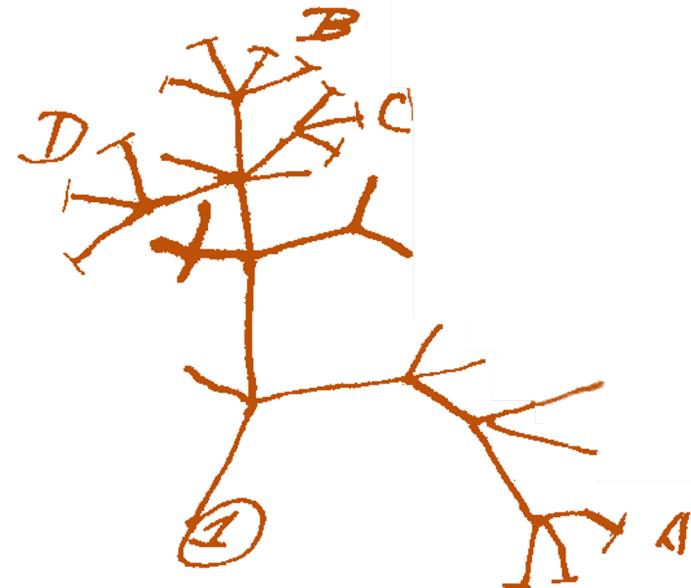


Phylogenetics: the science of evolutionary relationships

I think



BIOL 435/535: Bioinformatics
Feb 10, 2022

Classifying organisms: Pre-Darwin thought

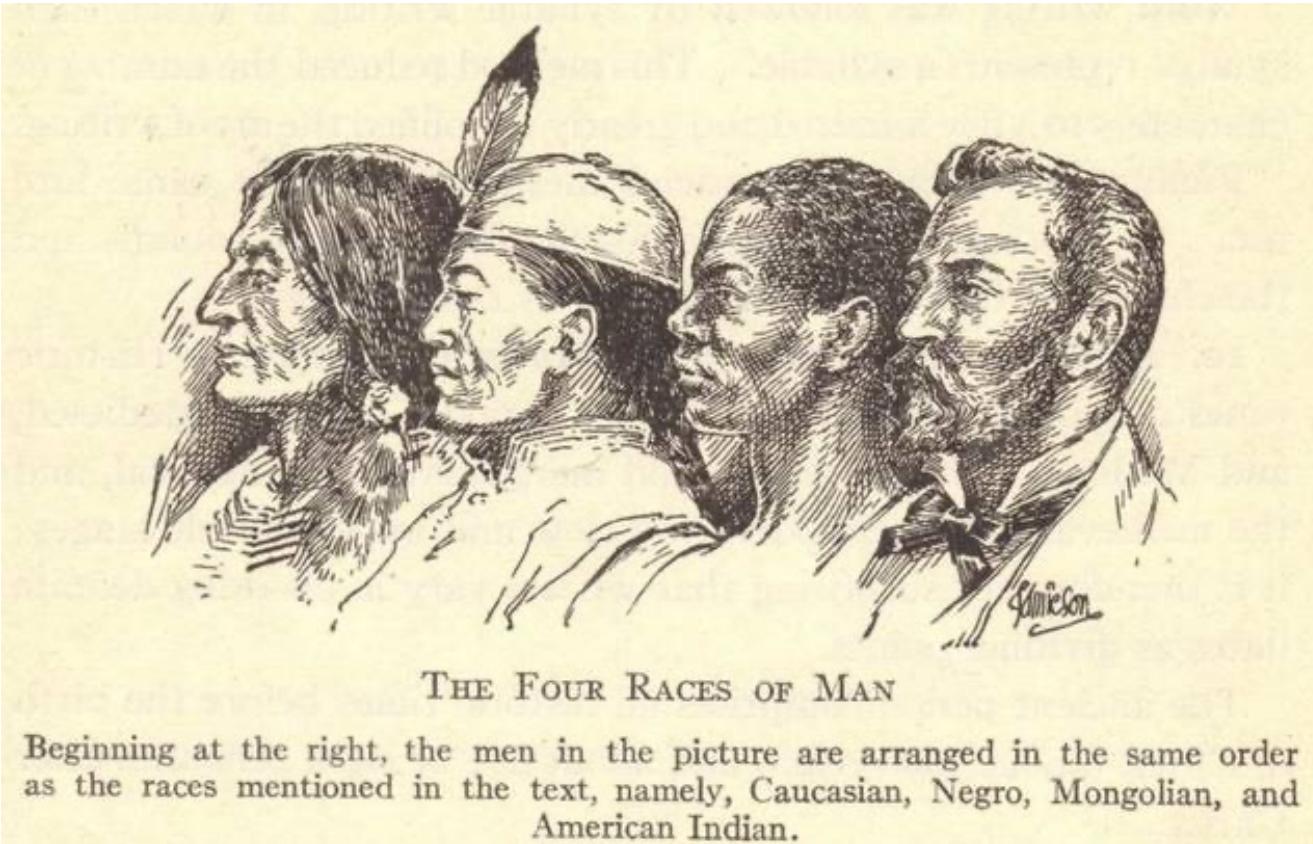


Peter Rabbit:
the Platonic cottontail

Plato – A species is defined by its ideal/perfect form, and variations observed in nature represent lower forms of existence

A form of **essentialism**, which is incompatible with evolution (i.e., change in form over time)

Classifying organisms: Pre-Darwin thought



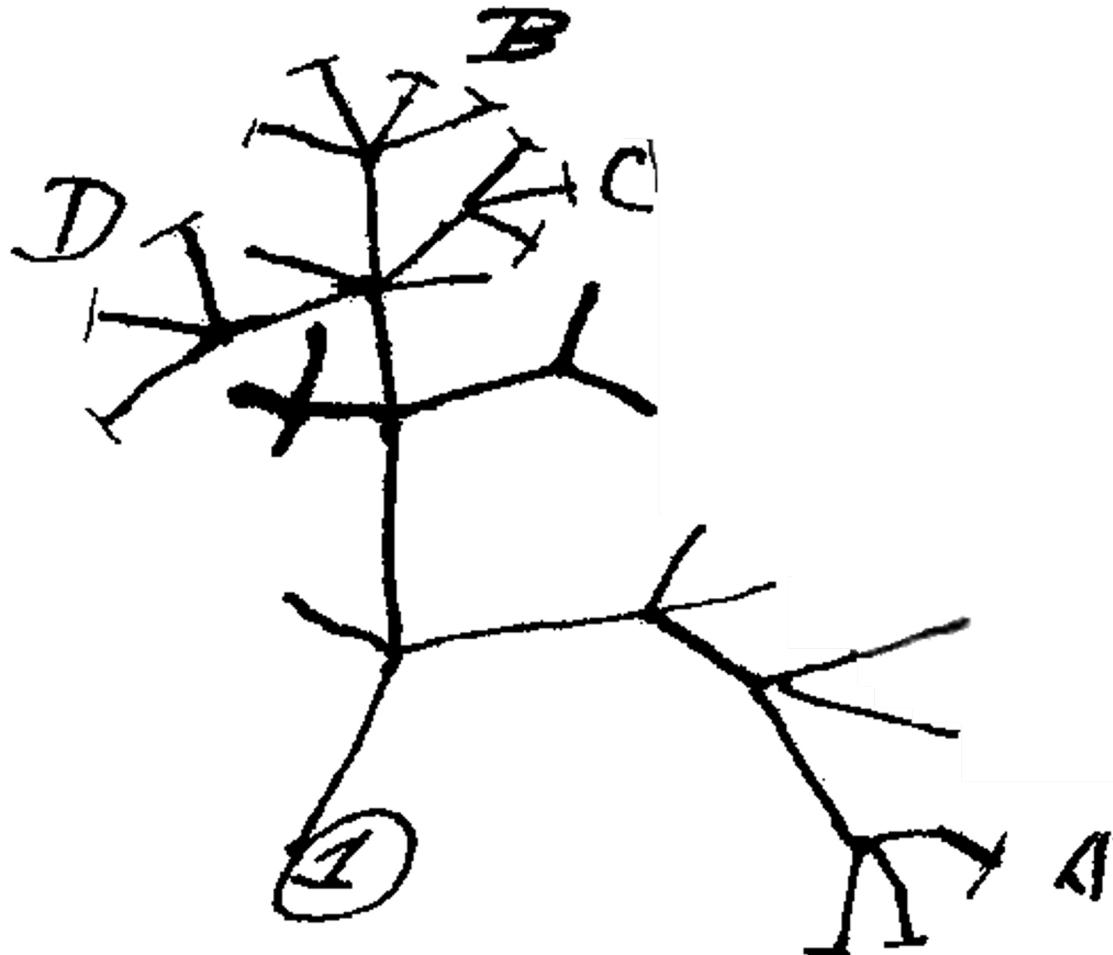
"(1) American: red, biliary, straight – governed by customs; (2) European: white, sanguine, muscular – governed by customs; (3) Asian: yellow (pale), melancholic, stiff – governed by opinion; and (4) African: black, phlegmatic, stiff – governed by chance"

