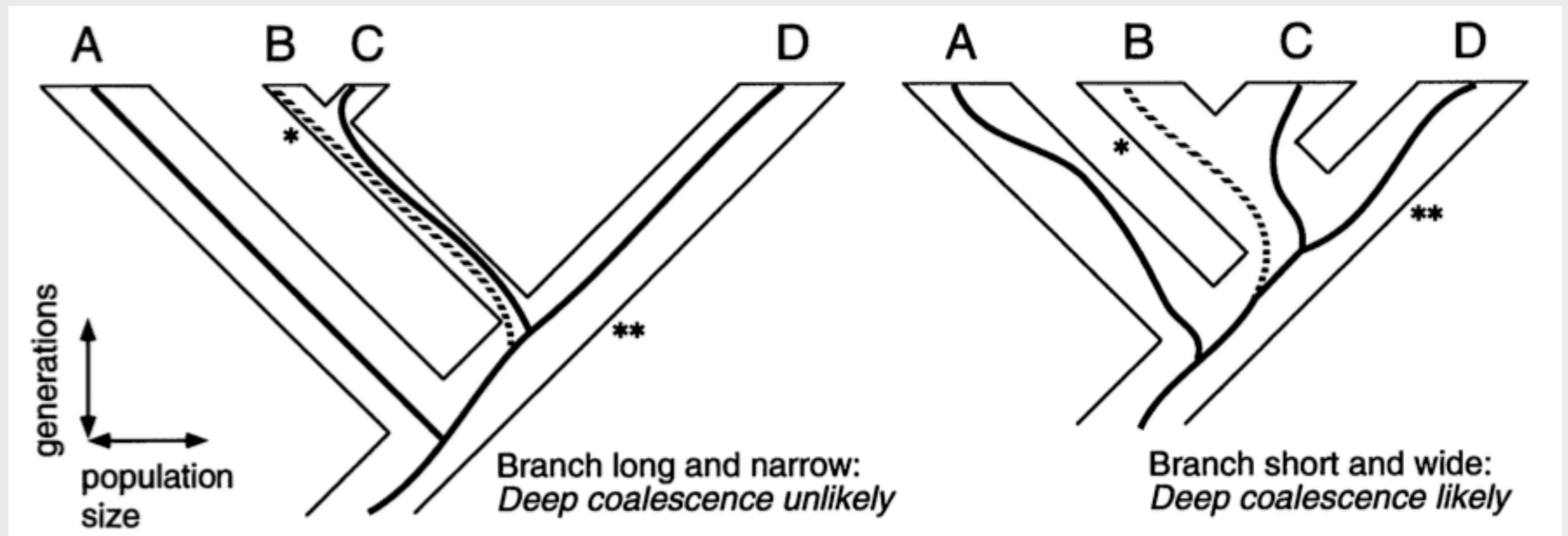


Phylogenetics 5

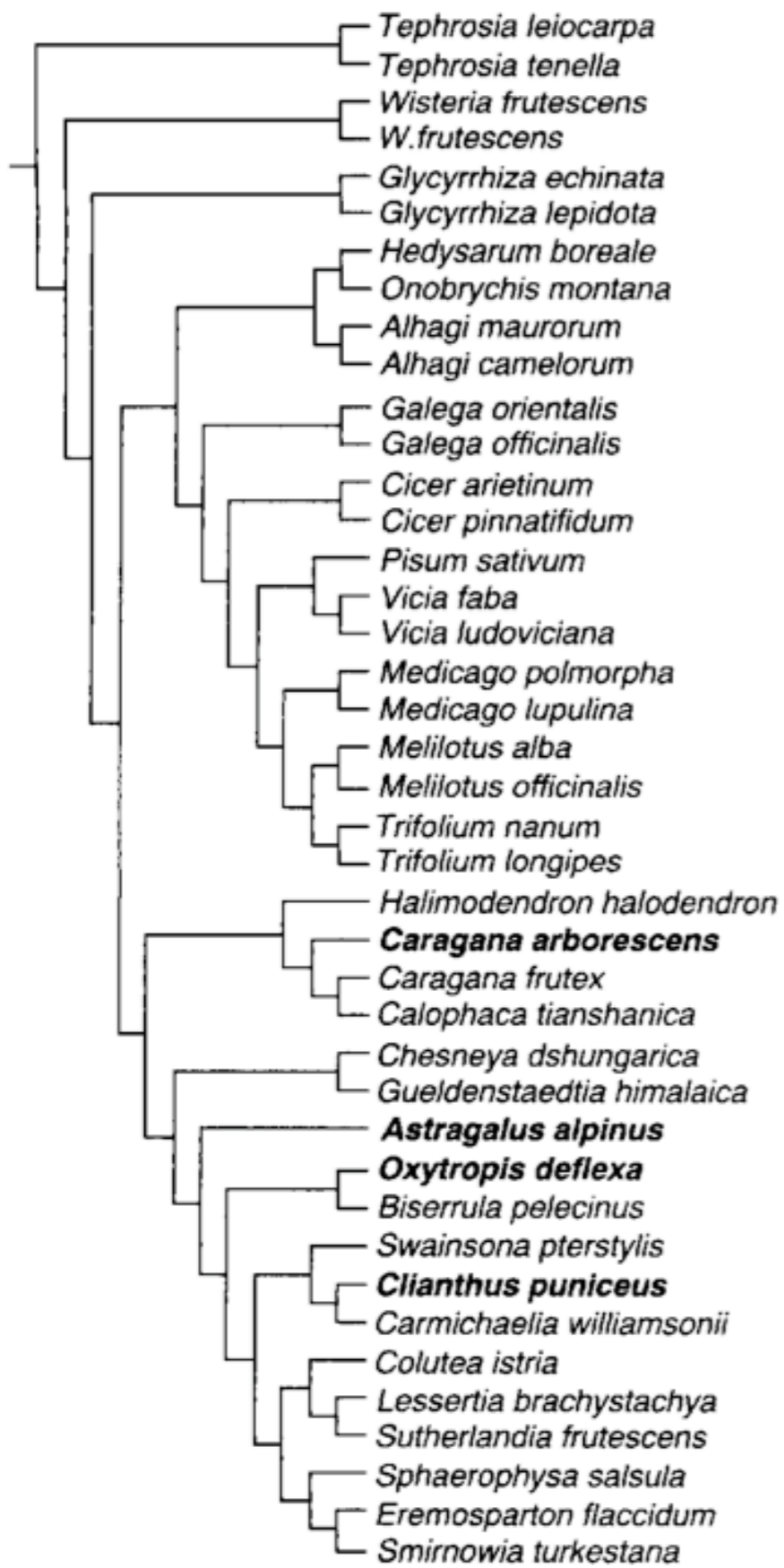


Brian O'Meara

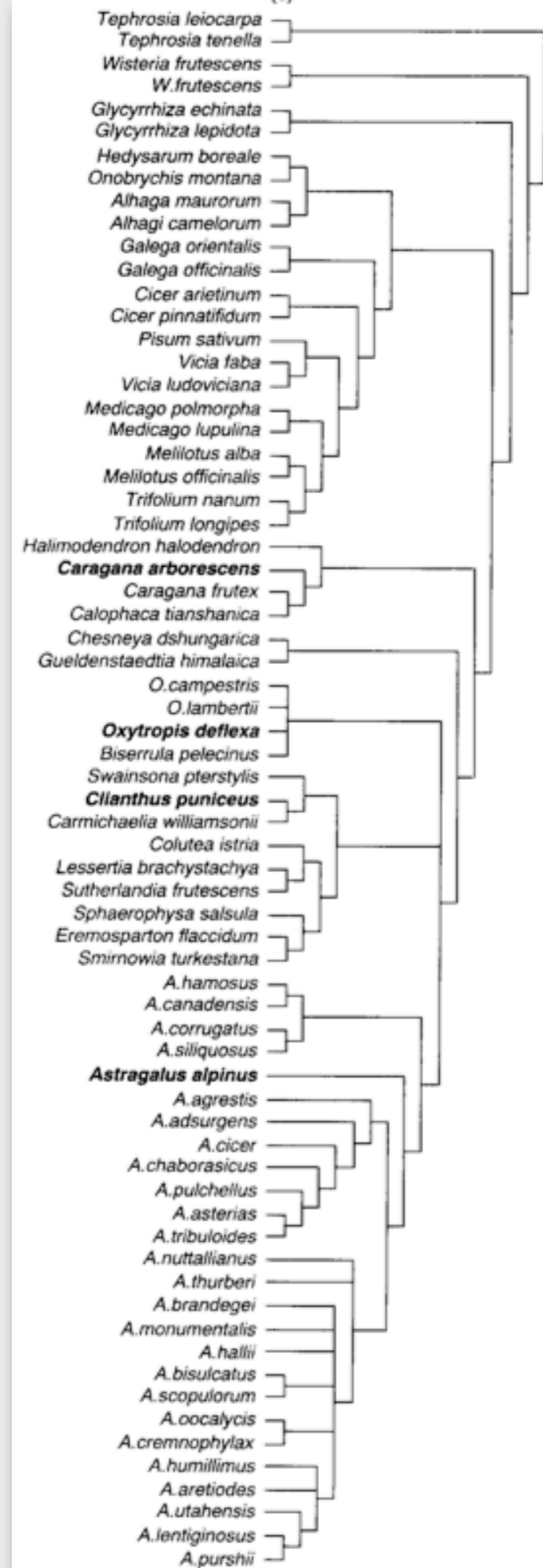
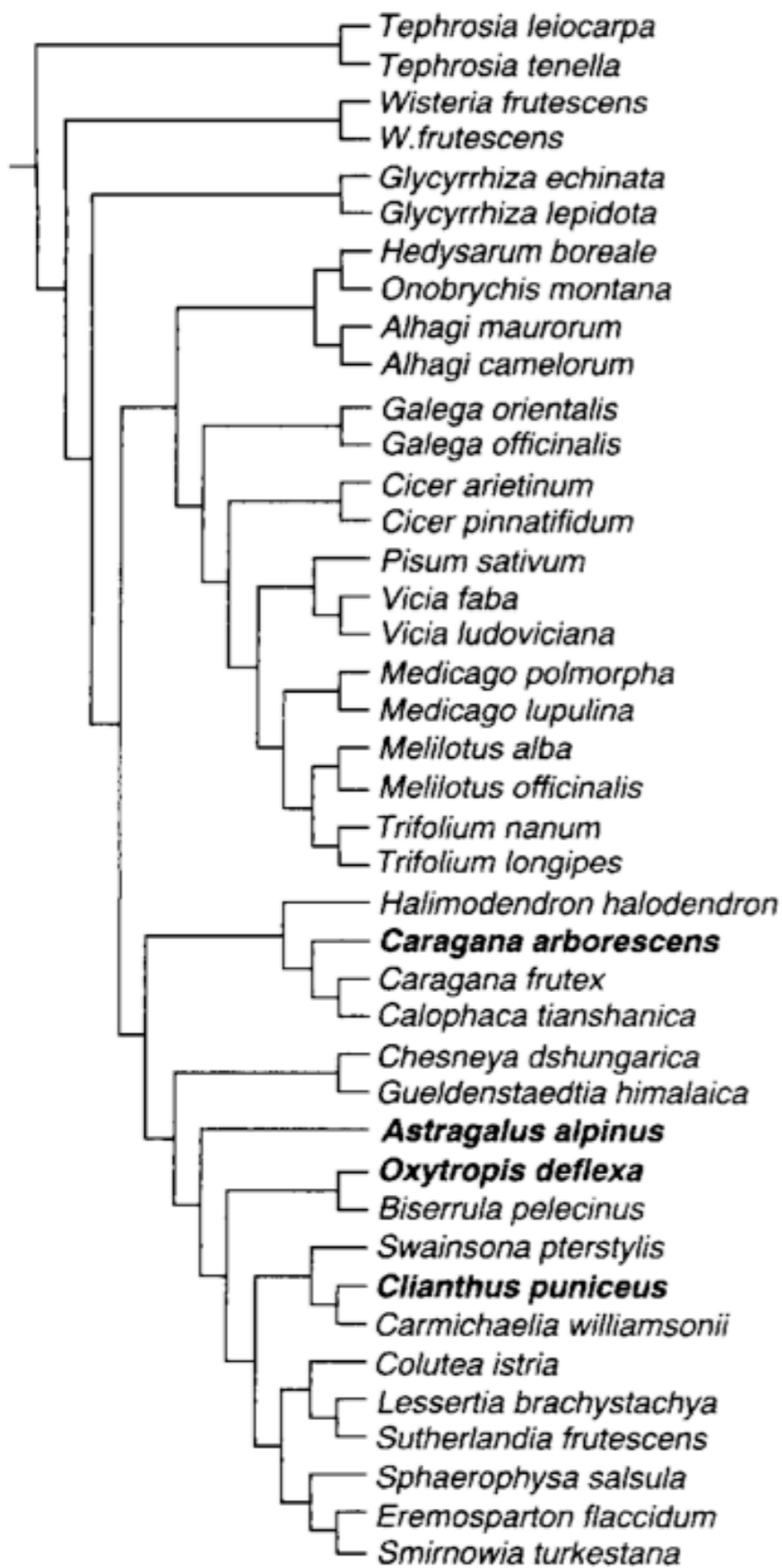
<http://www.brianomeara.info>

