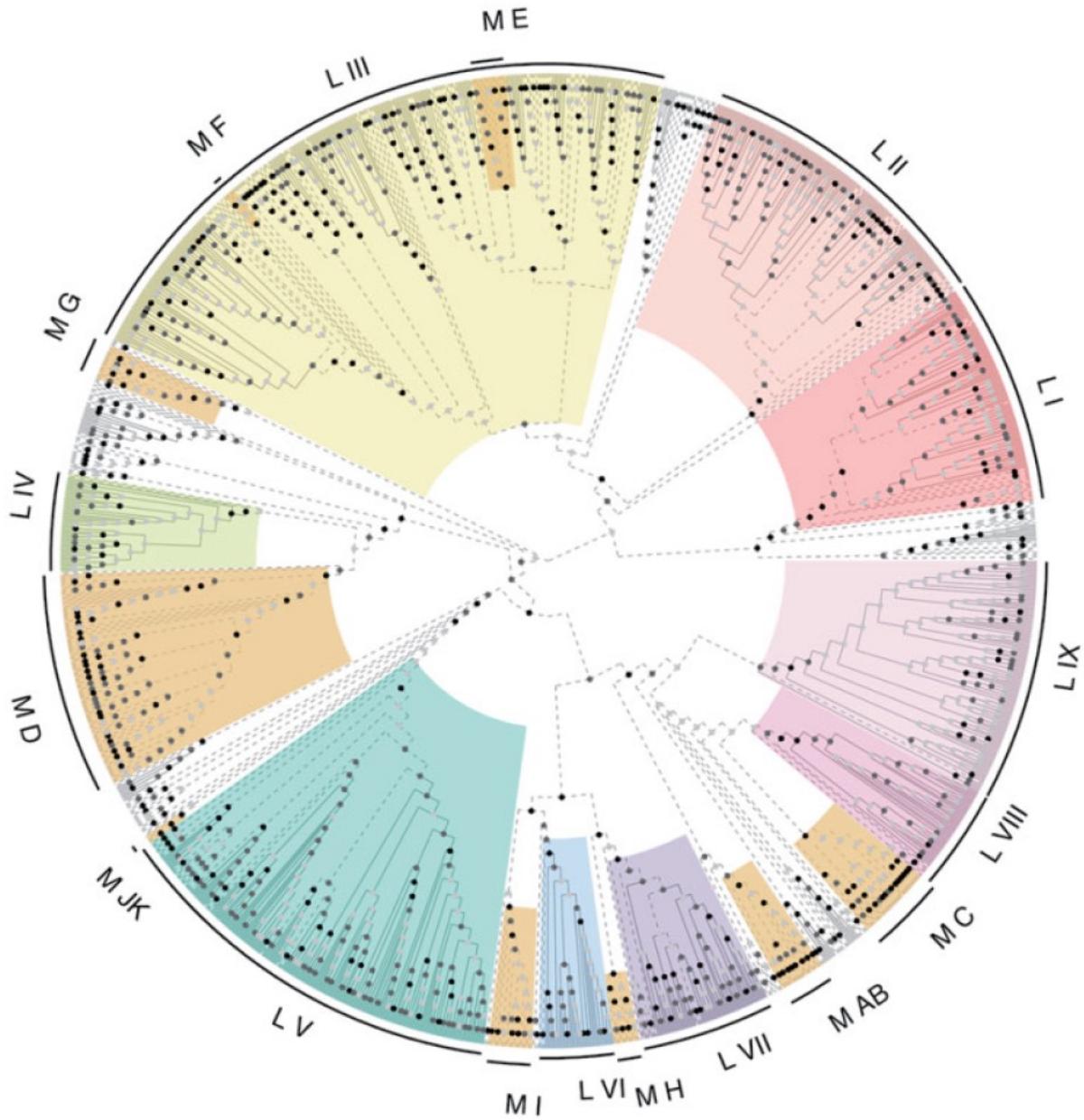
- 100%
- 70%-100%
- < 70%

Visualization examples



As trees get large, you might have to get more creative with how information is displayed

Visualization examples



As trees get large, you might have to get more creative with how information is displayed

Learning Goals

Newick: The tree file format

Bootstrapping

Other support values

Visualization examples

Visualization tools

Visualization tools

Figtree

The screenshot shows a GitHub repository page for `rambaut/figtree`. The repository is public and has 16 watchers, 77 forks, and 272 stars. It contains 3 branches and 7 tags. The `Code` tab is selected. The repository was last updated on Dec 12, 2021, with 304 commits. The commit history includes several pull requests and updates to the `Graphics`, `WebRoot`, `doc`, and `examples` directories.

rambaut / figtree Public

Code Issues 78 Pull requests 2 Actions Projects Security Insights

master 3 branches 7 tags

Go to file Add file Code

rambaut Merge pull request #175 from jonchang/java-8 ... f9085b6 on Dec 12, 2021 304 commits

File	Commit Message	Time
Graphics	Mac/Linux/Windows packaging now done by ANT. Adding 'Color by' ...	13 years ago
WebRoot	add jwt version of figtree (including war task to create a war file)	12 years ago
doc	Update large_trees.md	5 years ago
examples	Added example trees	14 years ago

About

Automatically exported from code.google.com/p/figtree

Readme 272 stars 16 watching 77 forks

Visualization tools

Otherwise, need to leverage tools in R

- “ape” package
- “phytools” package
- “ggtree” package

Visualization tools

Otherwise, need to leverage tools in R

- “ape” package
- “phytools” package
- “ggtree” package

Visualization tools

Otherwise, need to leverage:

- “ape” package
- “phytools” package
- “ggtree” package



End

We will finish the day by visualizing and making figures with our estimated phylogenies.

The number of computers may be limited