Attempts systematic classification of all life on earth

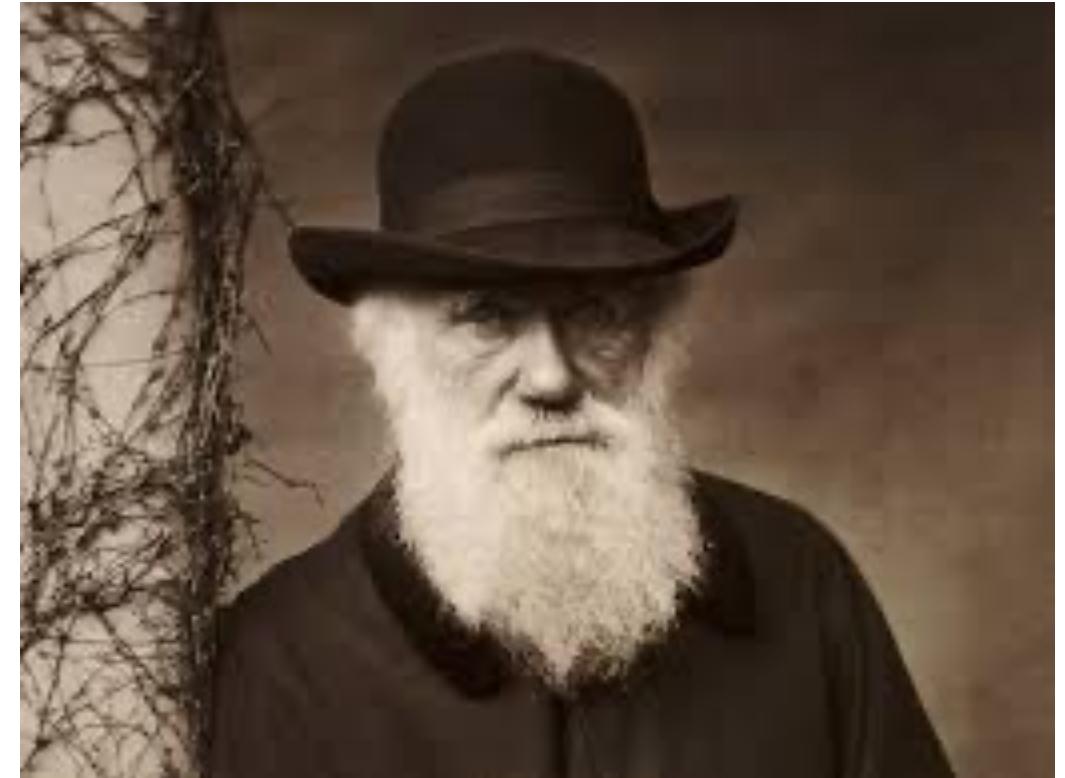
Development of binomial nomenclature

Racism perpetuated far and wide

I think

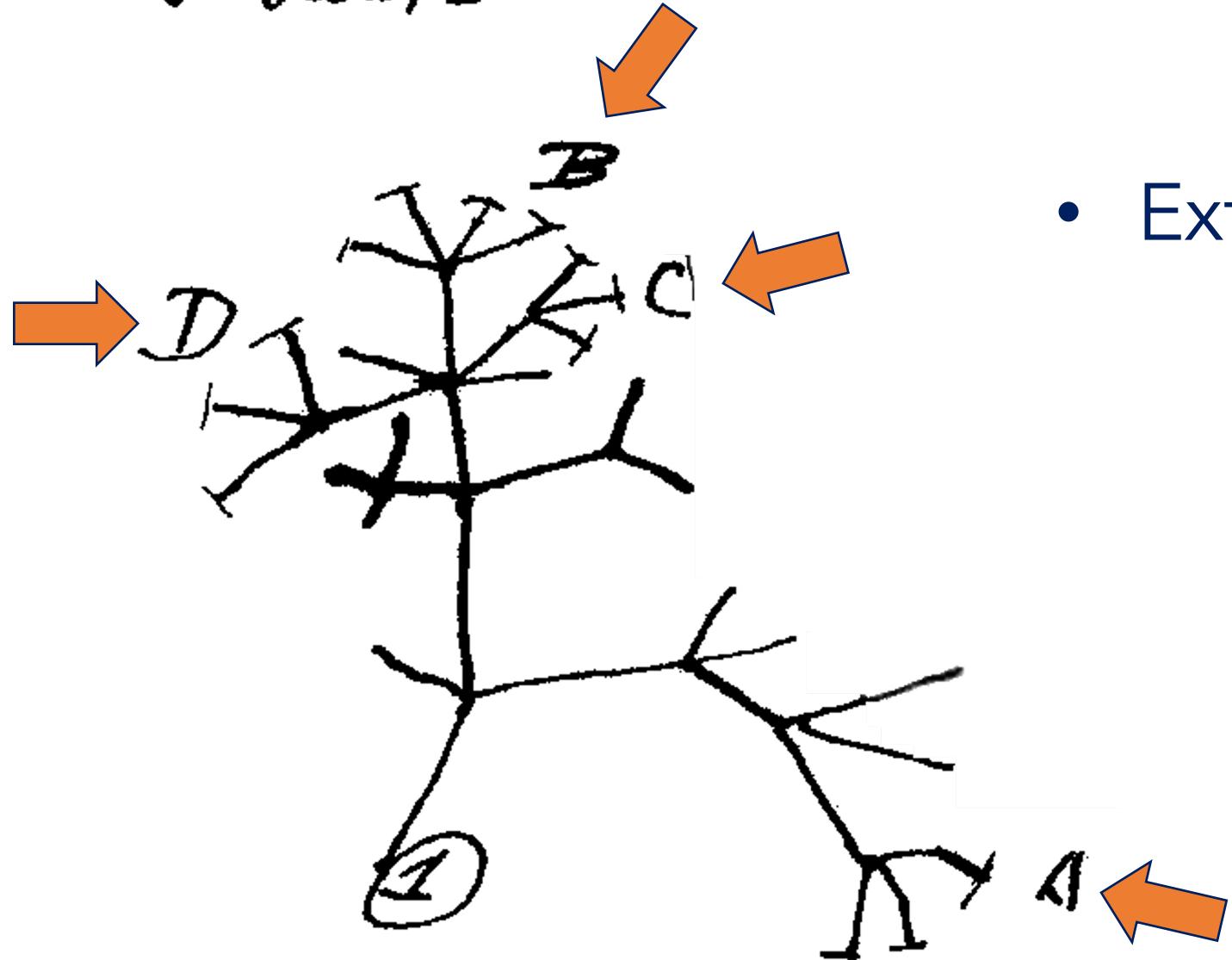


The first tree



Charles Darwin

I think

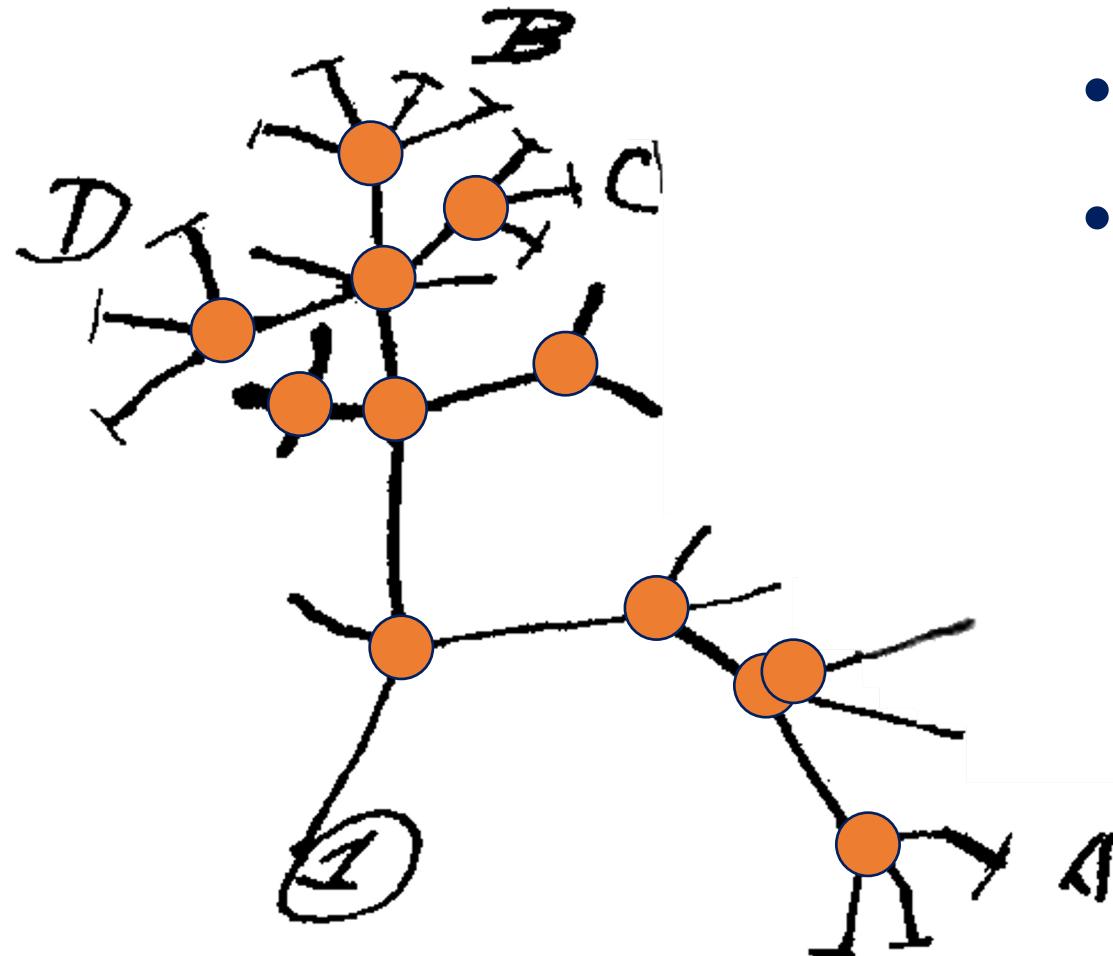


The first tree

- Extant taxa at the tips

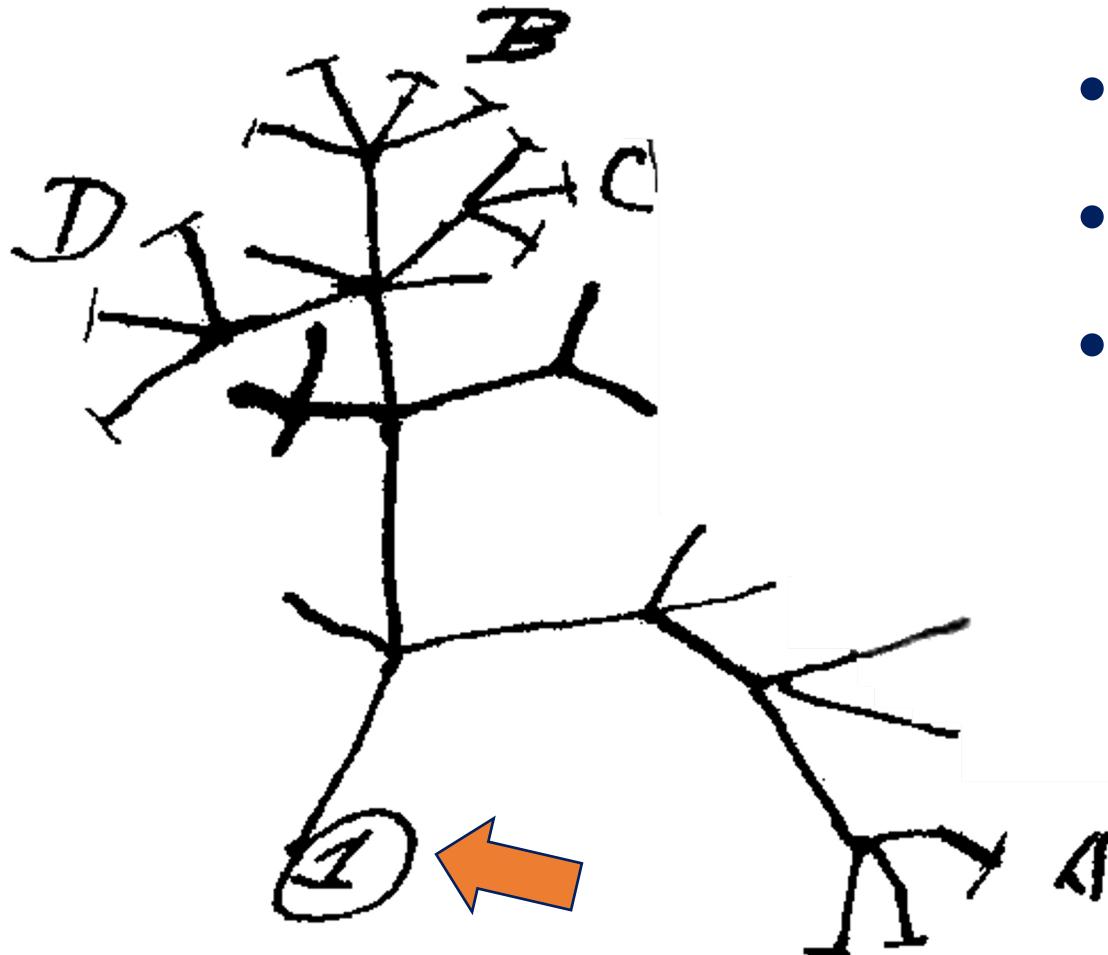
I think

The first tree



- Extant taxa at the tips
- “Splitting” – aka **speciation**

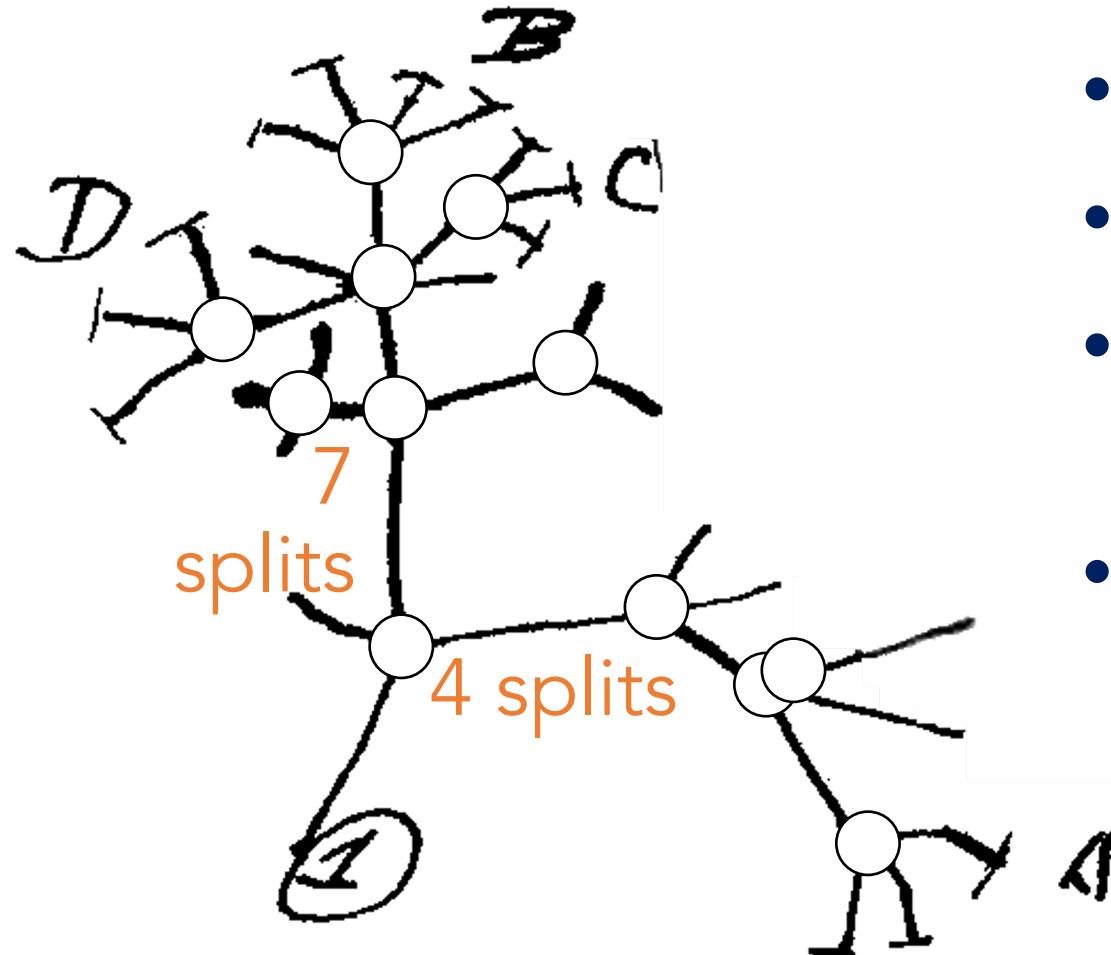
I think



The first tree

- Extant taxa at the tips
- “Splitting” – speciation
- Root – common ancestry

I think

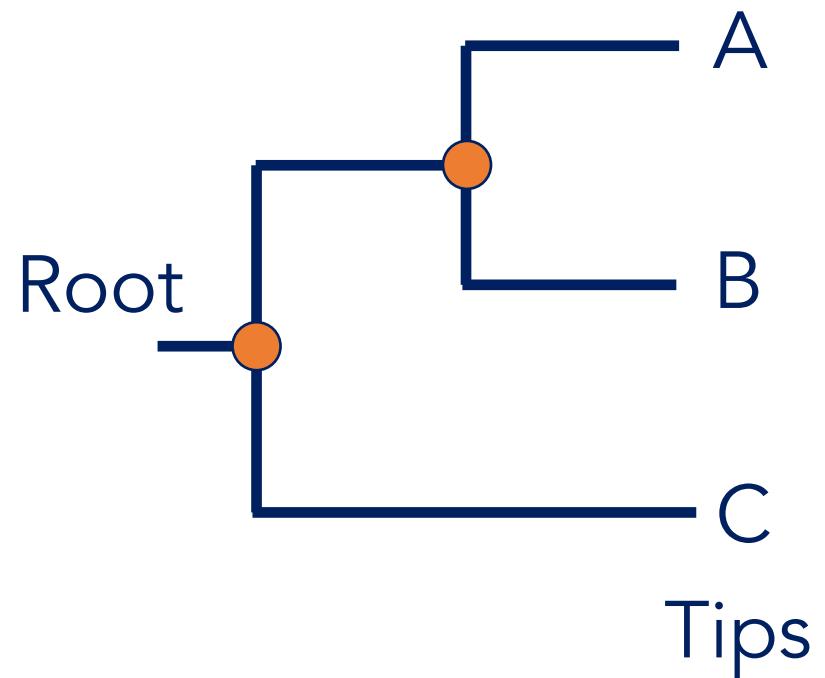


The first tree

- Extant taxa at the tips
- “Splitting” – speciation
- Root – inference of the common ancestor
- Some branches split more than others

Phylogenetic tree interpretation – “tree thinking”

- Trees are a method for hierarchical organization based on relatedness

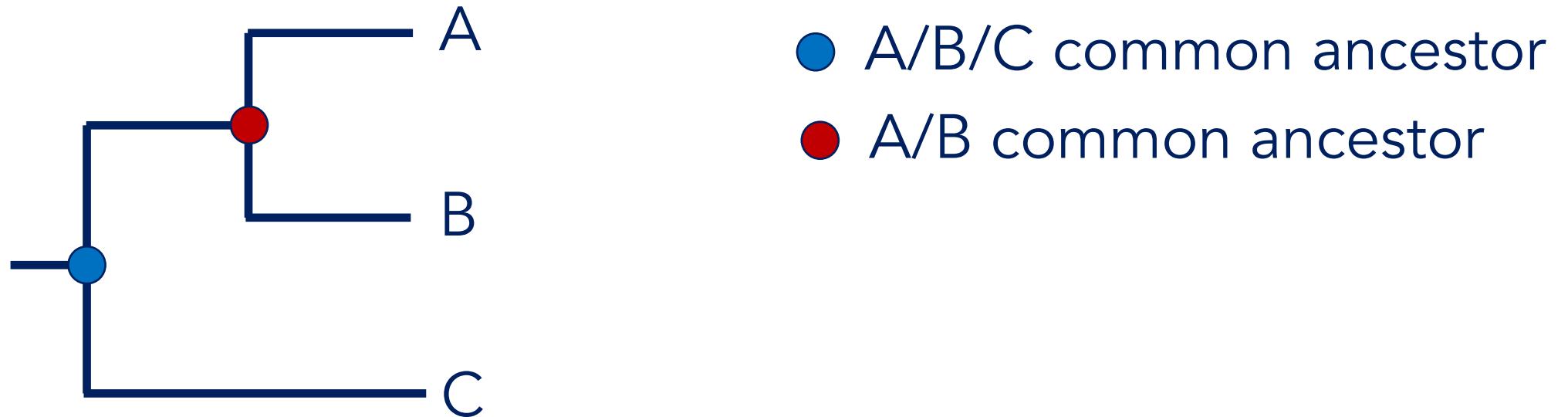


- Node = common ancestor
- Branch

Branching order = Tree Topology

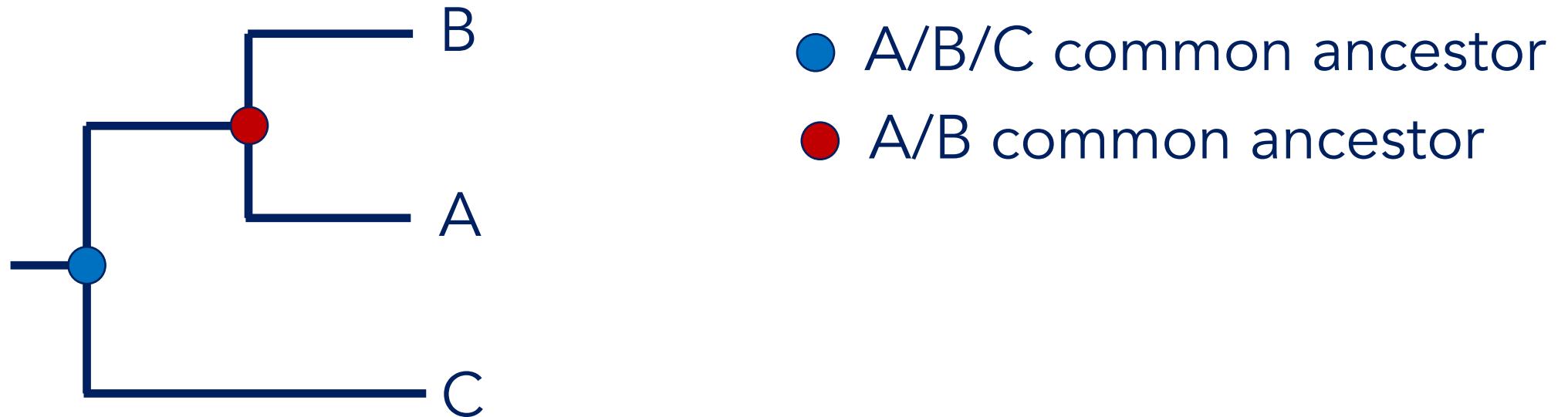
Phylogenetic tree interpretation – “tree thinking”

- Relatedness based on **most recent common ancestor**, not physical distance between tips



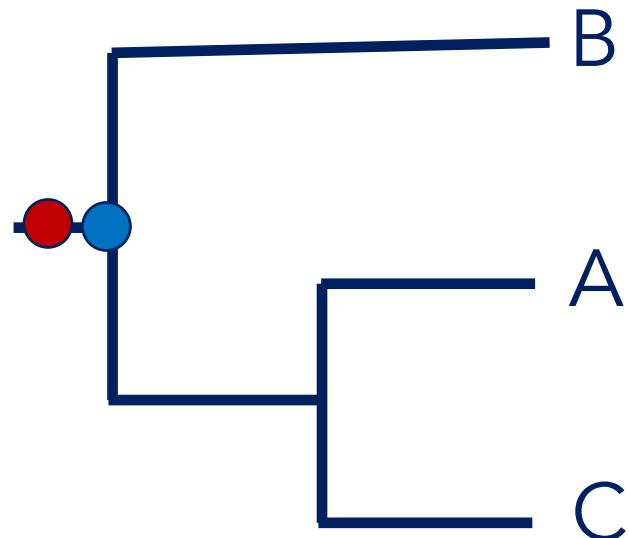
Phylogenetic tree interpretation – “tree thinking”

- Branches can rotate around nodes,
does not change relationships



Phylogenetic tree interpretation – “tree thinking”

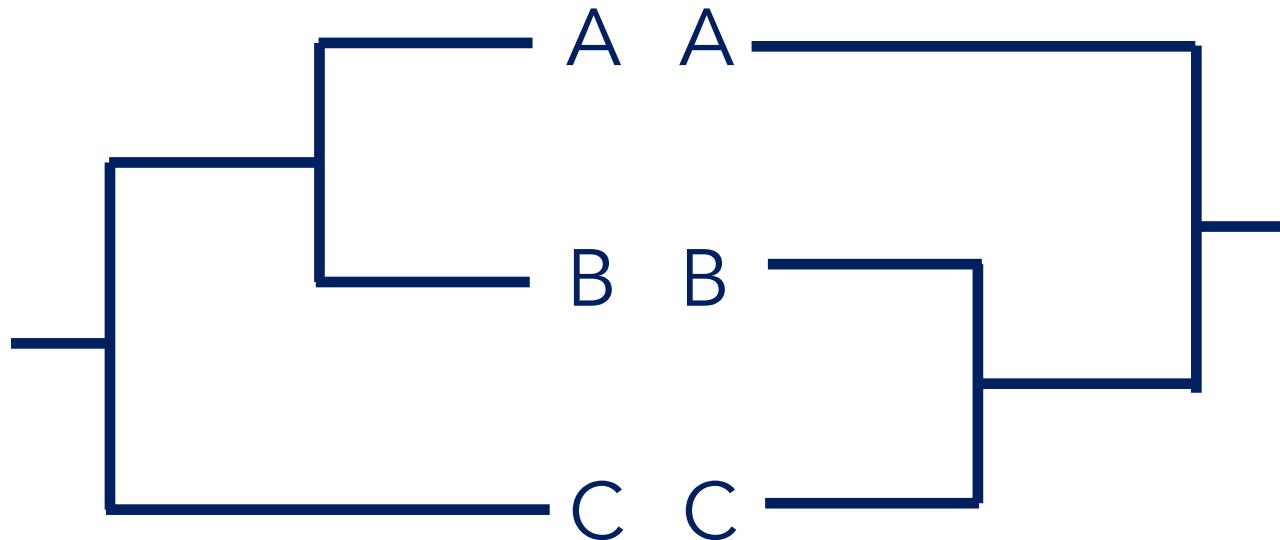
- Branches moving to different parts of the tree **does change relationships**