UT Knoxville EEB Graduate Core class, Fall 2010

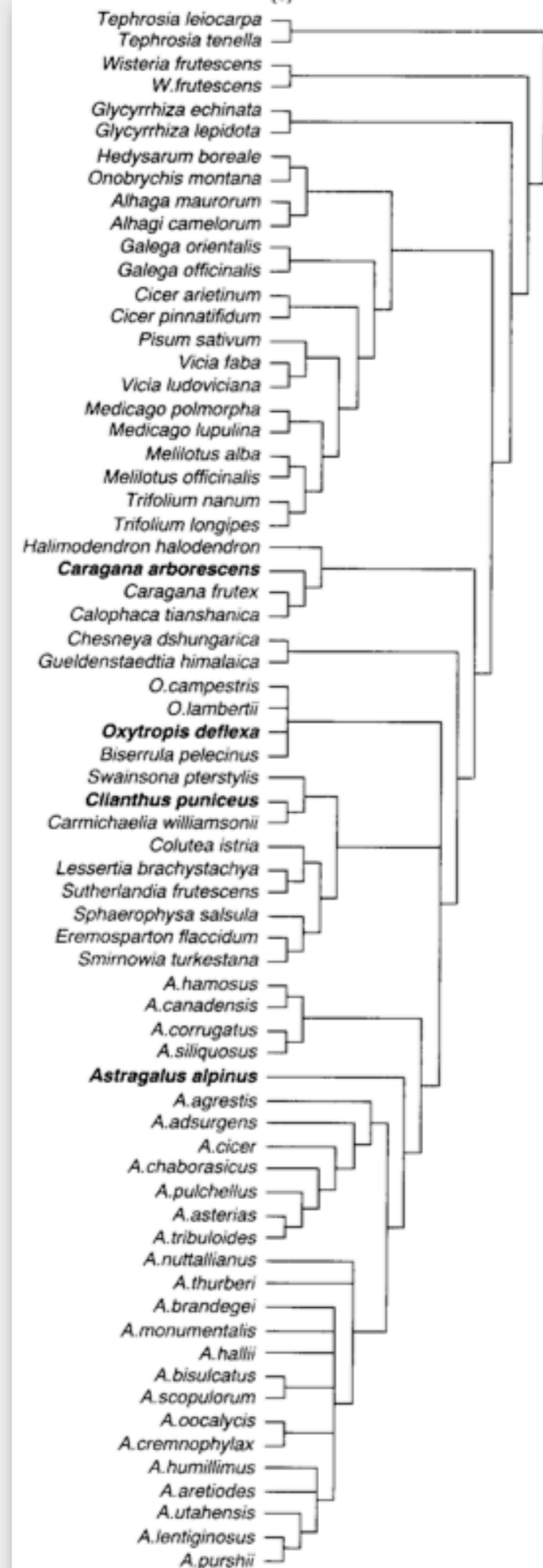
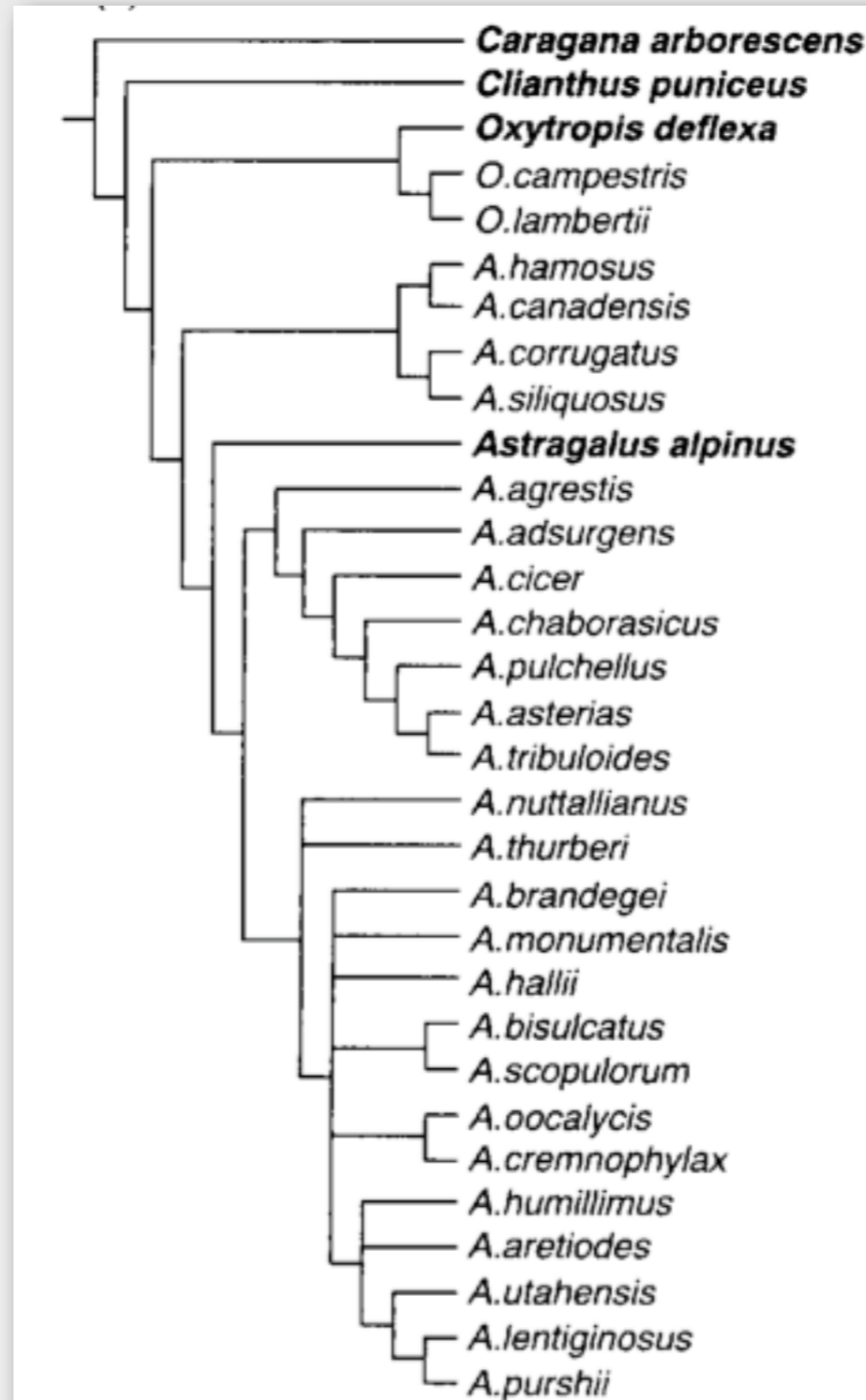
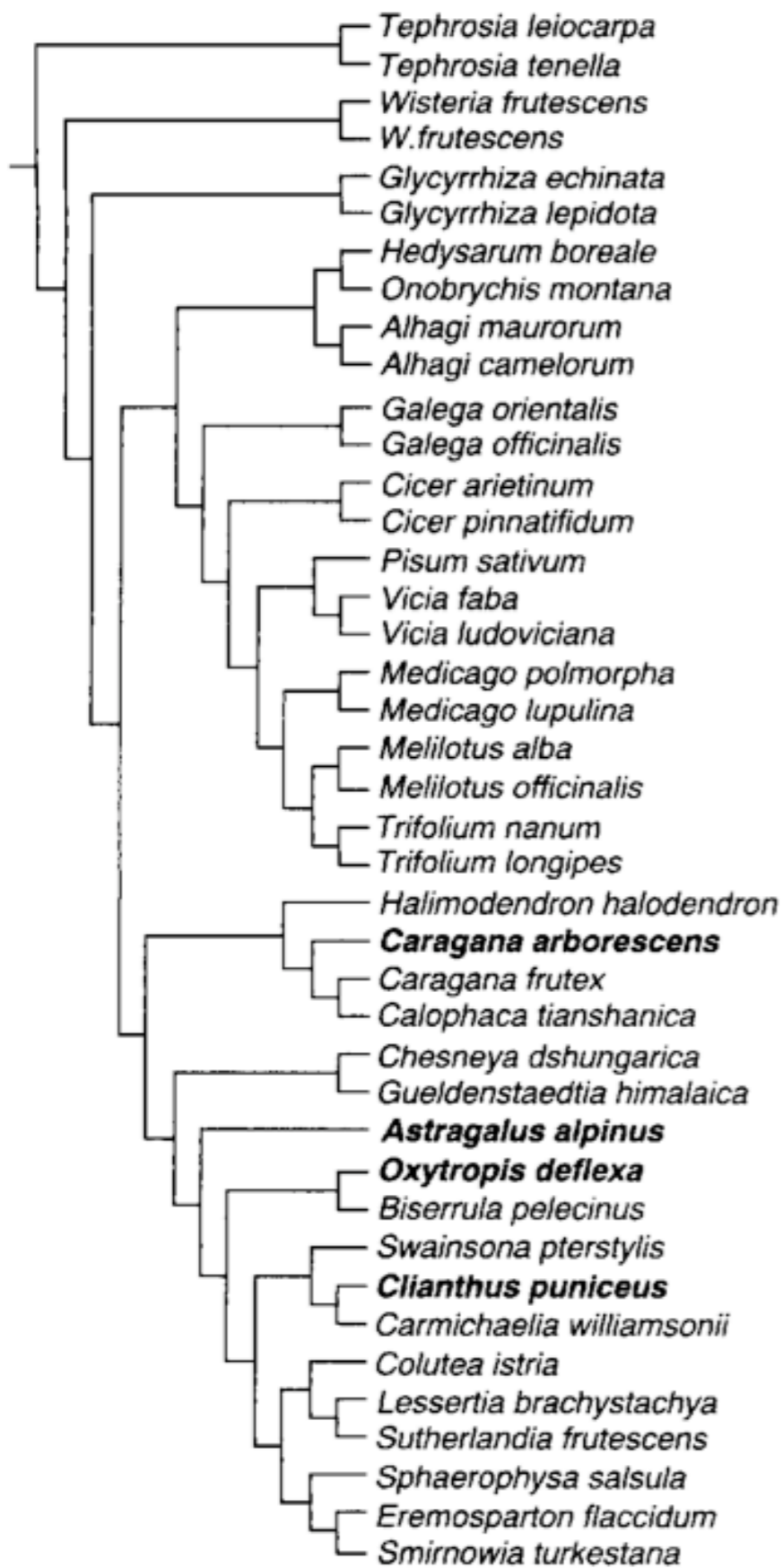
- Supertrees & Supermatrices
- Gene tree species tree
- Where to find trees