- A/B/C common ancestor
 - A/B common ancestor
- ||

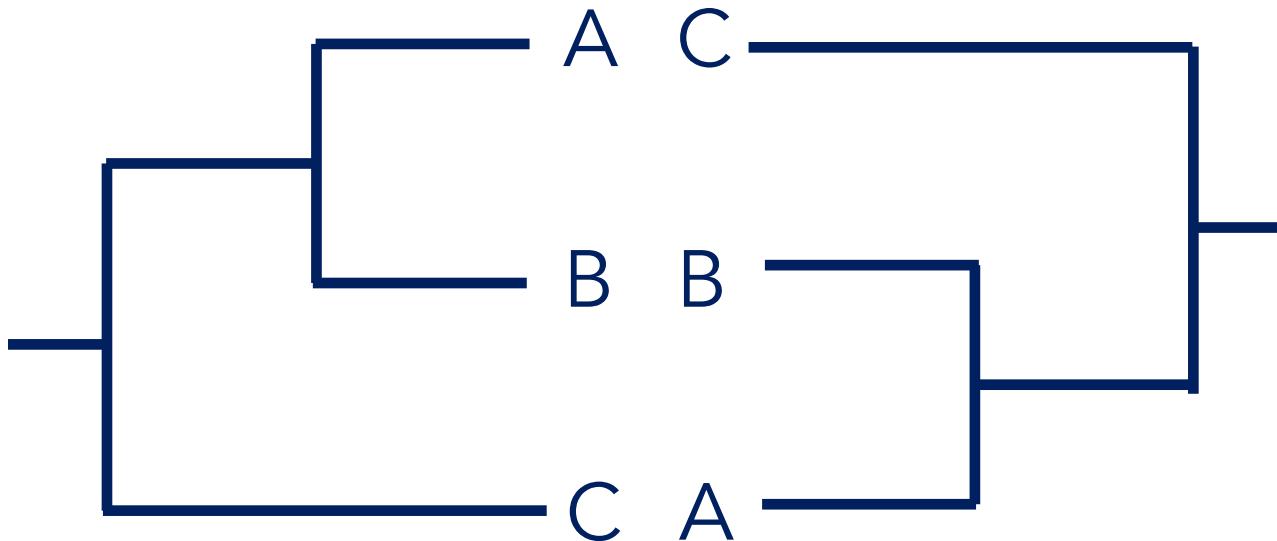
Phylogenetic tree interpretation – “tree thinking”

Are these trees the same?

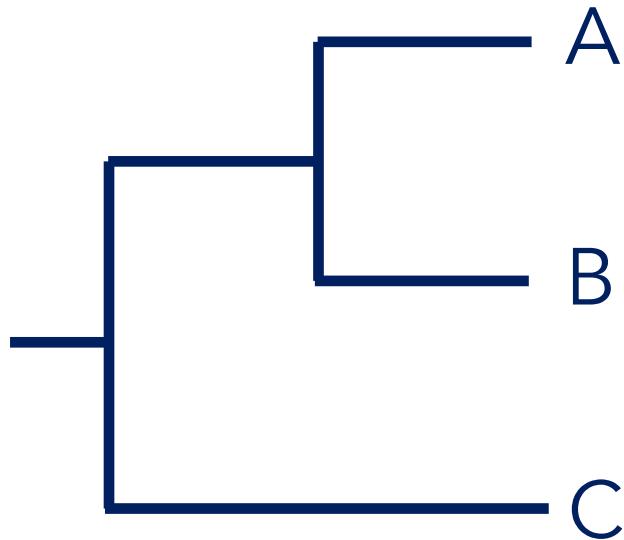


Phylogenetic tree interpretation – “tree thinking”

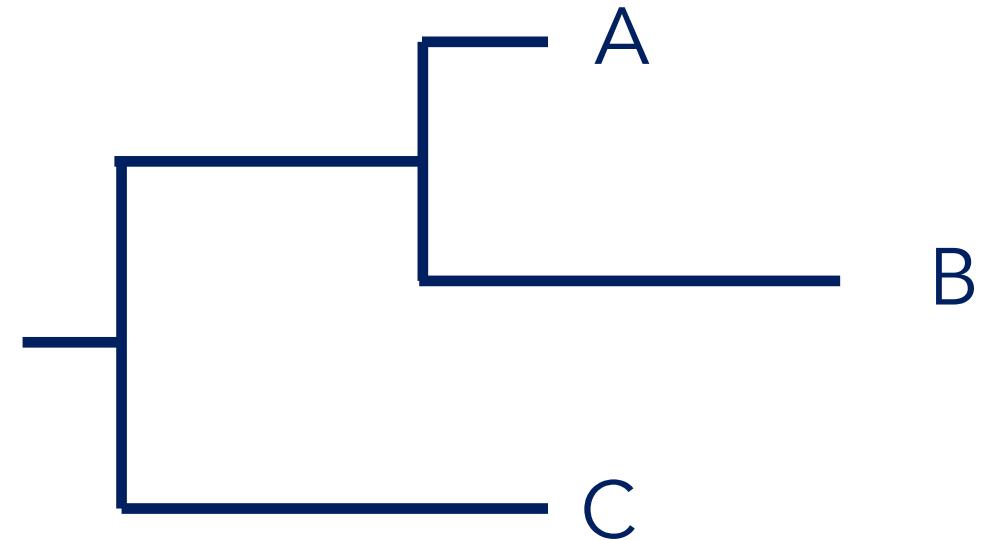
Are these trees the same?



Phylogenetic tree interpretation – “tree thinking”

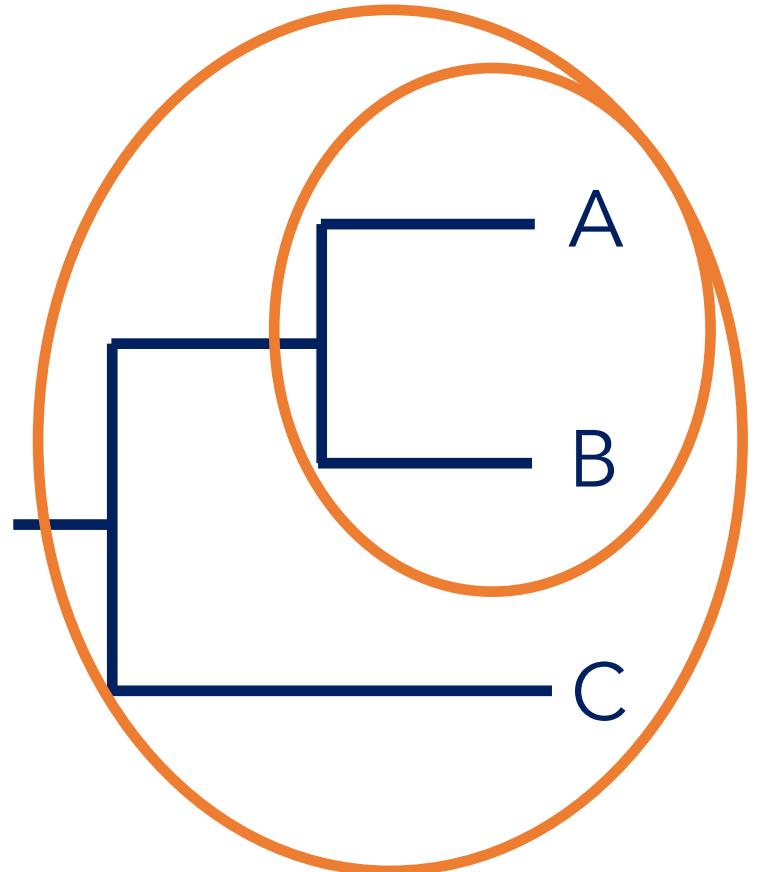


Cladogram
species relationships only



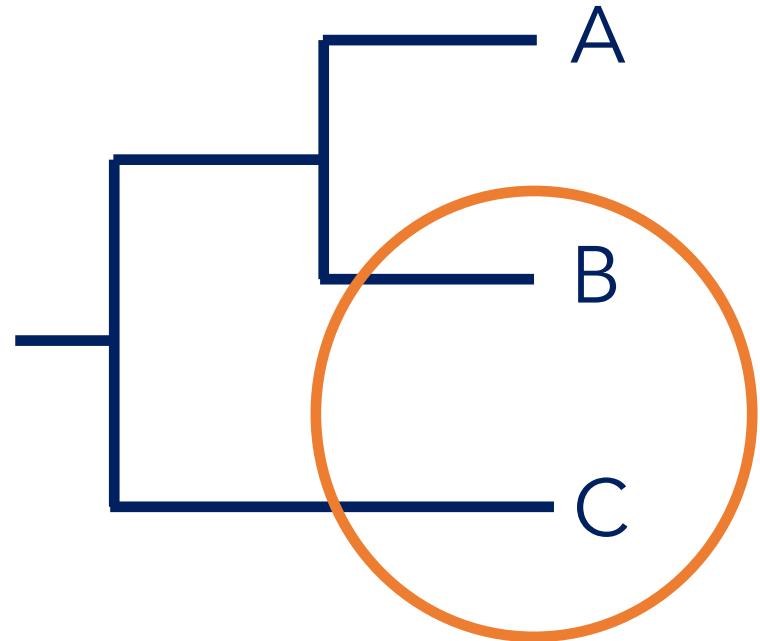
Phylogram
species relationships +
amount of evolution

Cladistics



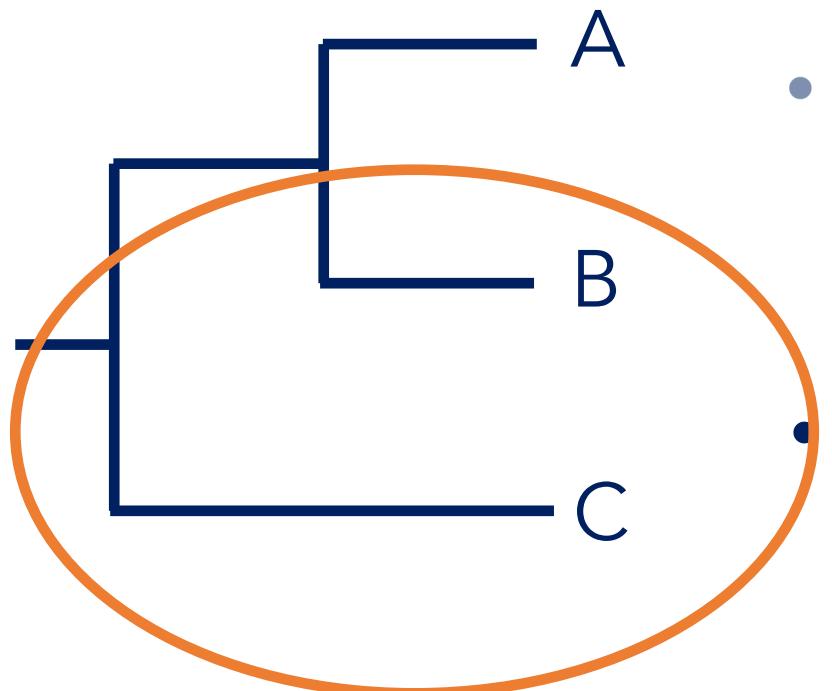
- Monophyletic –
A group of taxa representing all the descendants of a shared common ancestor

Cladistics



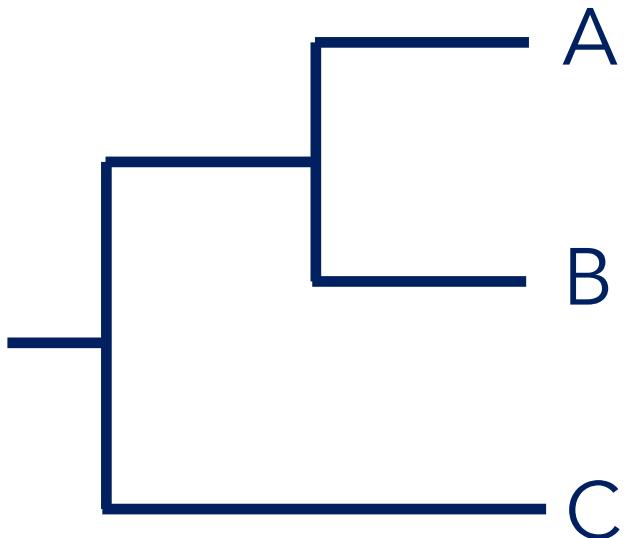
- Monophyletic –
A group of taxa representing all the descendants of a shared common ancestor
- Polyphyletic –
A group of taxa descended from multiple distinct common ancestors

Cladistics



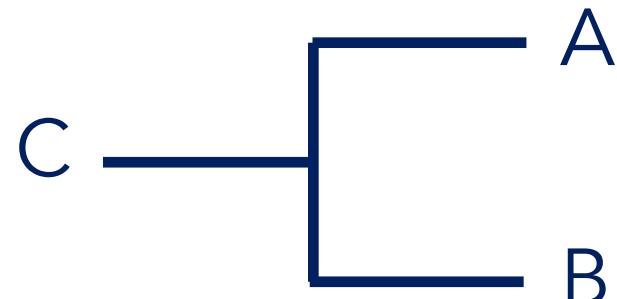
- Monophyletic –
A group of taxa representing all the descendants of a shared common ancestor
- Polyphyletic –
A group of taxa descended from multiple distinct common ancestors
- Paraphyletic –
A group of taxa representing some but not all of the descendants of a shared common ancestor

Into the Treospace



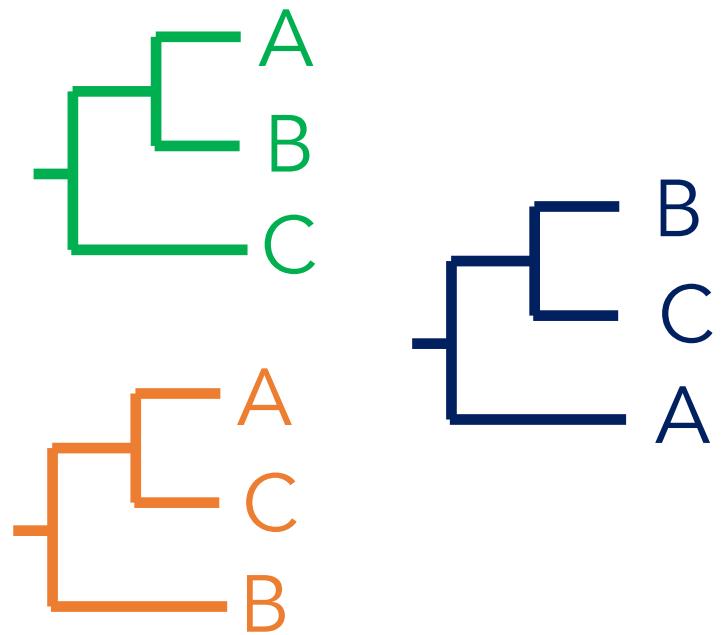
Rooted

Statement about the common ancestor

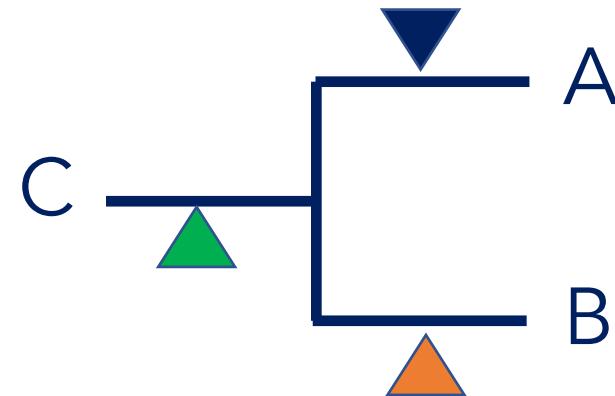


Unrooted

Into the Treospace



Rooted



Unrooted

Into the Treospace – number of possible trees

TABLE 1. THE NUMBERS OF ROOTED TREES WITH n LABELLED TIPS AND WITH UNLABELLED INTERIOR NODES. THE LEFT COLUMN COUNTS ALL TREES, THE RIGHT COLUMN ONLY BIFURCATING TREES.

n	All trees	Bifurcating trees
1	1	1
2	1	1
3	4	3
4	26	15
5	236	105
6	2,752	945
7	39,208	10,395
8	660,032	135,135
9	12,818,912	2,027,025
10	282,137,824	34,459,425
11	6,939,897,856	654,729,075
12	188,666,182,784	13,749,310,575
13	5,617,349,020,544	316,234,143,225
14	181,790,703,209,728	7,905,853,580,625
15	6,353,726,042,486,112	213,458,046,676,875
16	238,513,970,965,250,048	6,190,283,353,629,375
17	9,571,020,586,418,569,216	191,898,783,962,510,625
18	408,837,905,660,430,516,224	6,332,659,870,762,850,625
19	18,522,305,410,364,568,764,416	221,643,095,476,699,771,875
20	887,094,711,304,094,583,095,296	8,200,794,532,637,891,559,375
21	44,782,218,857,751,551,087,214,592	319,830,986,772,877,770,815,625
22	2,376,613,641,928,796,906,249,519,104	13,113,070,457,687,988,603,440,625

Into the Treospace – number of possible trees

TABLE 1. THE NUMBERS OF ROOTED TREES WITH n LABELLED TIPS AND WITH UNLABELLED INTERIOR NODES. THE LEFT COLUMN COUNTS ALL TREES, THE RIGHT COLUMN ONLY BIFURCATING TREES.

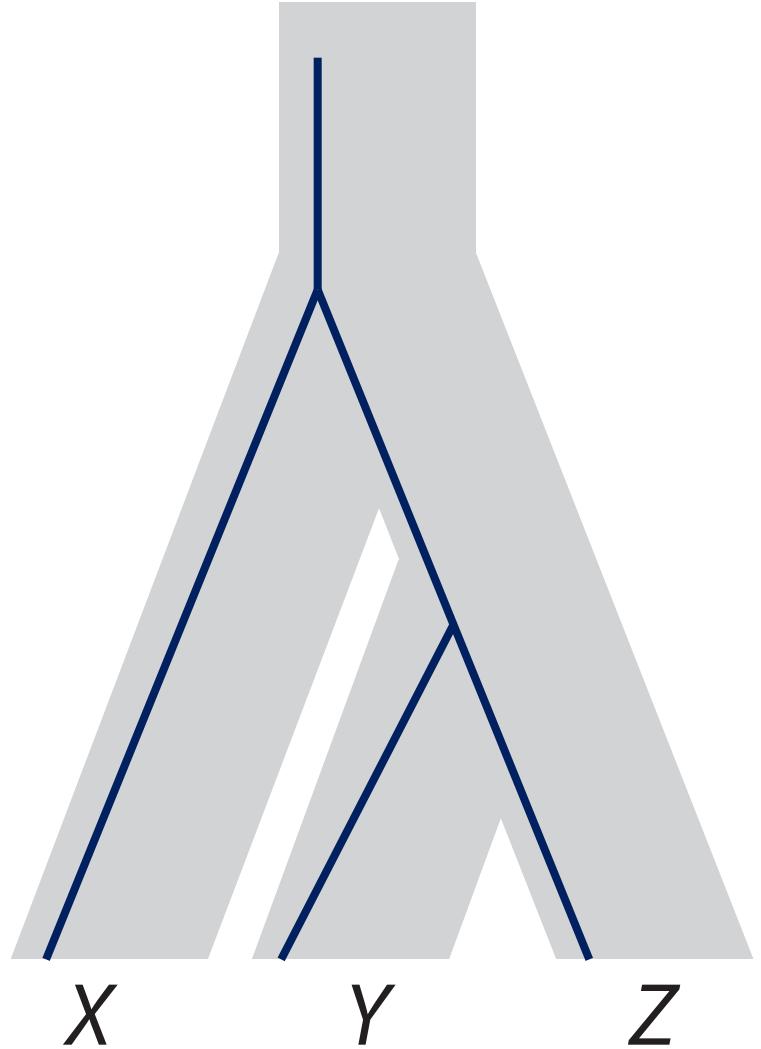
n	All trees	Bifurcating trees
1		1
2		1
3		4
4		26
5		236
6	2,752	945

With > 20 taxa, the treospace is **too big** to search every possible tree

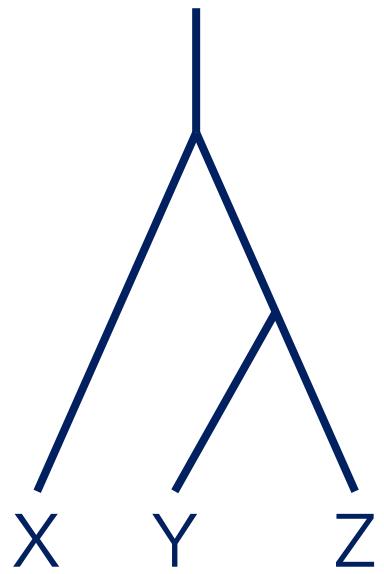
n	All trees	Bifurcating trees
15	6,353,726,042,486,112	213,458,046,676,875
16	238,513,970,965,250,048	6,190,283,353,629,375
17	9,571,020,586,418,569,216	191,898,783,962,510,625
18	408,837,905,660,430,516,224	6,332,659,870,762,850,625
19	18,522,305,410,364,568,764,416	221,643,095,476,699,771,875
20	887,094,711,304,094,583,095,296	8,200,794,532,637,891,559,375
21	44,782,218,857,751,551,087,214,592	319,830,986,772,877,770,815,625
22	2,376,613,641,928,796,906,249,519,104	13,113,070,457,687,988,603,440,625

Into the Treospace

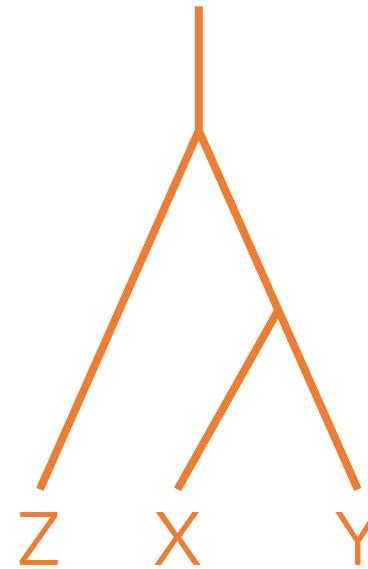
Species trees vs. gene trees



Gene 1 Tree

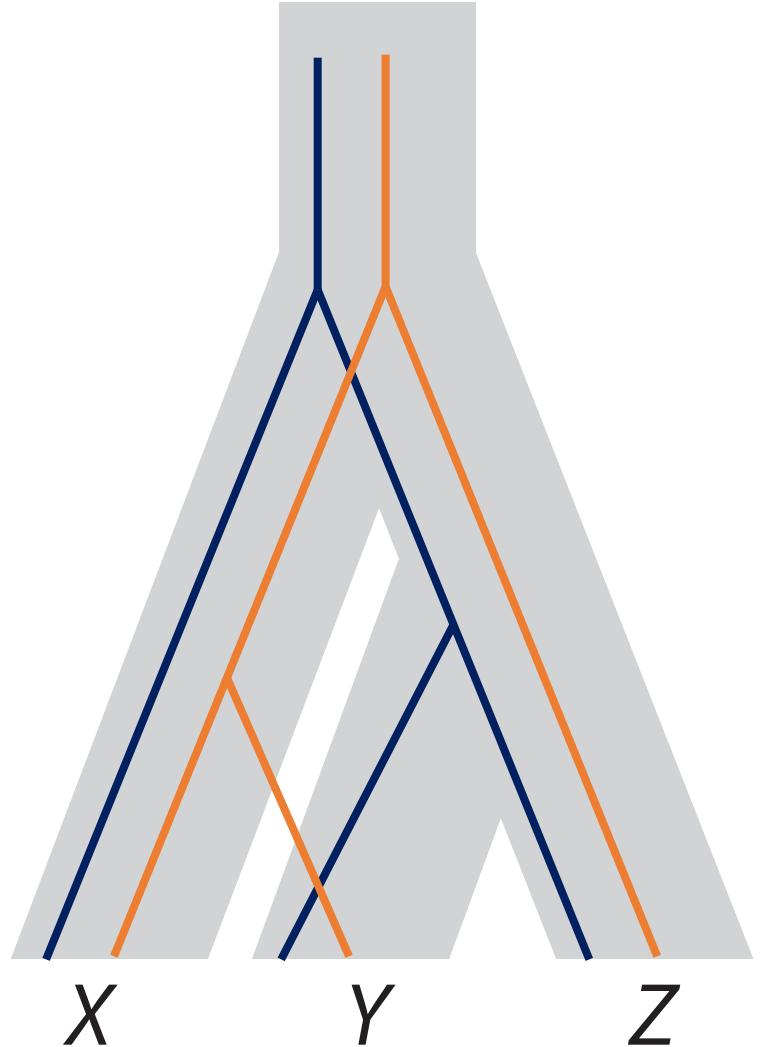


Gene 2 Tree



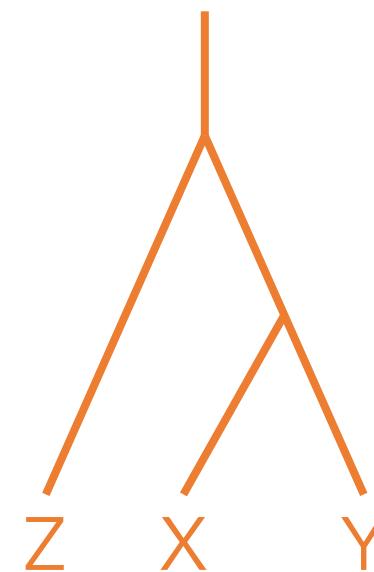
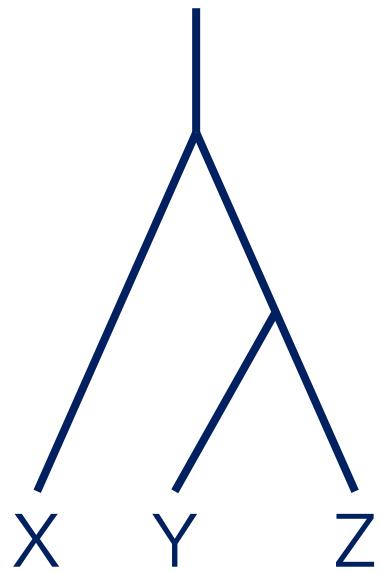
Into the Treospace