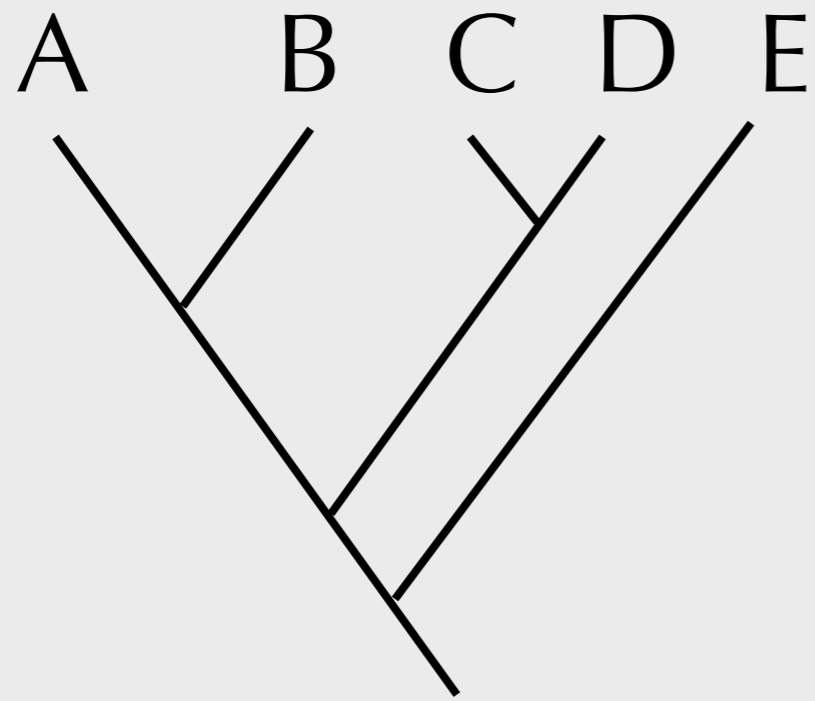


Sanderson et al. Phylogenetic supertrees: Assembling the trees of life.
 Trends Ecol Evol (1998) vol. 13 (3) pp. 105-109

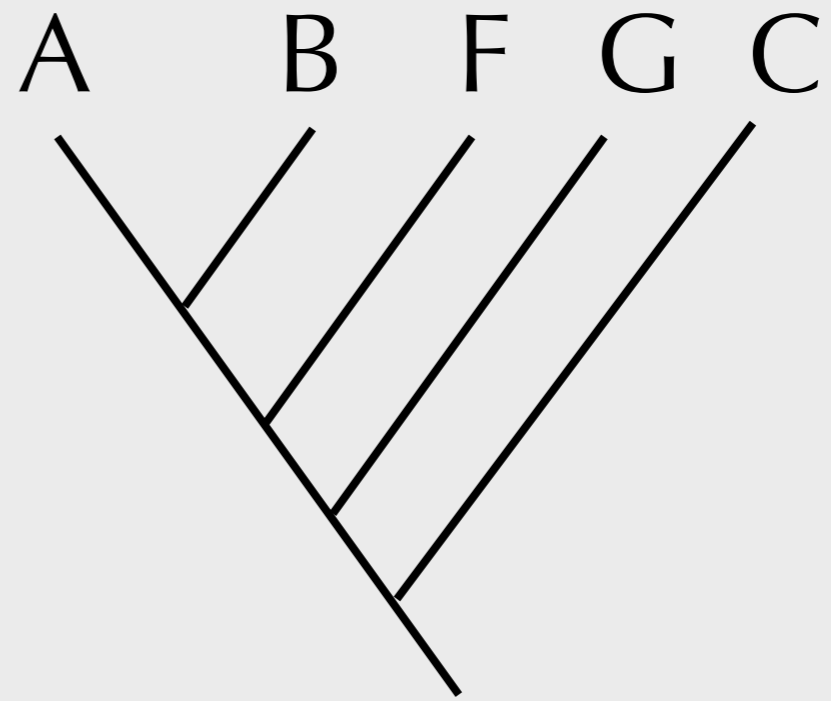


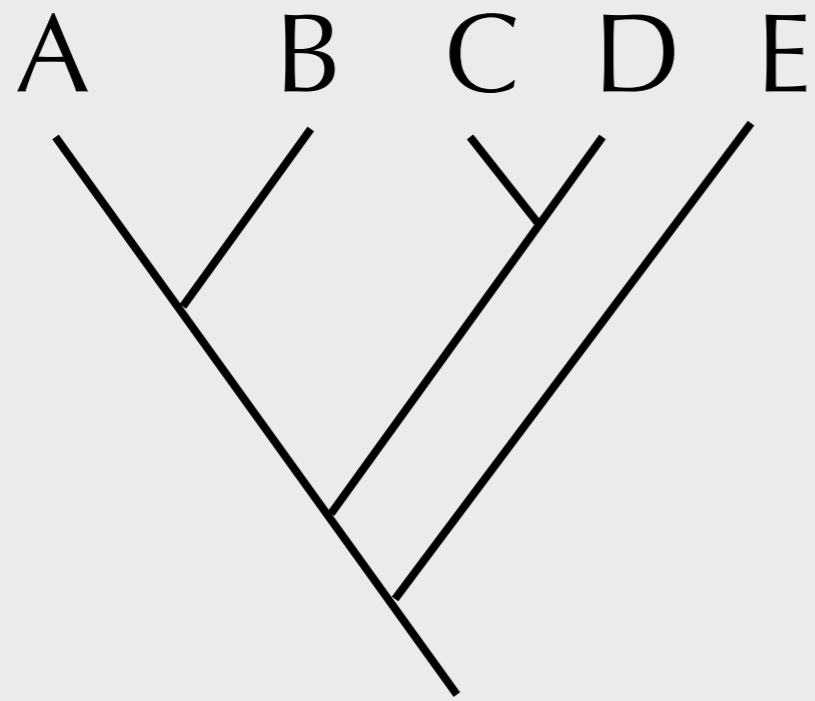
Sanderson et al. Phylogenetic supertrees: Assembling the trees of life. Trends Ecol Evol (1998) vol. 13 (3) pp. 105-109



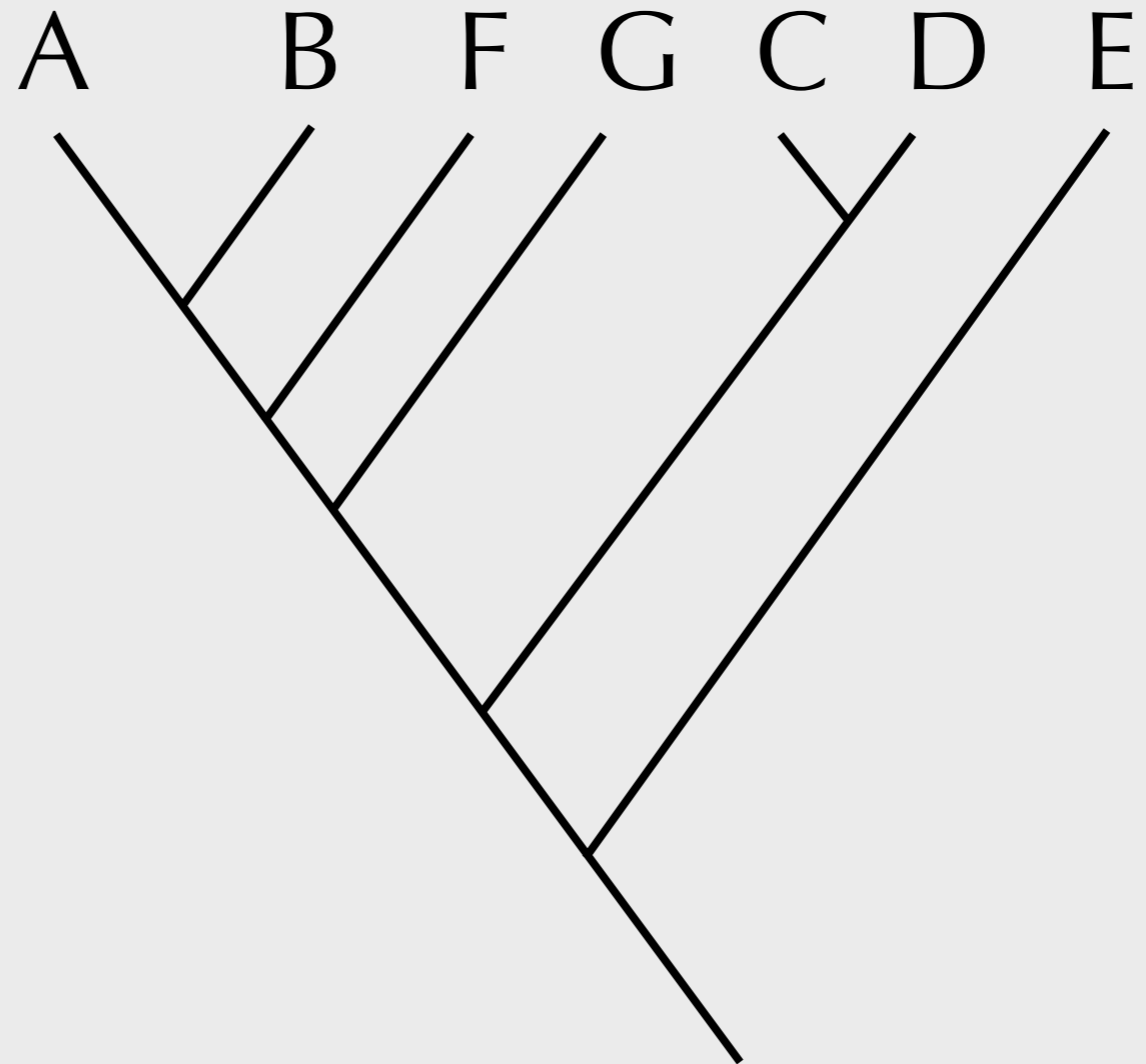
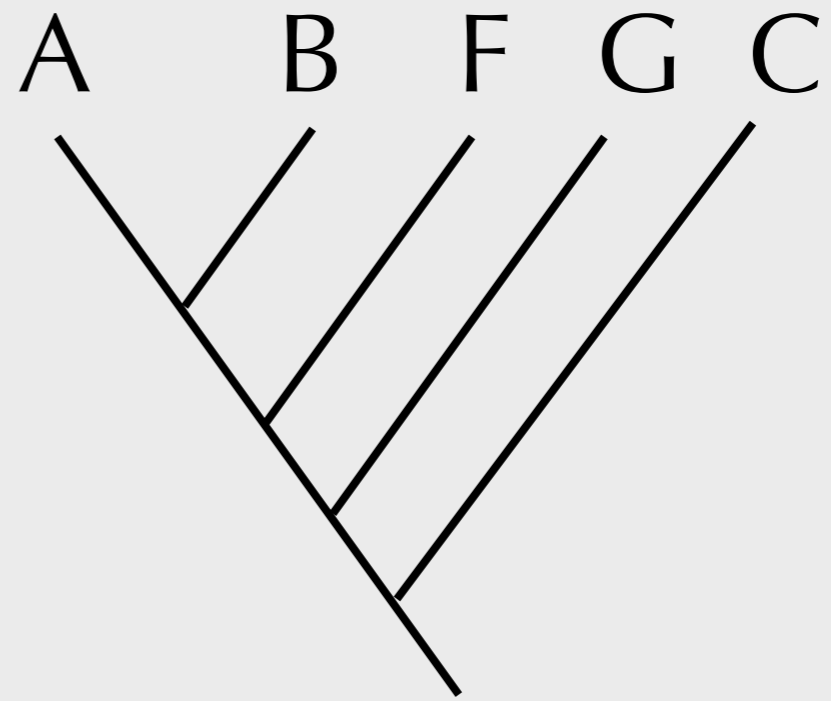