Species trees vs. gene trees



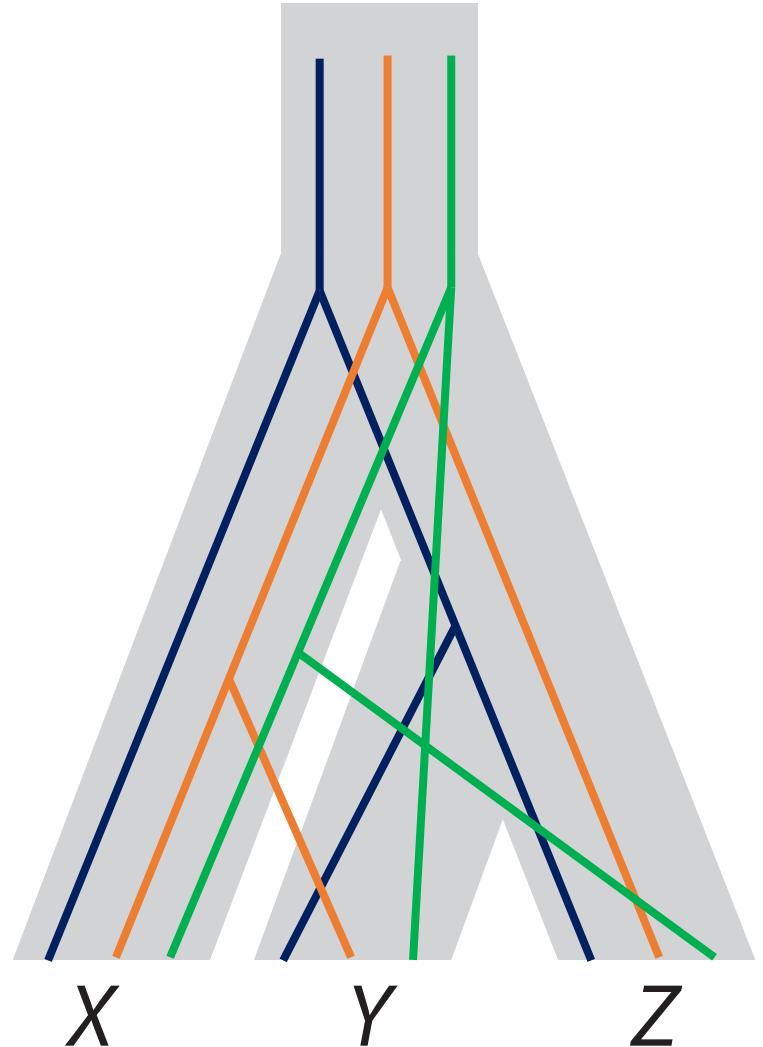
Gene 1 Tree

Gene 2 Tree

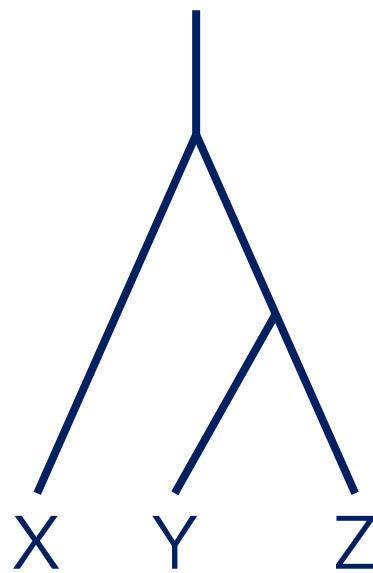


Into the Treospace

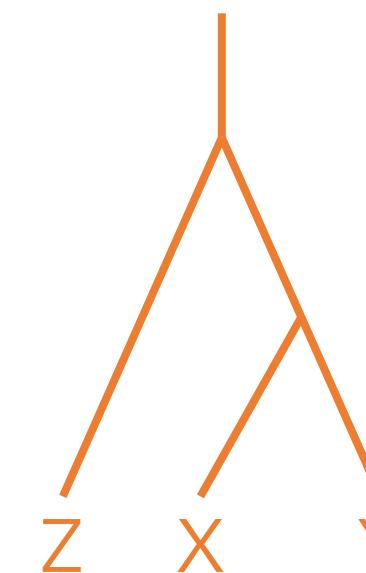
Species trees vs. gene trees



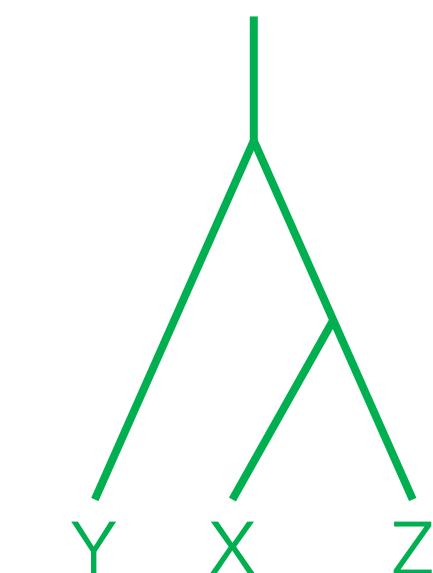
Gene 1 Tree



Gene 2 Tree

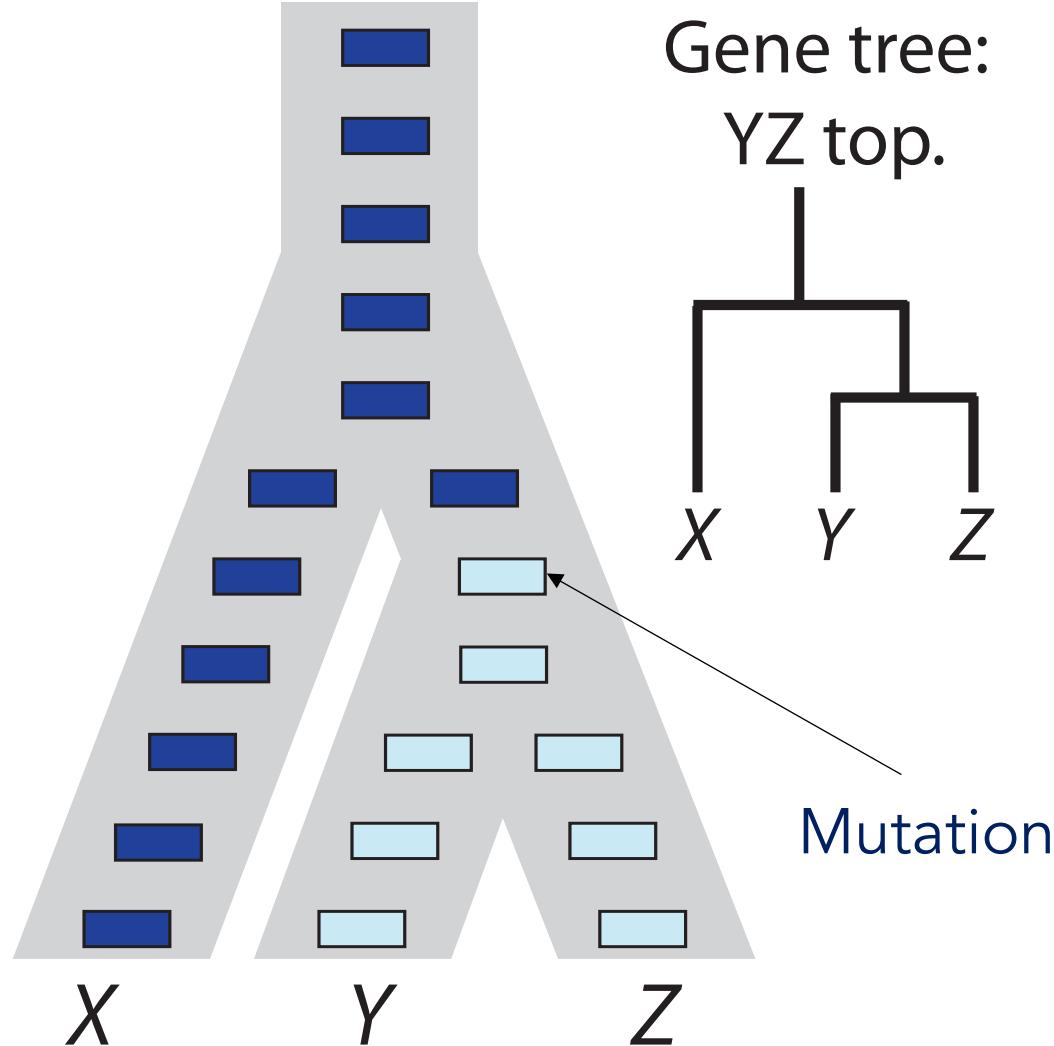


Gene 3 Tree



Evolutionary processes lead to incongruent gene trees

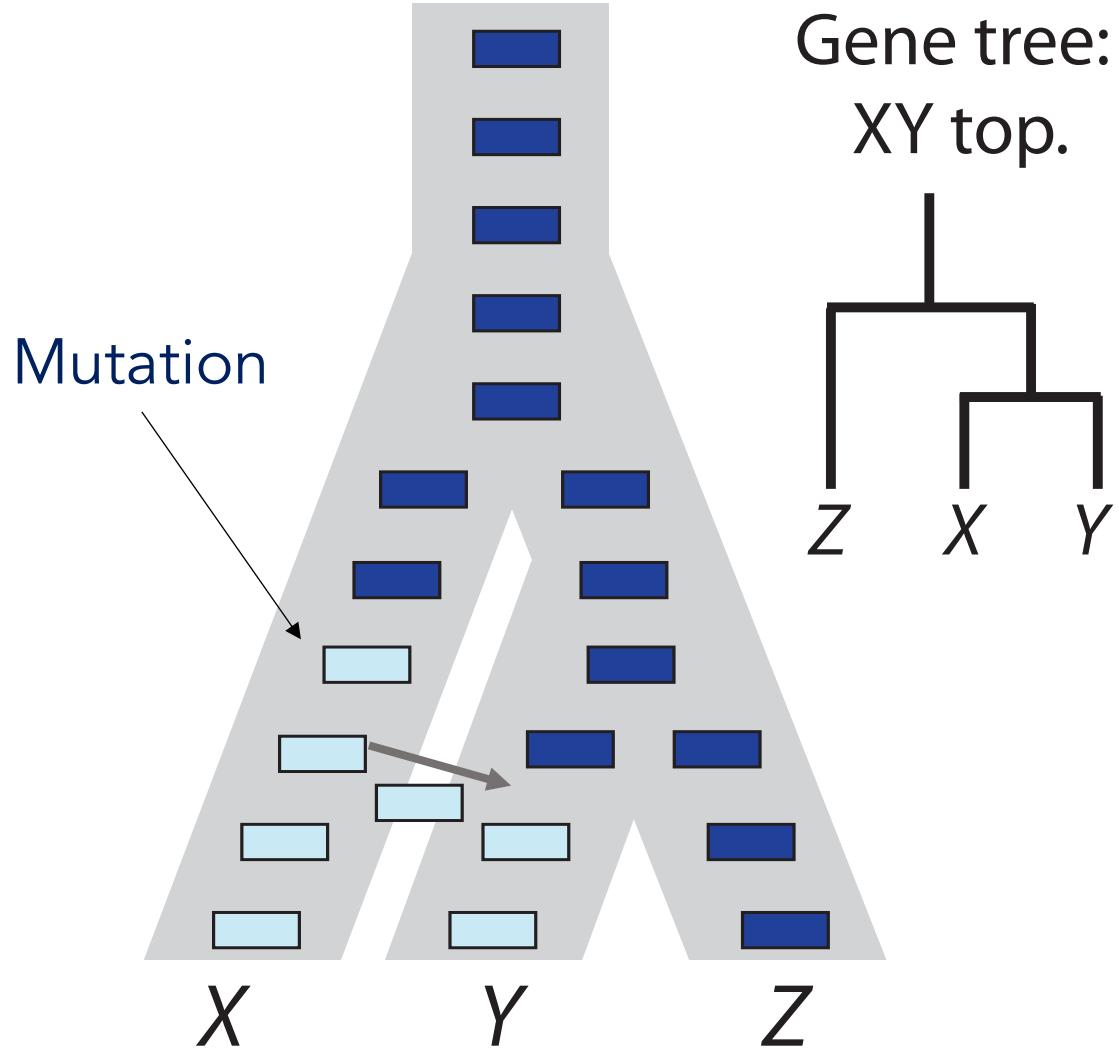
Speciation:



- Results in diverse species
- Gene trees will represent the true species branching order

Evolutionary processes lead to incongruent gene trees

Hybridization/introgression:



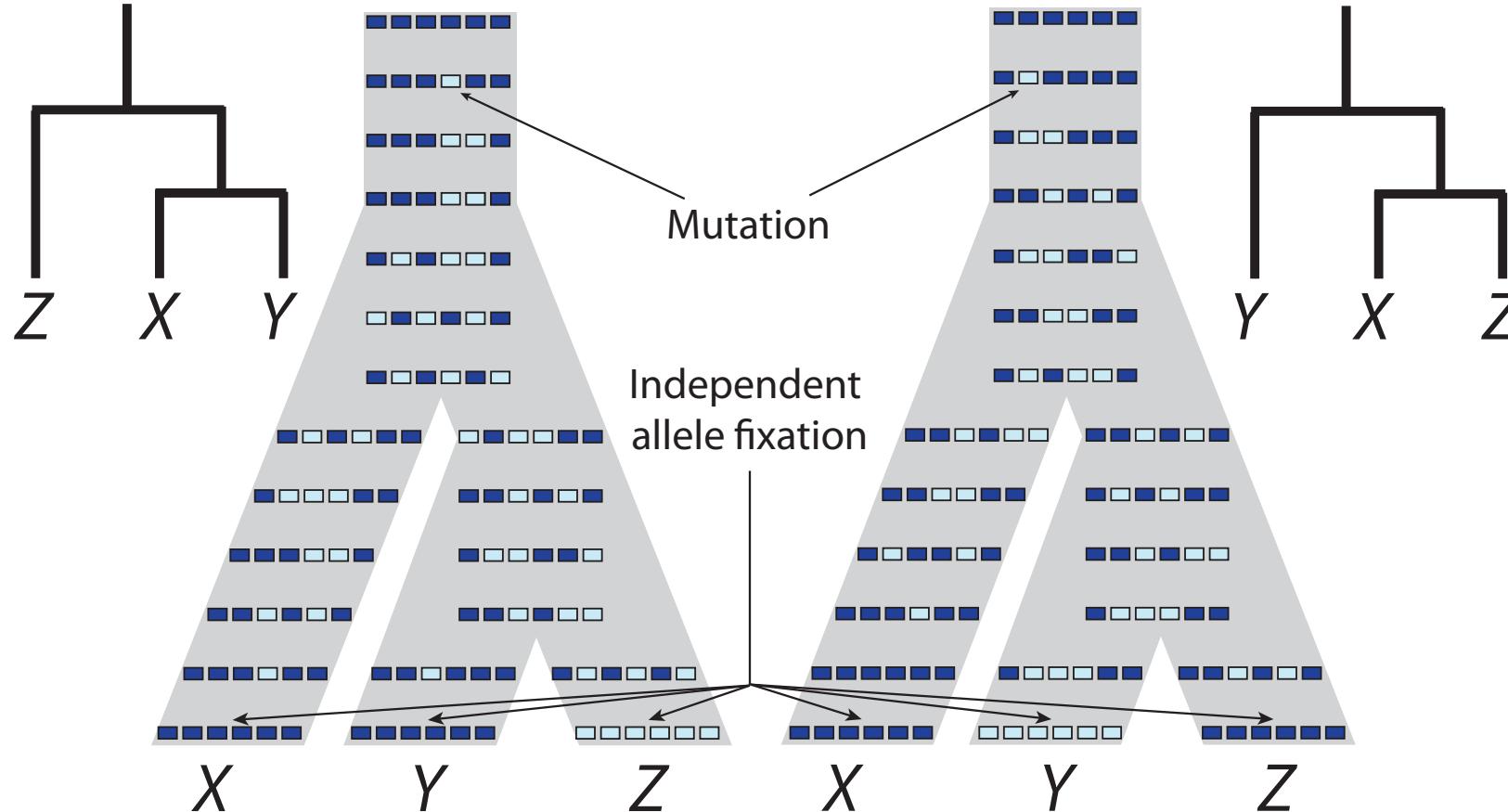
Gene tree:
XY top.

- Results in transfer of traits between species
- Hybridization occurs naturally in ~10% of plants
 - 22 of the world's 25 most important crops
- Occurs artificially during plant breeding

Evolutionary processes lead to incongruent gene trees

Incomplete lineage sorting (ILS):

Gene tree:
XY top.



Gene tree:
XZ top.

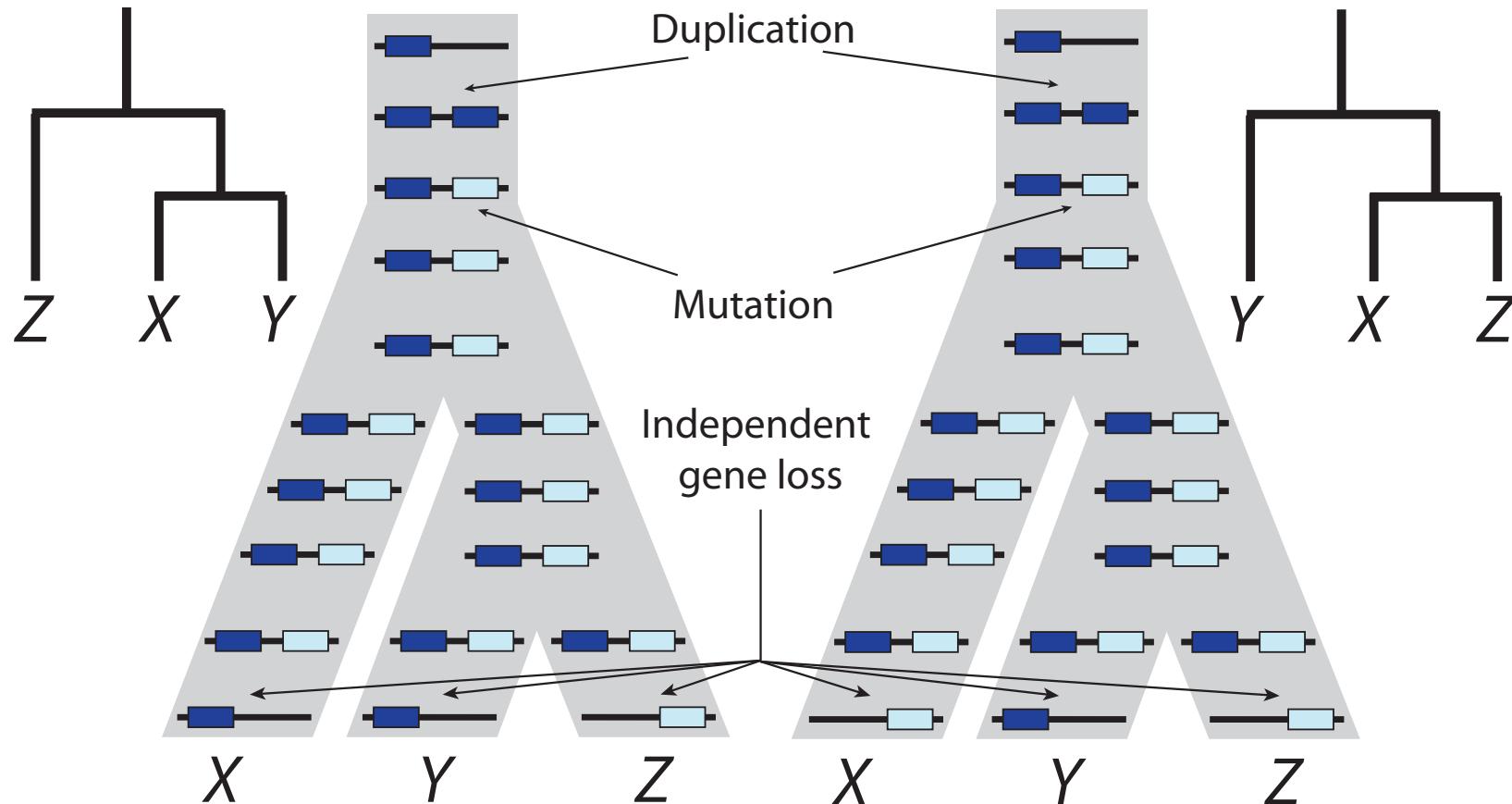
- Prevalent during rapid speciation
- The null hypothesis when studying introgression

Evolutionary processes lead to incongruent gene trees

Gene duplication:

Gene tree:

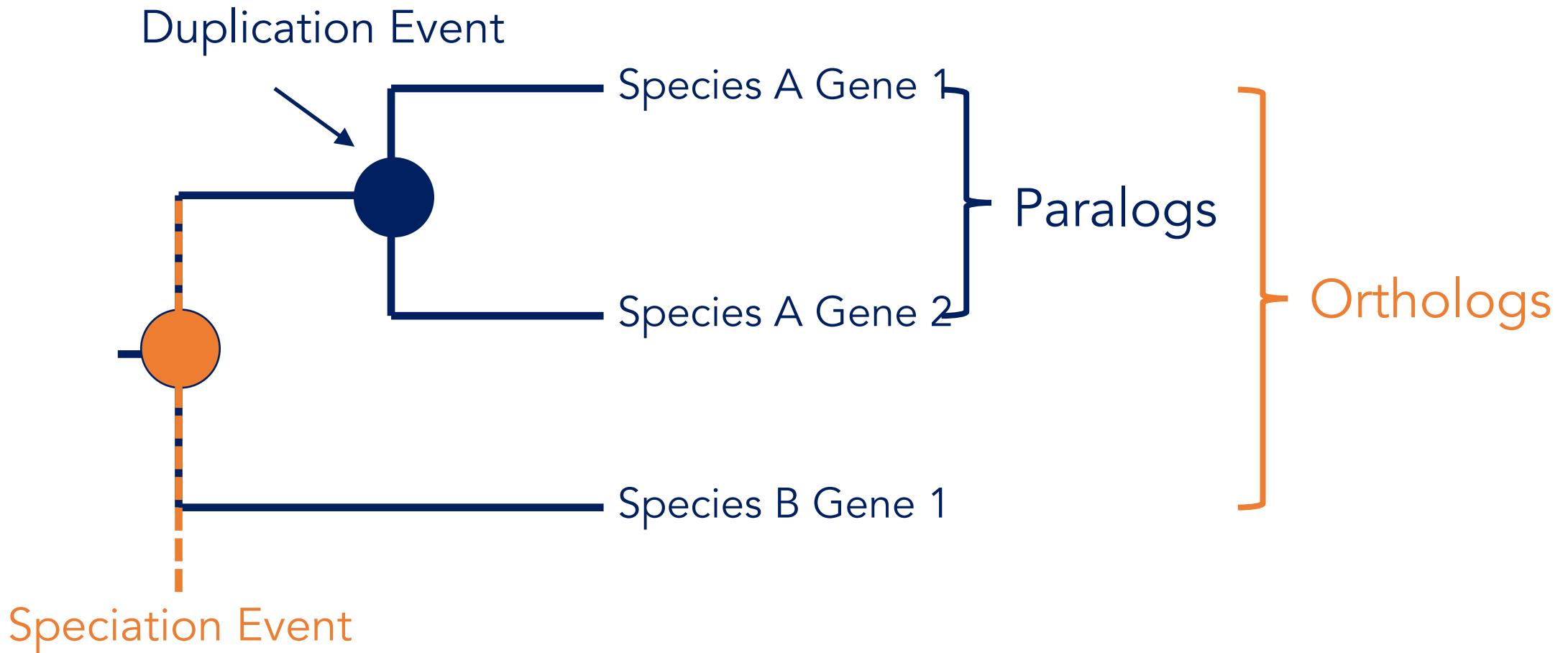
XY top.