+





+



A B C D E

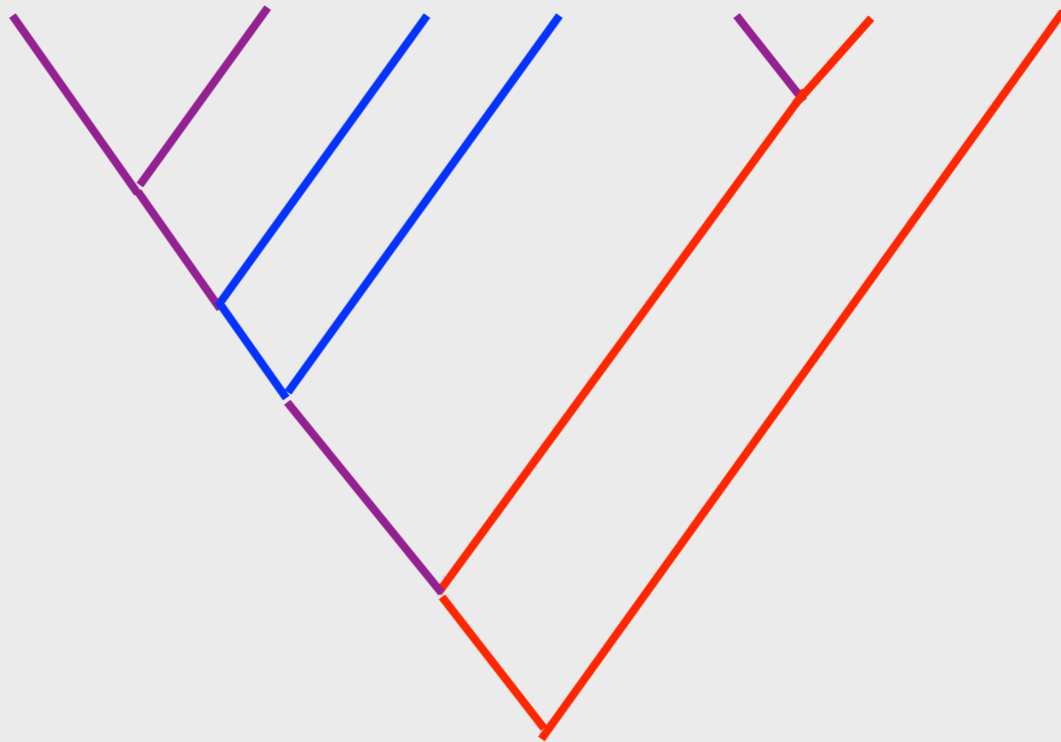


+

A B F G C

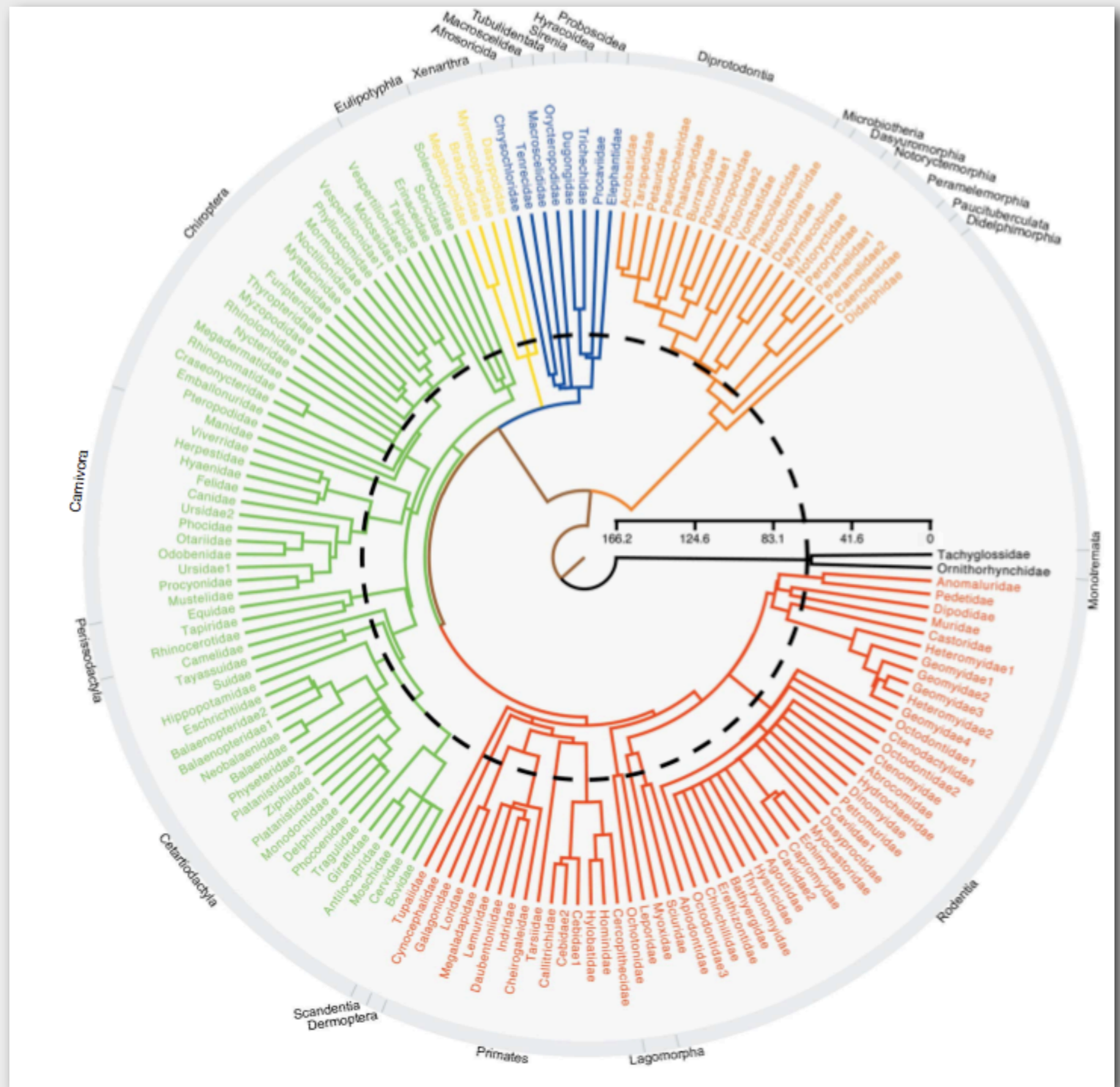


A B F G C D E

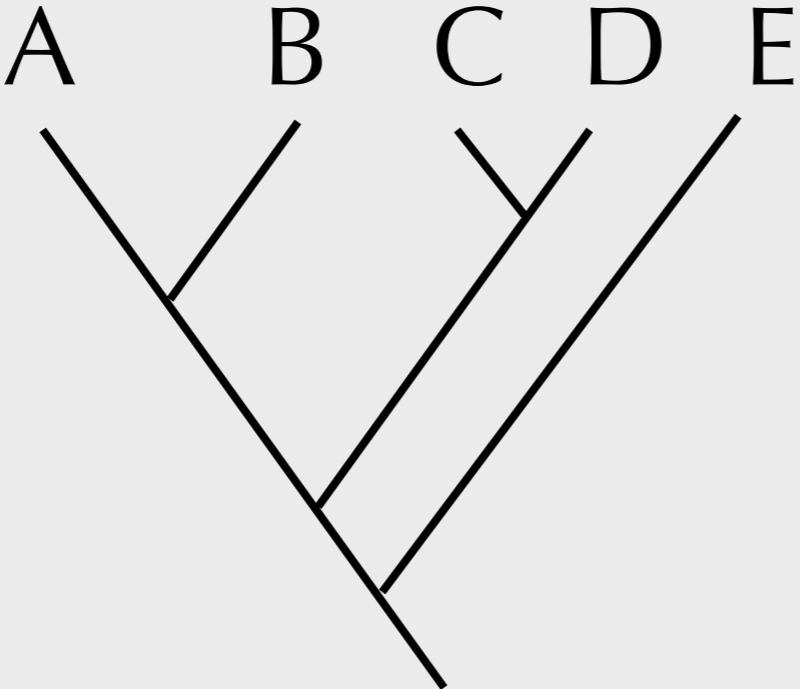


Supertree: combine two or more trees
with sets of taxa that do not overlap
completely

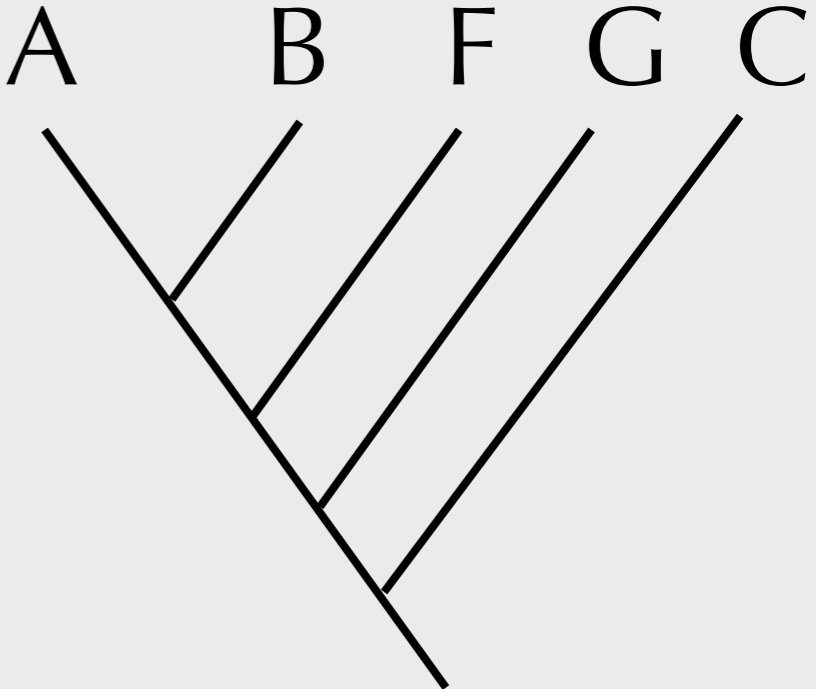
99% of mammal species
(summary tree shown)



Matrix representation with parsimony (MRP)

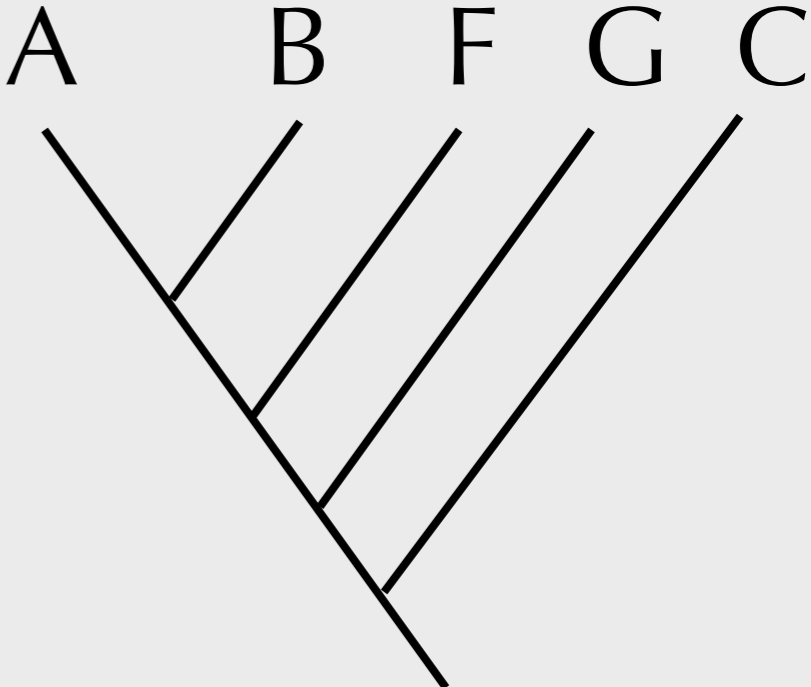
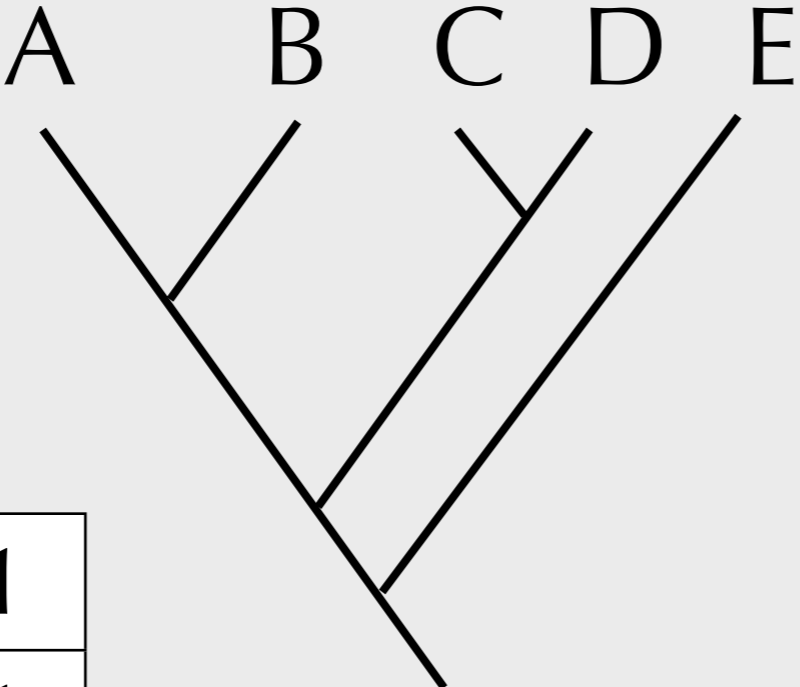