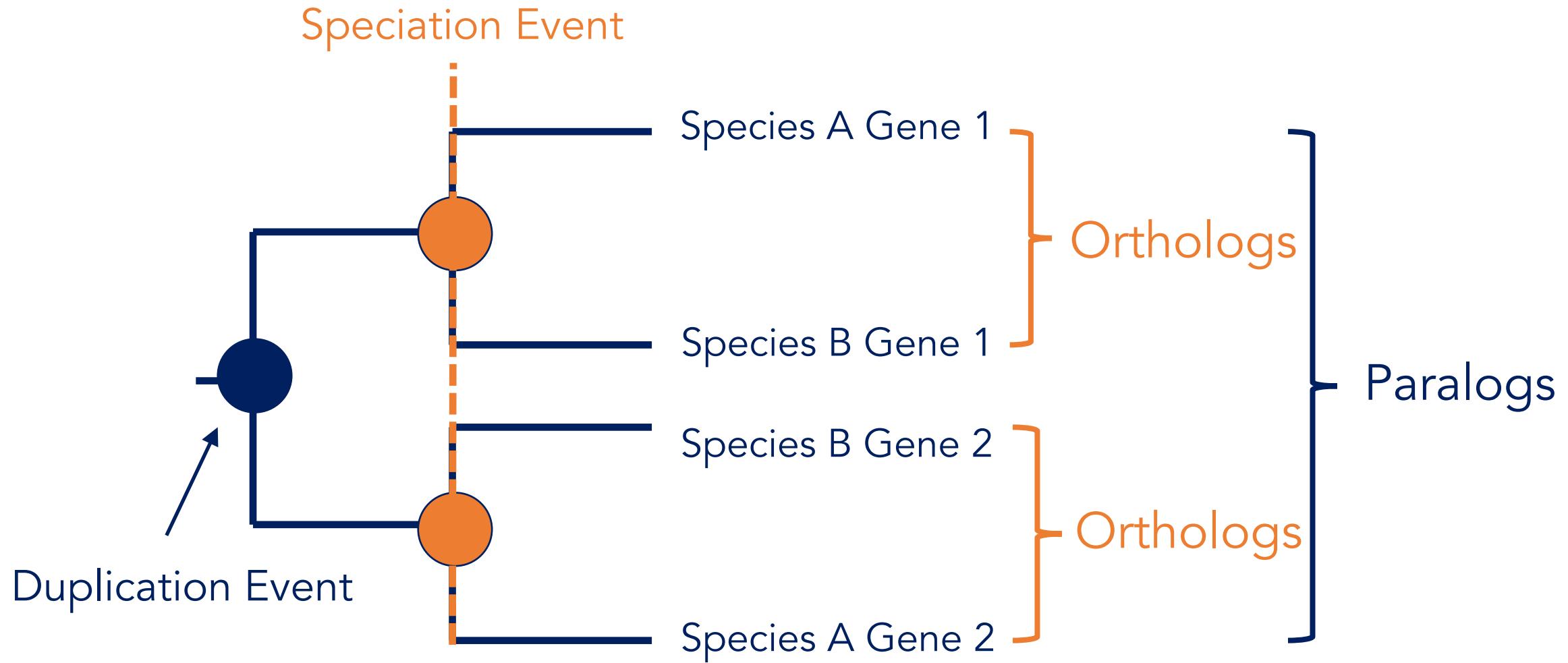
Gene tree:
XZ top.

- Underlies the evolution of novel traits
- ALL seed plants show evidence of whole genome duplication
 - All crops

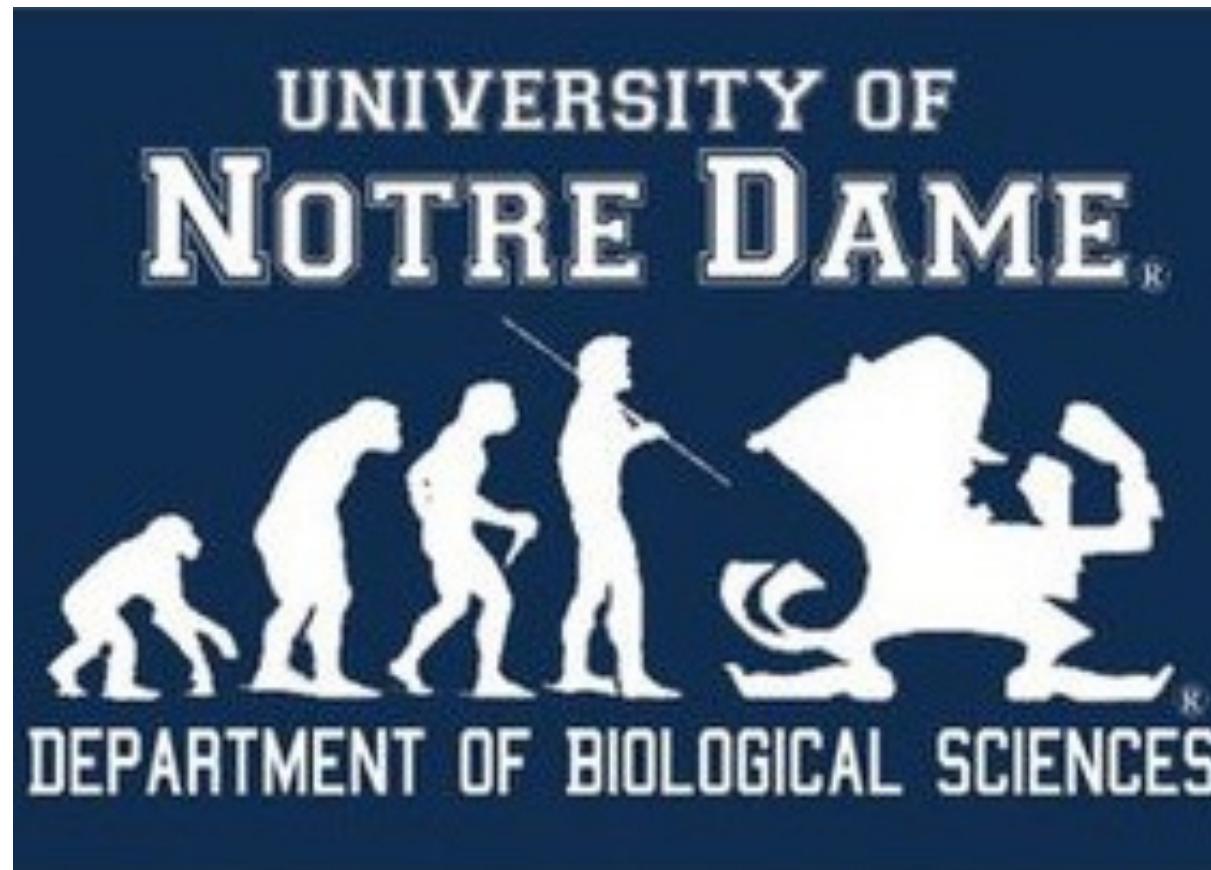
Orthologous or paralogous?



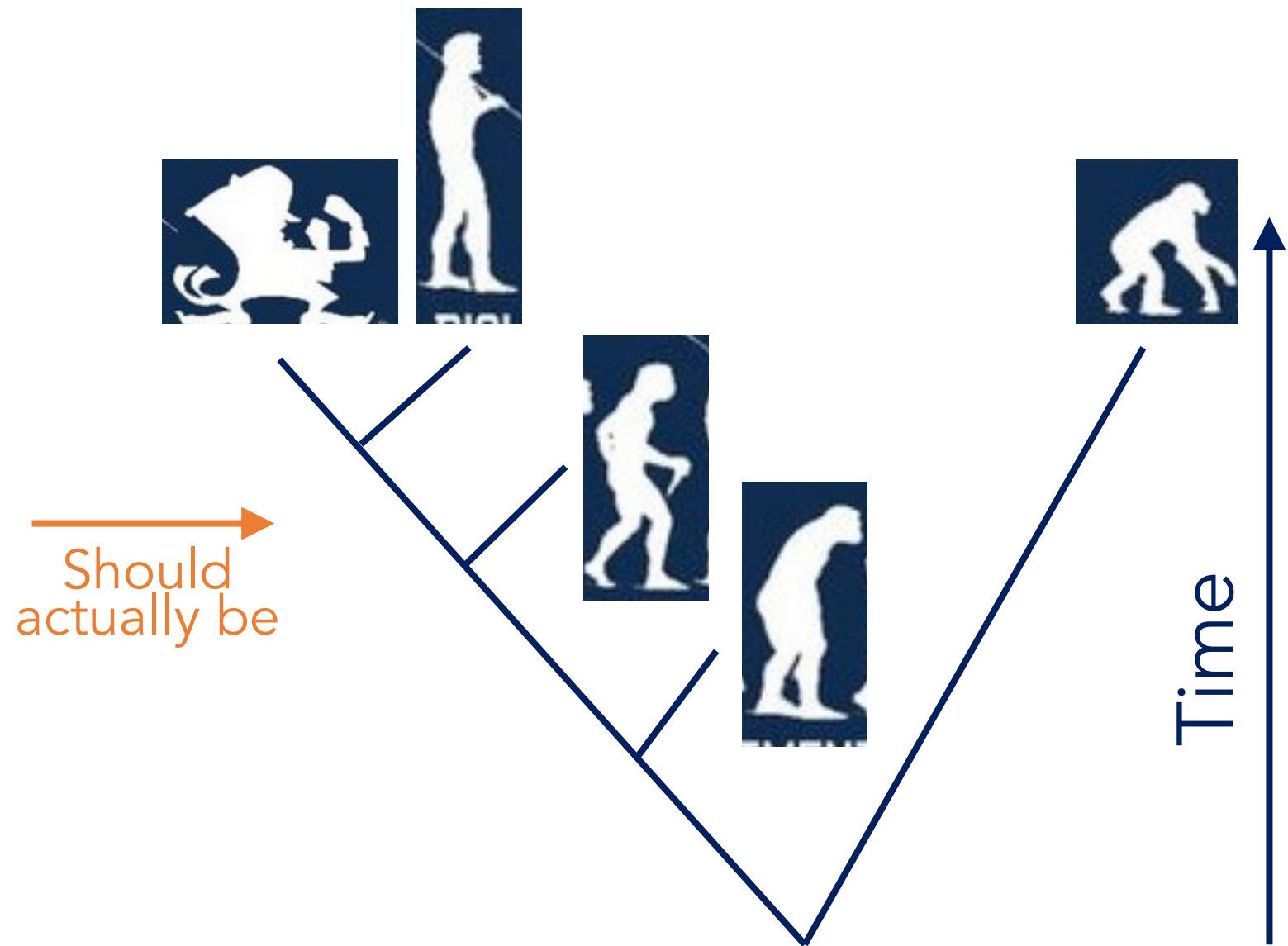
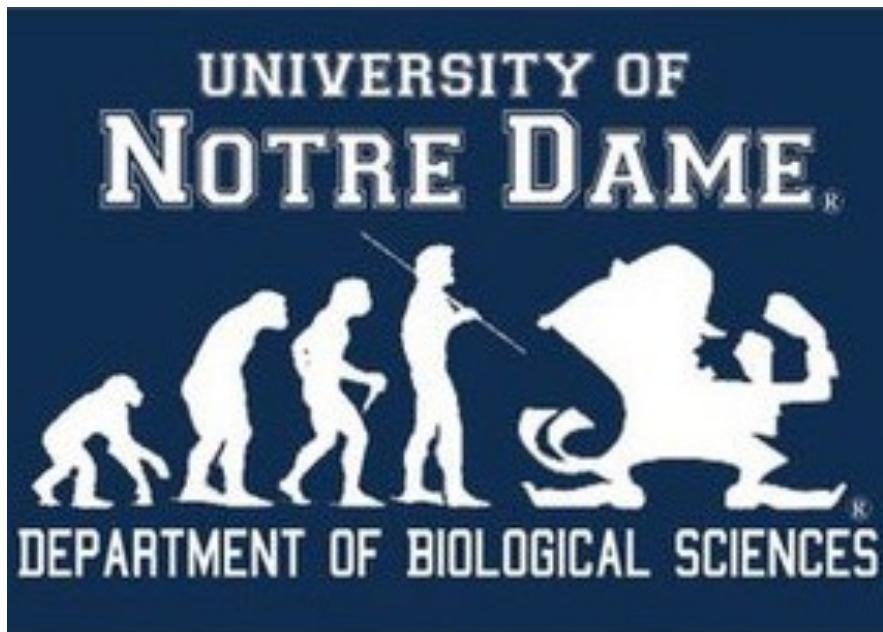
Orthologous or paralogous?



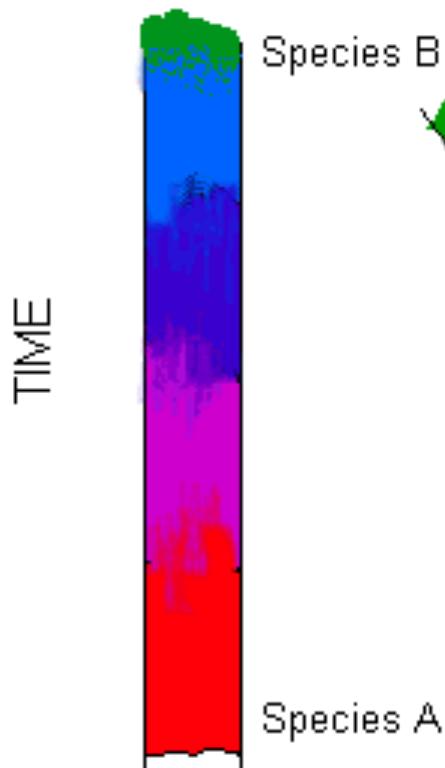
Pinnacles of evolution



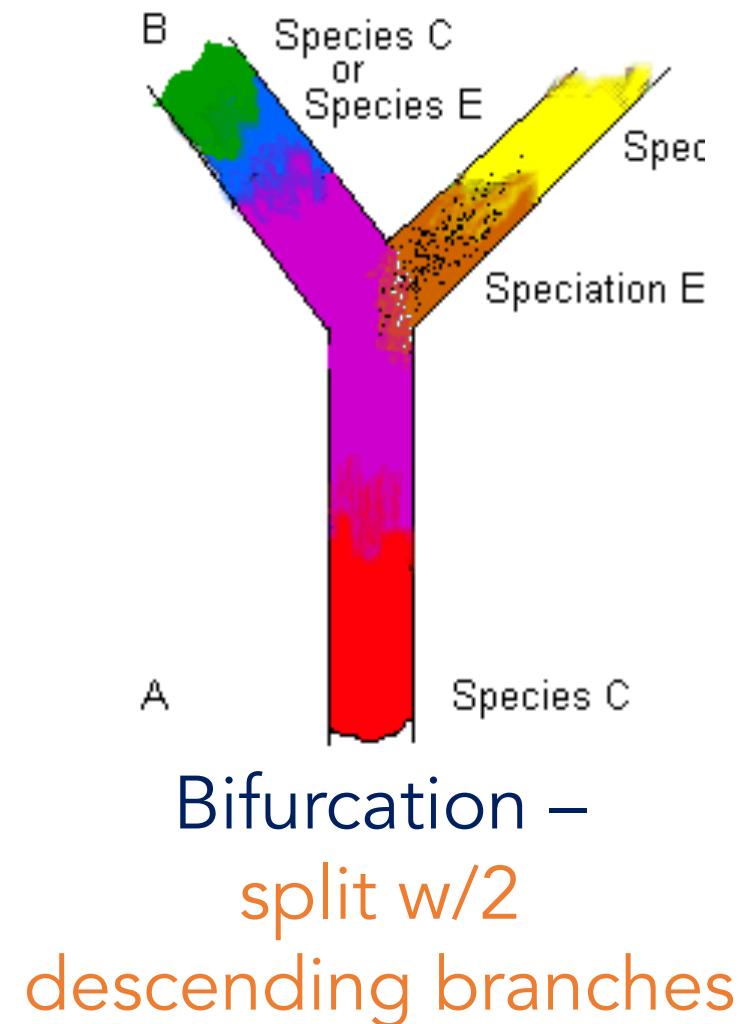
Pinnacles of evolution



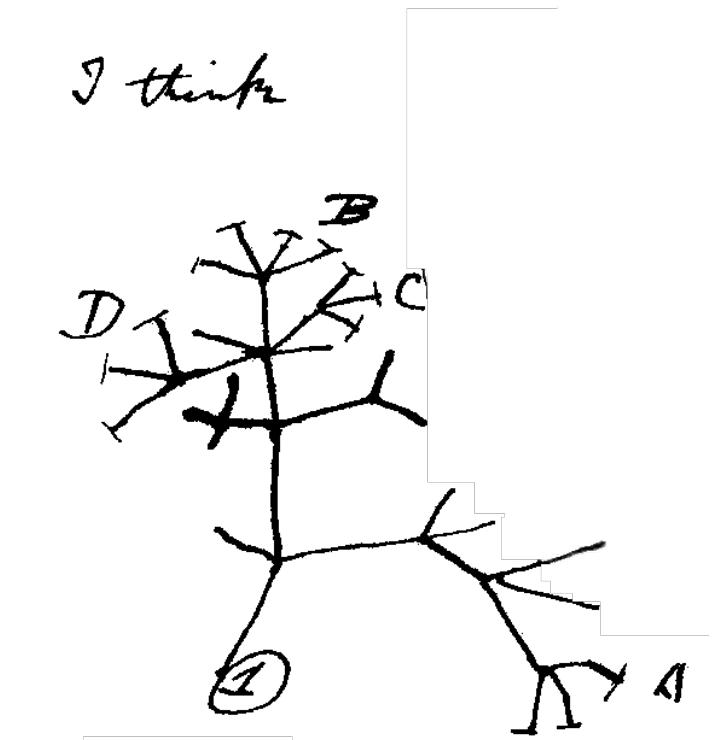
Anagenesis, Bifurcation, & Polytomies



Anagenesis –
no split



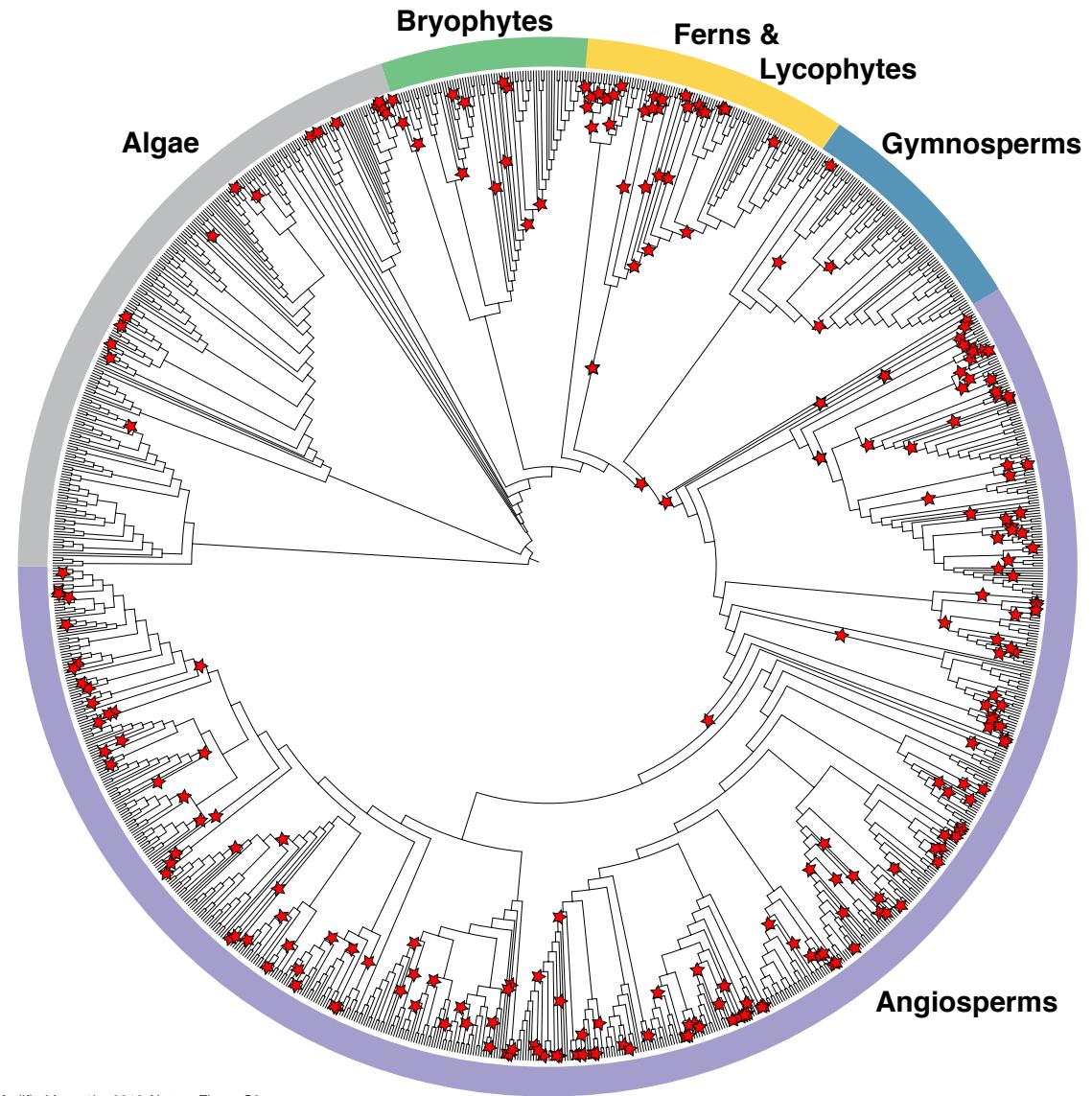
Bifurcation –
split w/2
descending branches



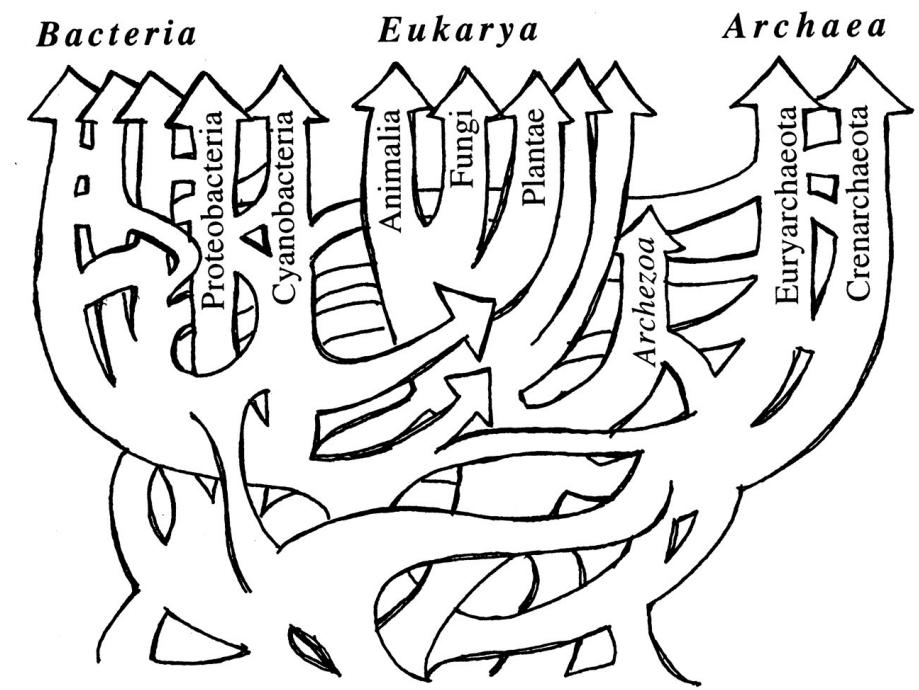
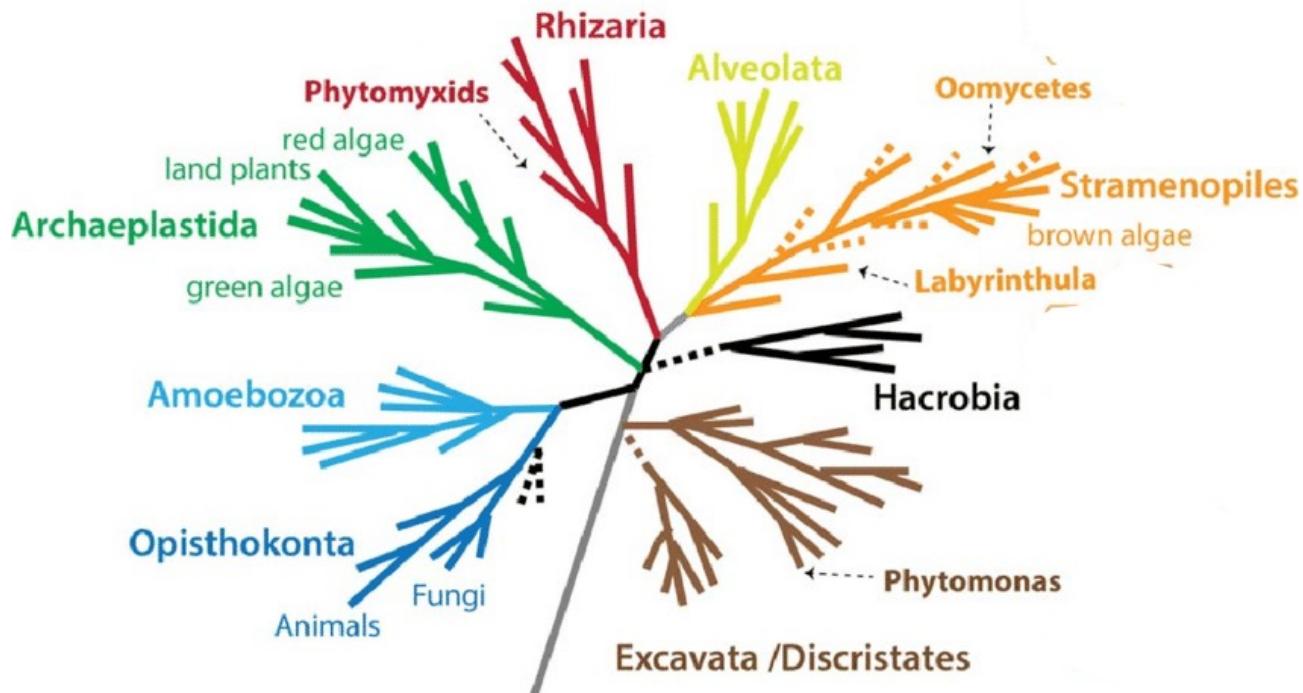
Polytomy –
Split w/ ≥ 3
descending branches

Using phylogenetics to understand nature

- Cataloging the diversity of life
- Identifying & understanding evolutionary transitions
- Understanding evolutionary processes & mechanisms



Cataloguing the diversity of life: who is related to whom

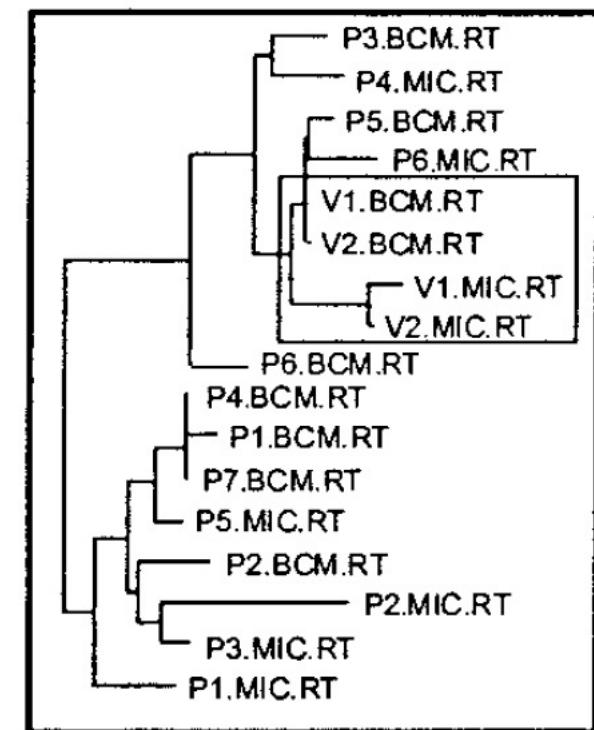


Cataloguing the diversity of life: who is related to whom

Molecular evidence of HIV-1 transmission in a criminal case

Michael L. Metzker^{**}, David P. Mindell[#], Xiao-Mei Liu^{*\$}, Roger G. Ptak^{##}, Richard A. Gibbs*, and David M. Hillis^{**}

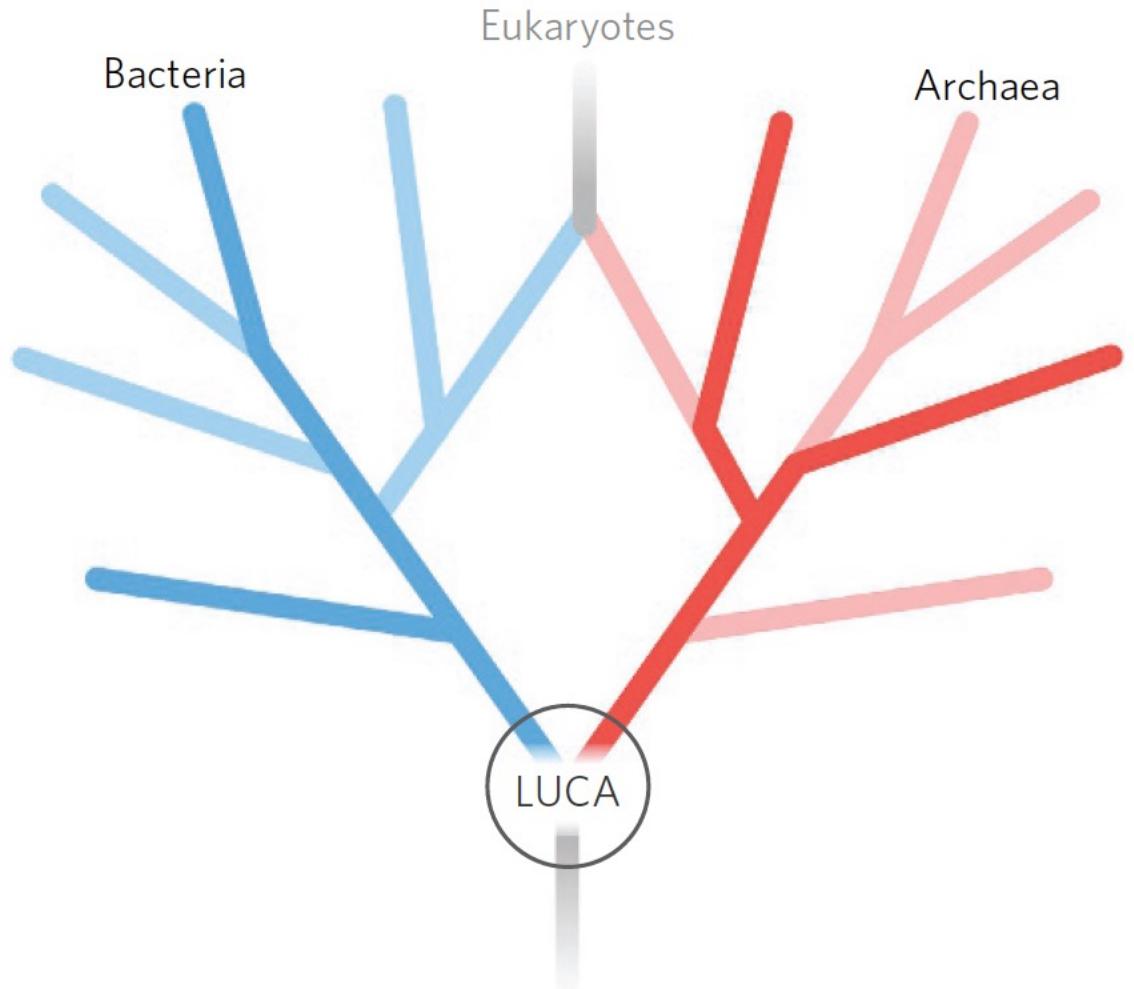
A gastroenterologist was convicted of attempted second-degree murder by injecting his former girlfriend with blood or blood-products obtained from an HIV type 1 (HIV-1)-infected patient under his care. Phylogenetic analyses of HIV-1 sequences were admitted and used as evidence in this case, representing the first use of phylogenetic analyses in a criminal court case in the United States. Phylogenetic analyses of HIV-1 reverse transcriptase and env DNA sequences isolated from the victim, the patient, and a local population sample of HIV-1-positive individuals showed the victim's HIV-1 sequences to be most closely related to and nested within a lineage comprised of the patient's HIV-1 sequences. This finding of paraphyly for the patient's



Cataloguing the diversity of life: when did it happen?