A	1	0	1
B	1	0	1
C	0	1	1
D	0	1	1
E	0	0	0



A	1	1	1
B	1	1	1
C	0	0	0
F	0	1	1
G	0	0	1

Matrix representation with parsimony (MRP)

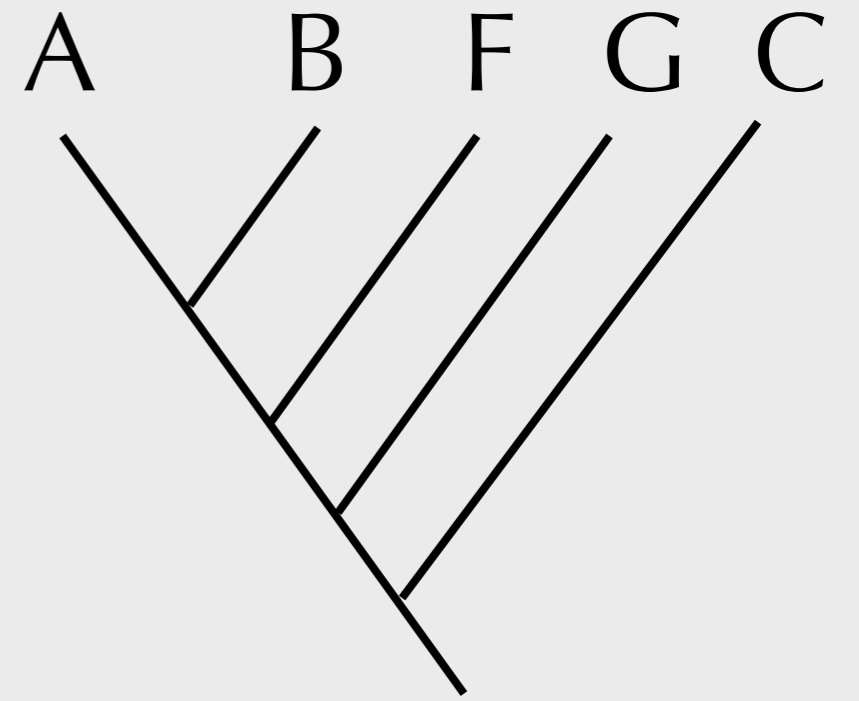
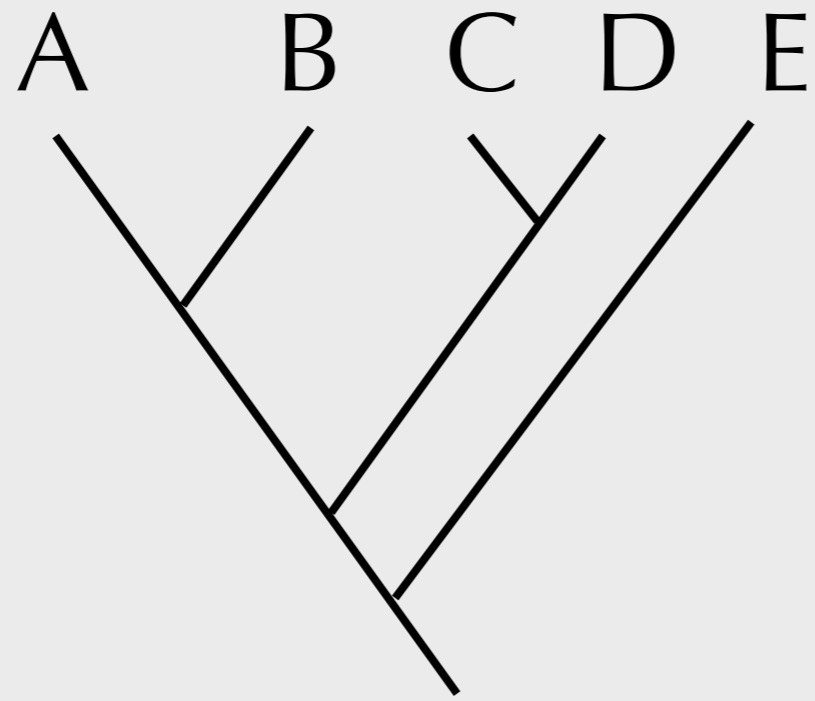


A	1	0	1
B	1	0	1
C	0	1	1
D	0	1	1
E	0	0	0

A	1	1	1
B	1	1	1
C	0	0	0
F	0	1	1
G	0	0	1

A	1	0	1	1	1	1
B	1	0	1	1	1	1
C	0	1	1	0	0	0
D	0	1	1	?	?	?
E	0	0	0	?	?	?
F	?	?	?	0	1	1
G	?	?	?	0	0	1

Supermatrix



A	A	A	T	C	G	C	G	G	C	A	G	C	A	A	T	A
B	A	T	T	C	G	C	G	G	C	A	G	C	A	A	T	A
C	A	A	T	T	G	C	G	G	C	A	G	C	G	T	T	T
D	A	A	T	T	G	C	G	G	?	?	?	?	?	?	?	?
E	A	A	T	C	G	C	A	G	?	?	?	?	?	?	?	?
F	?	?	?	?	?	?	?	?	C	T	G	C	G	A	T	A
G	?	?	?	?	?	?	?	?	C	T	G	C	G	A	T	A

Supertree

Advantages

- Makes search for large trees faster (builds on previous searches)
- Can combine results from multiple kinds of data

Disadvantages

- Data redundancy (use same dataset for multiple trees)
- Combining raw data has advantages
- Biases in reconstruction (tree shape can matter)

- ~~Supertrees & Supermatrices~~
- Gene tree species tree
- Where to find trees

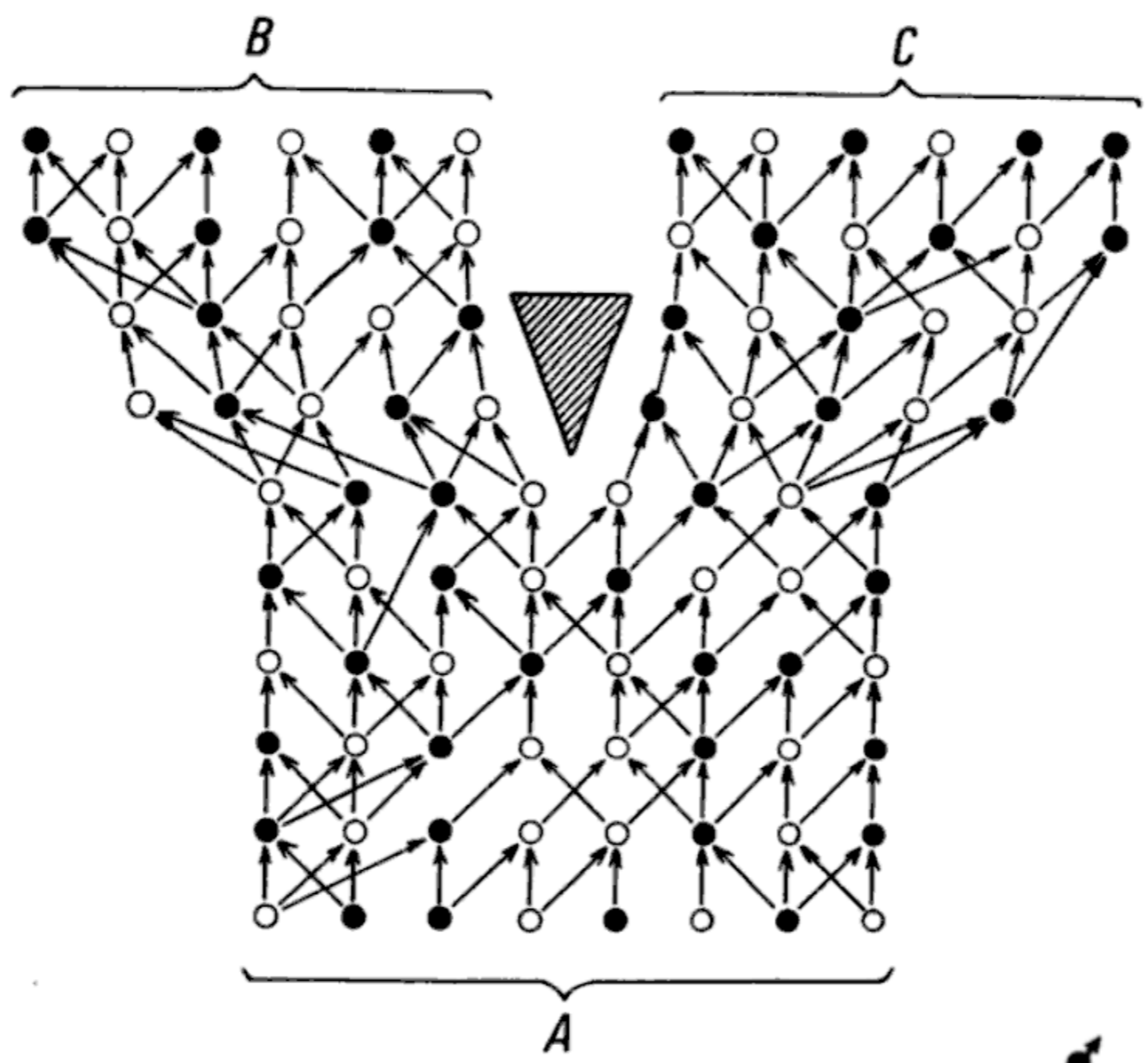
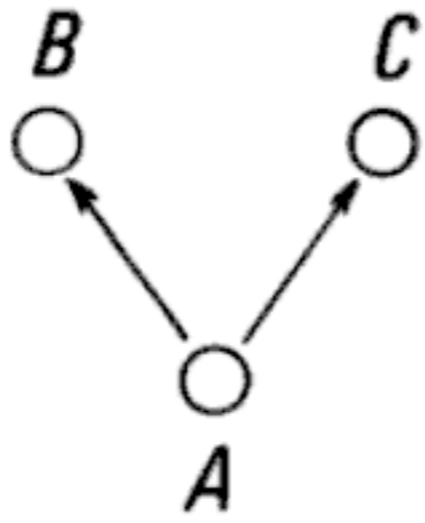


Figure 4. The process of species cleavage.

♂ ♀

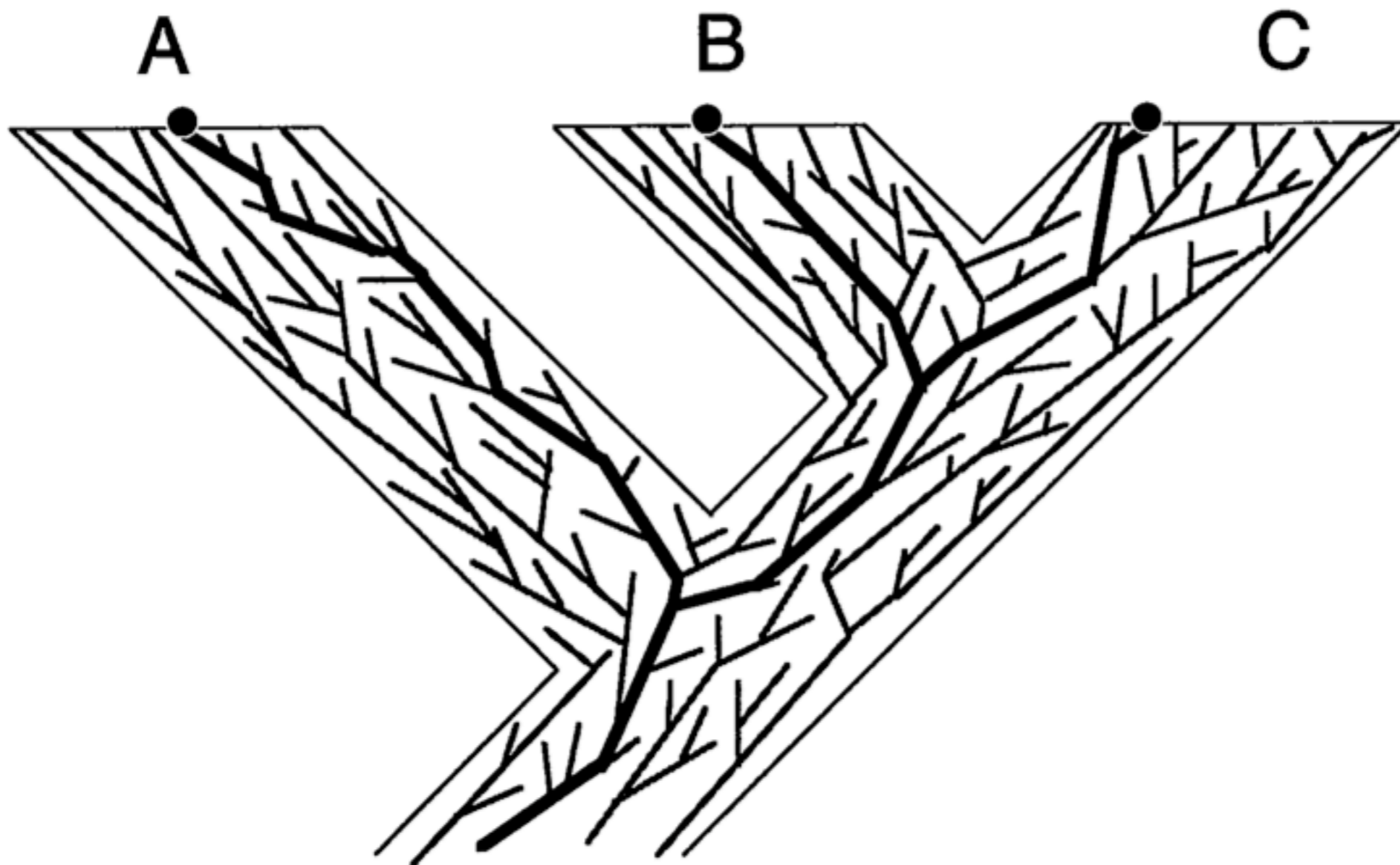


FIGURE 1. A gene tree contained within a species tree leading to three extant species: A, B, and C. Bold branches of gene tree show relationships among the sampled copies of the gene (●). Sampled copies from sister species B and C are sister copies.

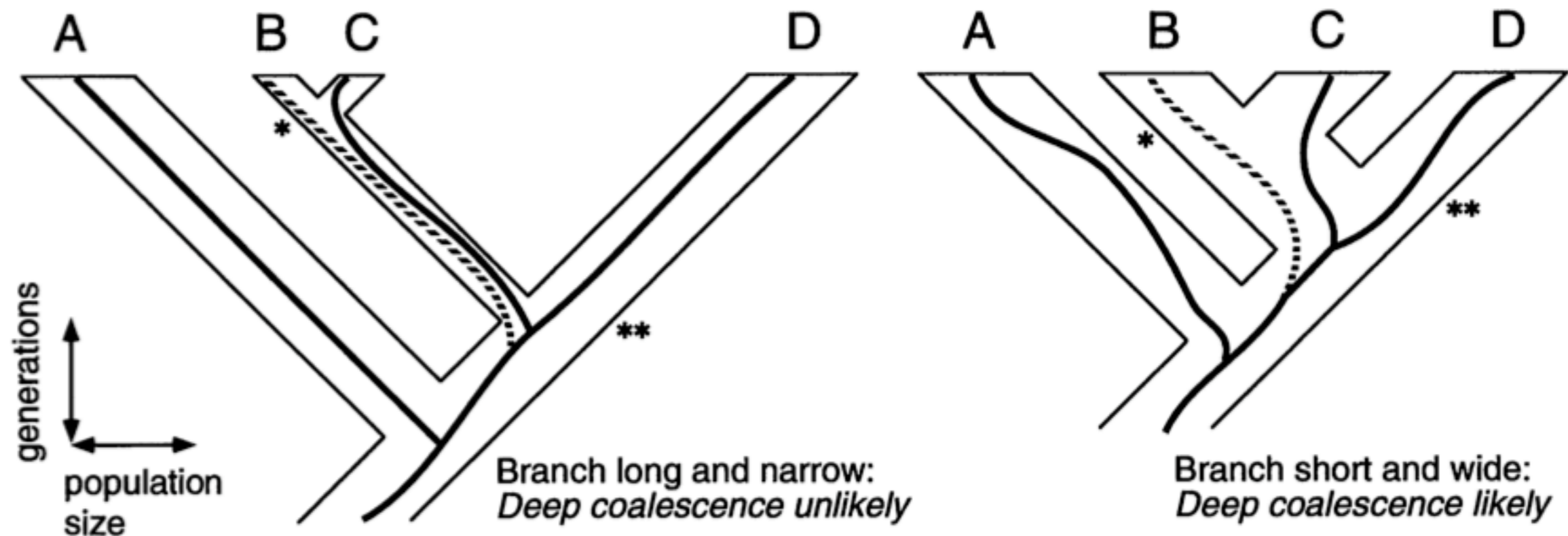


FIGURE 4. Lineage sorting (deep coalescence). Described in a time-forward sense as lineage sorting, an ancestral polymorphism at ** is retained through a lineage to the next speciation event at *, where different forms are sampled in different descendant species. Described in a time-backward sense as deep coalescence, two gene copies from species B and C meet at * but fail to coalesce until deeper than the speciation event at **, at which point the gene from C coalesces first with the gene from D. Failure to coalesce is more likely the shorter (in generations) and wider (in effective population size) the branch is between ** and *.