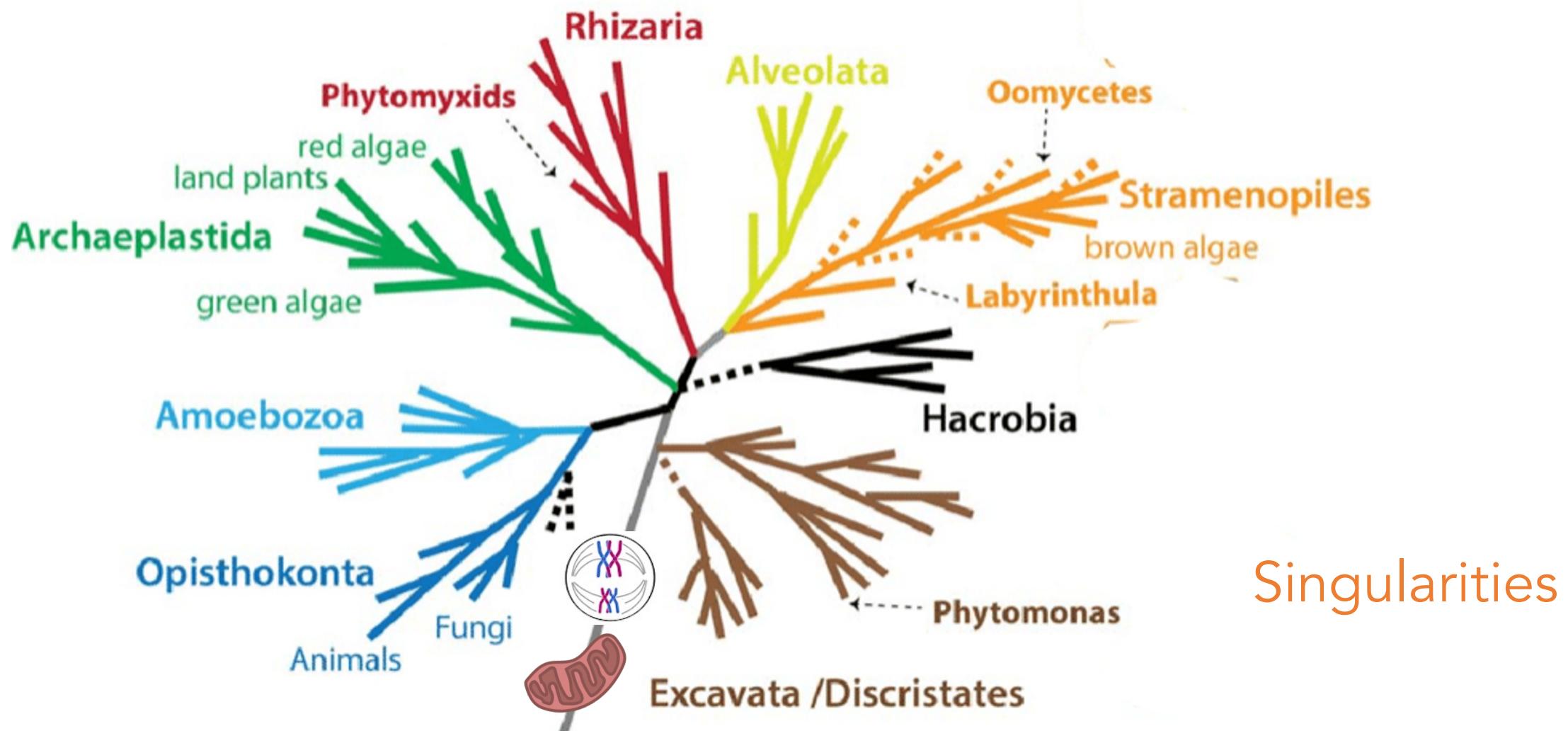


Identifying & understanding evolutionary transitions

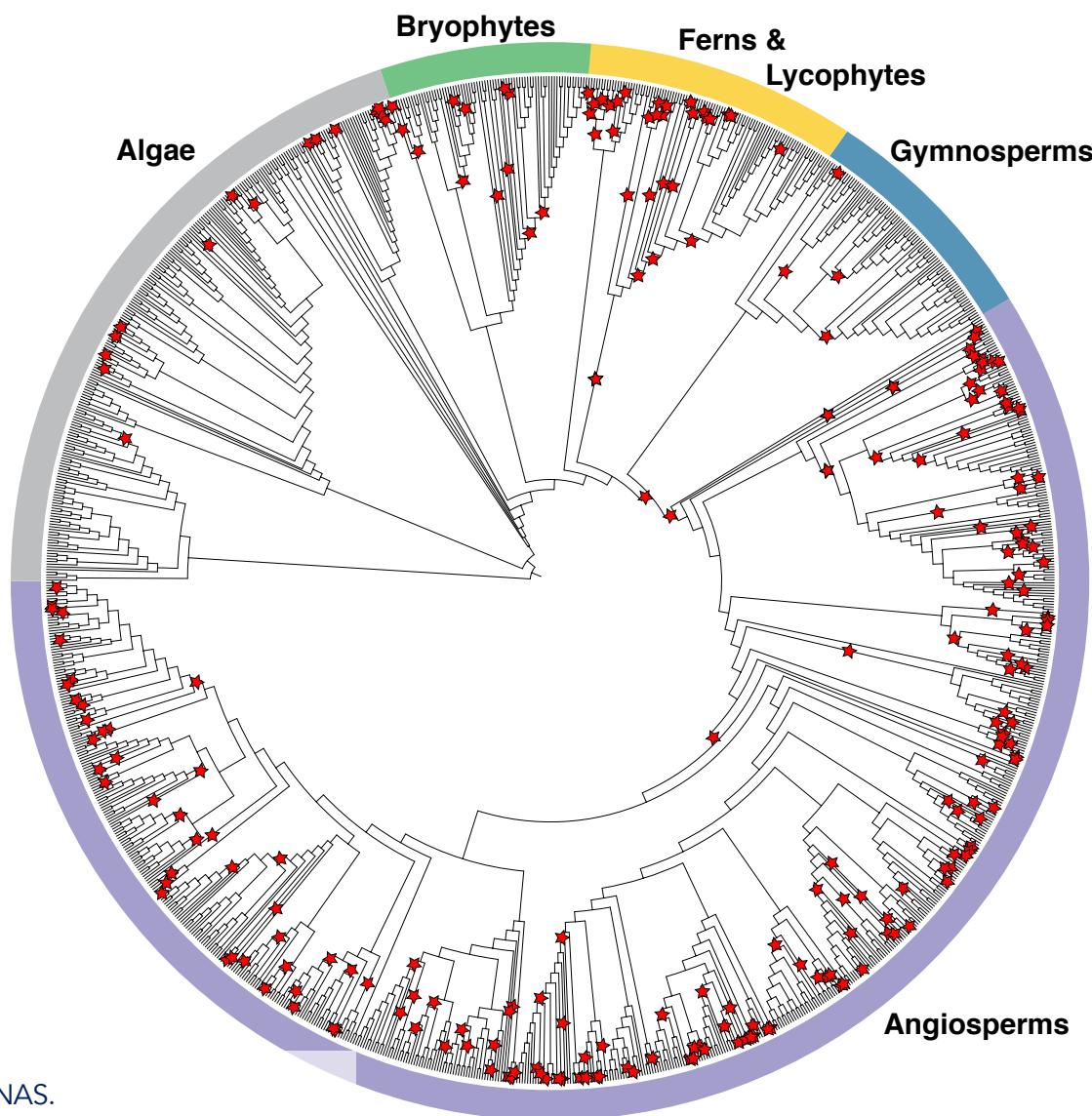


Ancestral State Reconstruction
Last Universal Common Ancestor

Identifying & understanding evolutionary transitions

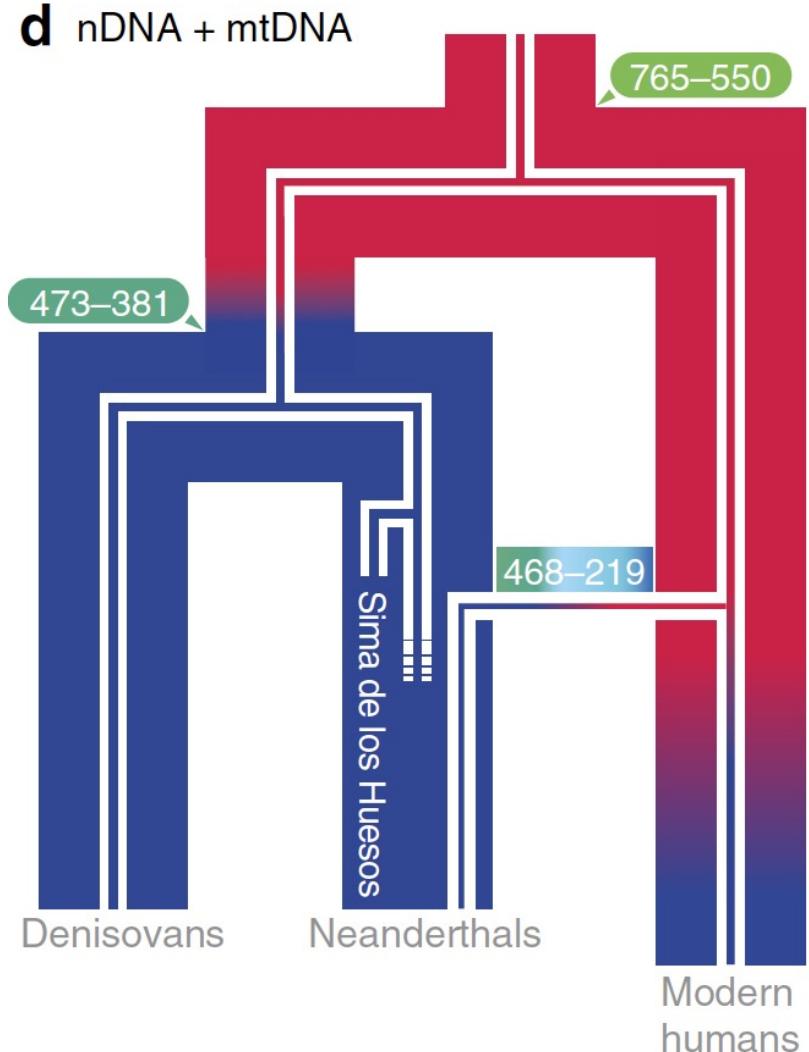


Identifying & understanding evolutionary transitions



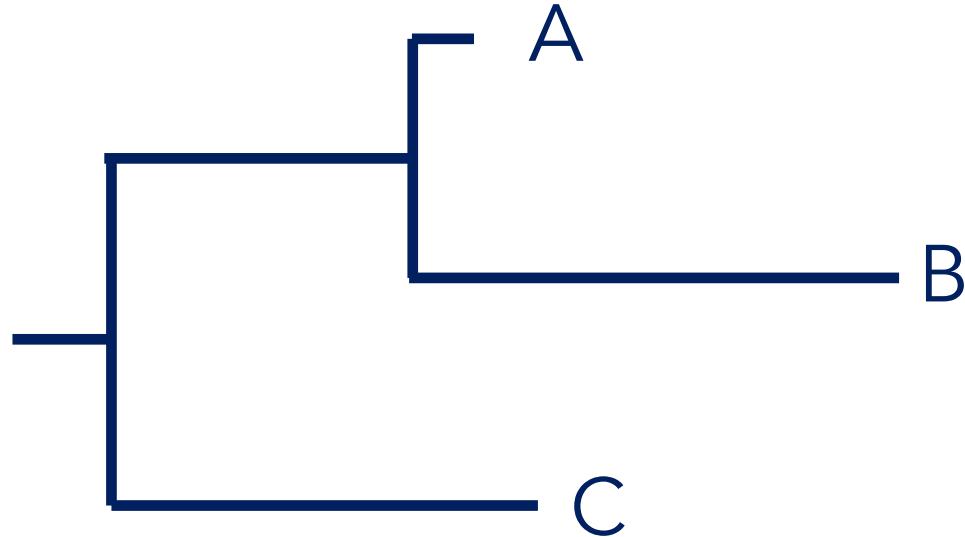
Preponderance/
consequences of mutations

Identifying & understanding evolutionary transitions



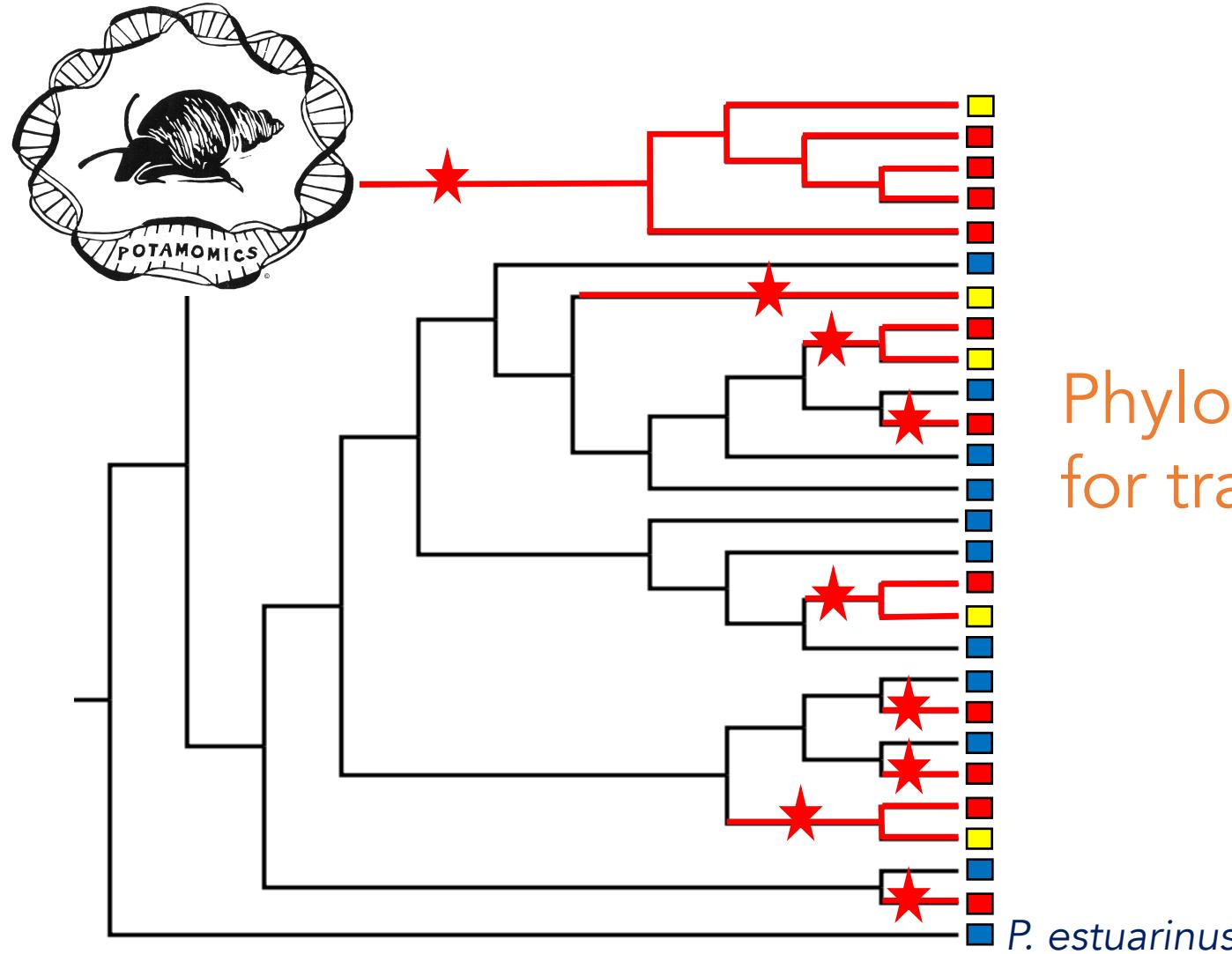
Signatures of
hybridization/introgression

Understanding evolutionary processes & mechanisms

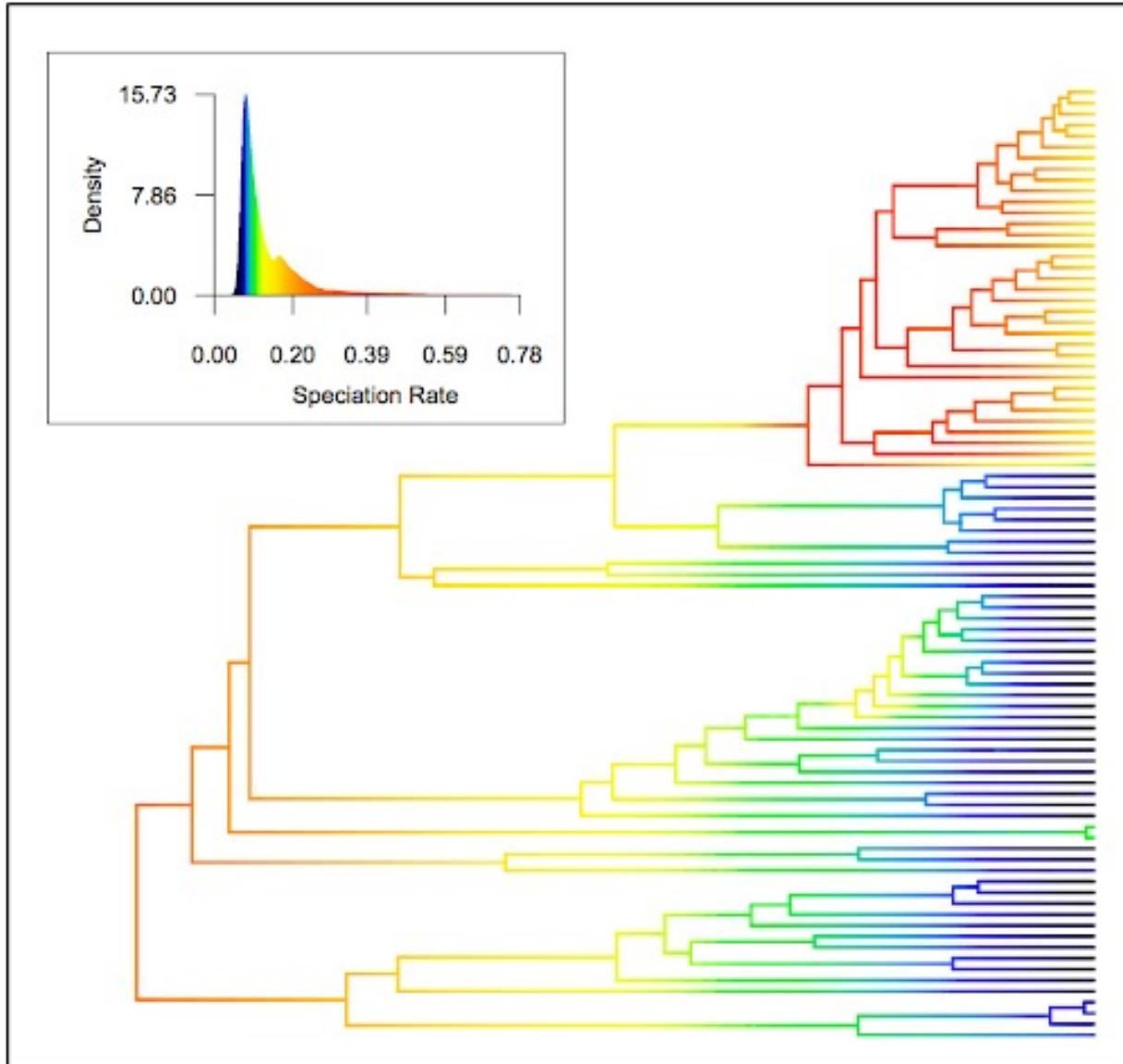


Evolutionary rates, type and mode of natural selection

Understanding evolutionary processes & mechanisms



Understanding evolutionary processes & mechanisms



Speciation/Extinction Rates

Next up: methods for tree inference

HW: Tree Thinking (Due 2/20)

Distance-based methods

- NJ, UPGMA

I think

Parsimony

Maximum Likelihood

Bayesian