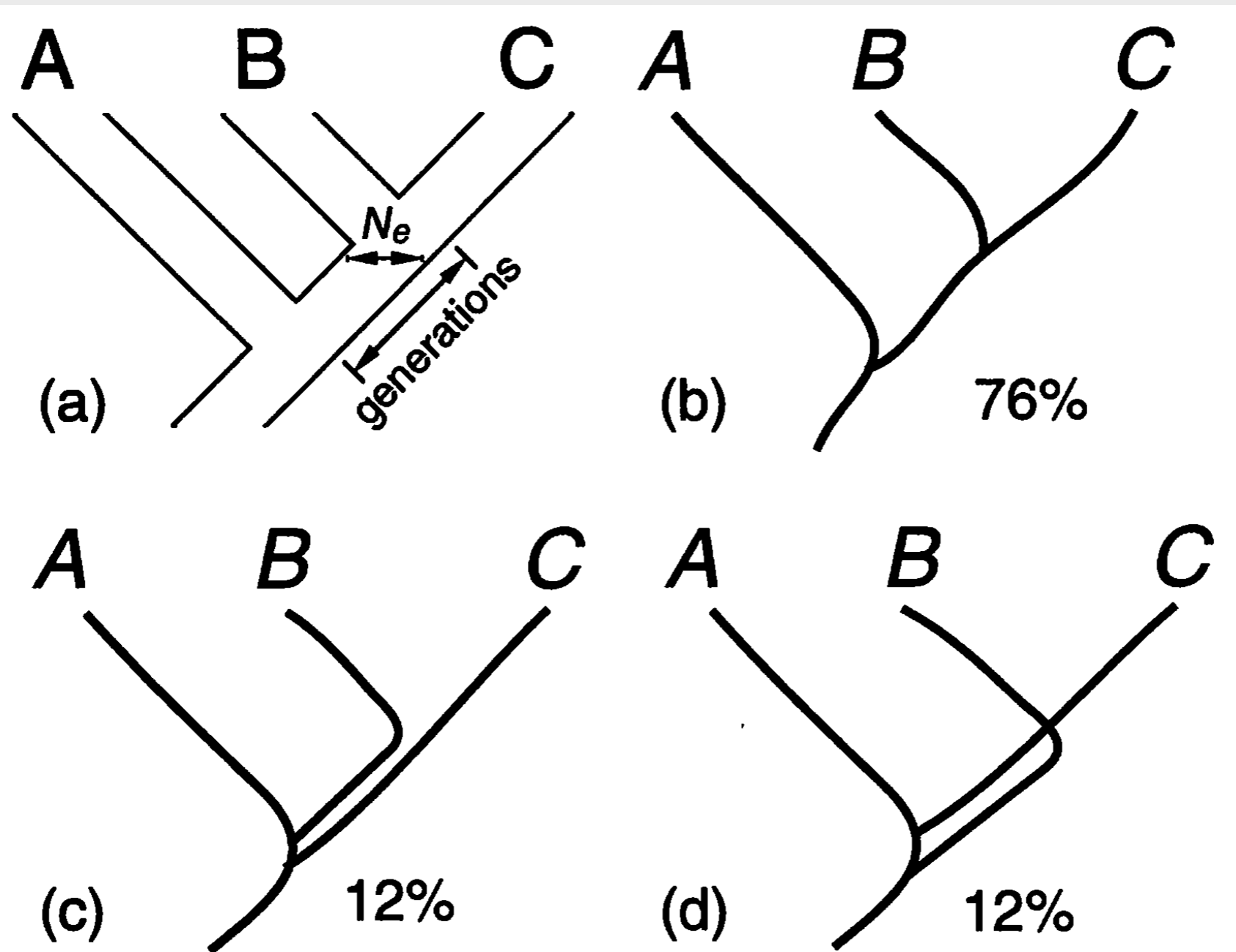


FIGURE 8. A tree of the successive splitting of a population (a) and three sampled gene trees (b–d). If the length (in generations) of the labeled branch is about twice the width of the branch (N_e = effective population size), then the gene trees should occur with the frequencies shown.

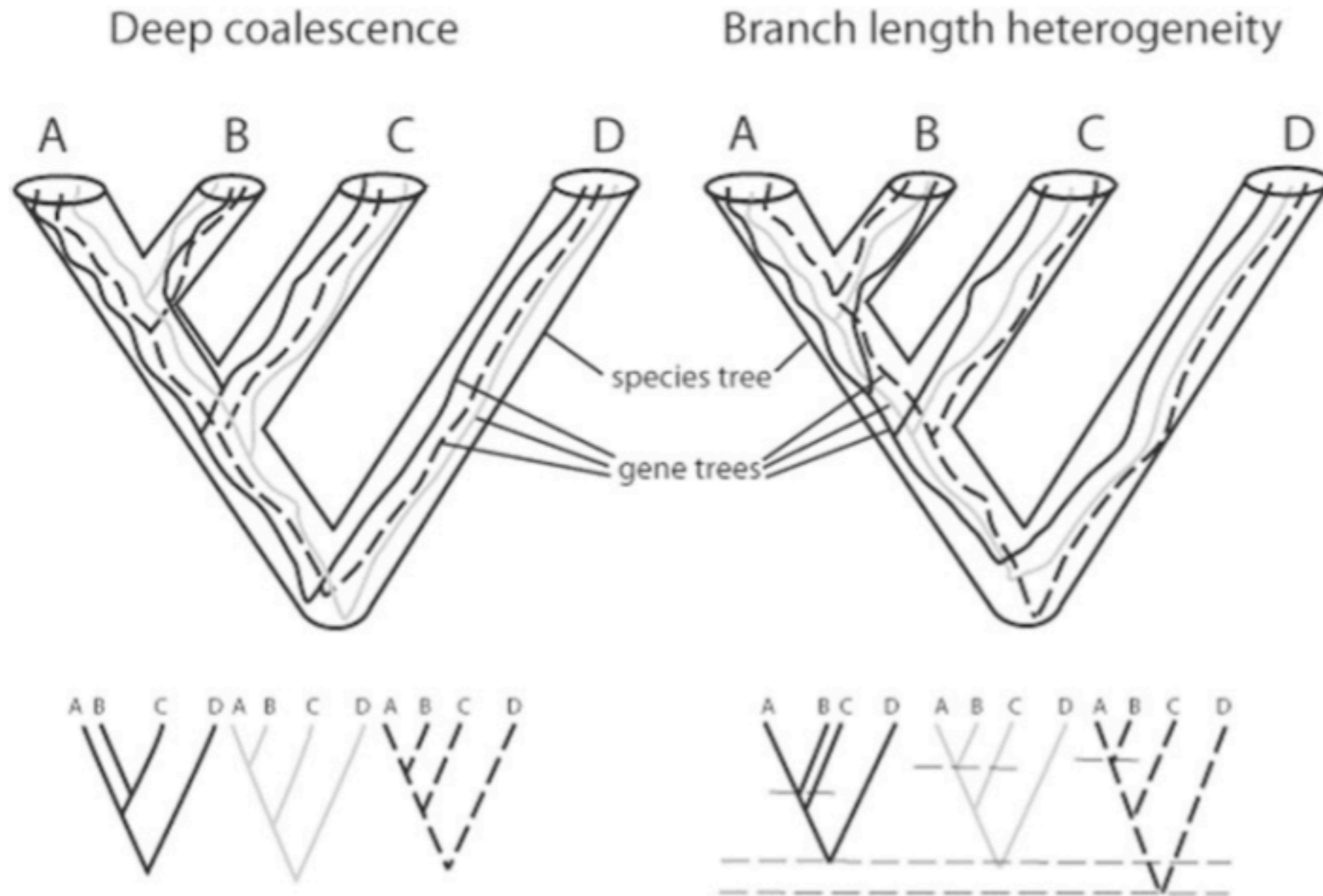


Figure 1. Distinction between deep coalescence and branch length heterogeneity as sources of gene tree heterogeneity and gene tree/species tree conflict. Example species trees are shown at the top, with constituent gene trees in the bottom row; taxa from which gene trees are sampled are given as A, B, C, and D. Whereas deep coalescence emphasizes topological differences between gene and species trees, branch length heterogeneity emphasizes branch length differences between gene and species trees and variation among genes in branch length, without topological variation. Branch length heterogeneity is ubiquitous and will be important for impacting site distributions in DNA sequences when effective population sizes and species tree branch lengths are large enough to permit substantial variation in coalescence times without deep coalescence. Heterogeneity in branch lengths among constituent gene trees is indicated by the dashed lines in the lower right panel.

- **Gene tree parsimony:** find species tree that minimizes number of duplications (Slowinski et al., 1997)
- **BEST:** find species tree with maximum posterior probability, under a model with DNA substitutions on gene trees and gene trees evolving within species trees (Liu & Pearl 2007, Edwards et al., 2007)
- **STEM:** find species tree that maximizes likelihood of input gene trees (inferred in another program) (Kubatko et al. 2009)
- ***BEAST:** multispecies coalescent (Heled & Rambaut, 2010)

- ~~Supertrees & Supermatrices~~
- ~~Gene tree species tree~~
- Where to find trees

- The literature: many trees published every day. Probably have to email the authors for the tree file.
- TreeBase.org: has many topologies (branch lengths prohibited (!)), also raw data
- mor (<http://mor.clarku.edu/>): continually-updated tree of fungi
- Phylomatic: hand-curated plant tree
- Phylota browser: has sets of aligned data for many groups, also trees from basic analyses
- Build your own: get data, run software. Or build supertree.
- Collaborate: lots of people (here and elsewhere) make trees. Find a buddy!

Resources for learning more

- Inferring Phylogenies by Joseph Felsenstein (half the field comes from his work)
- Journals: *Systematic Biology*, *Evolution*, *American Naturalist*, *Science*, *Nature*, *Molecular Phylogenetics & Evolution* all publish phylogenetic methods and trees (other places, too, of course)
- Other courses at UTK (and perhaps a reading group).
- Bodega Bay Workshop in Applied Phylogenetics (in CA, I often teach there), Woods Hole Workshop on Molecular Evolution (in MA)
- Felsenstein's list of phylogenetic software (<http://evolution.genetics.washington.edu/phylip/software.html>), my TreeTapper.org site