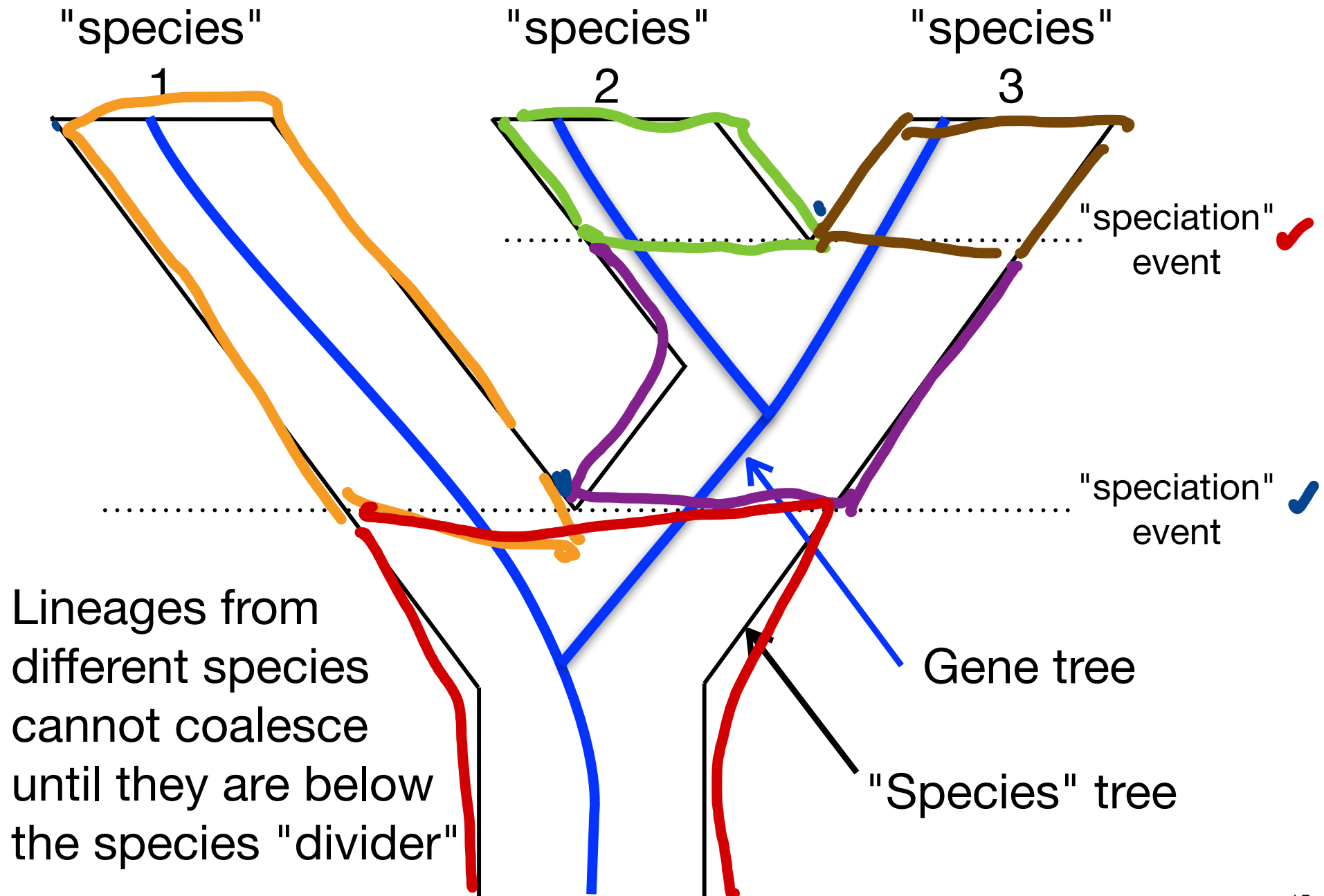
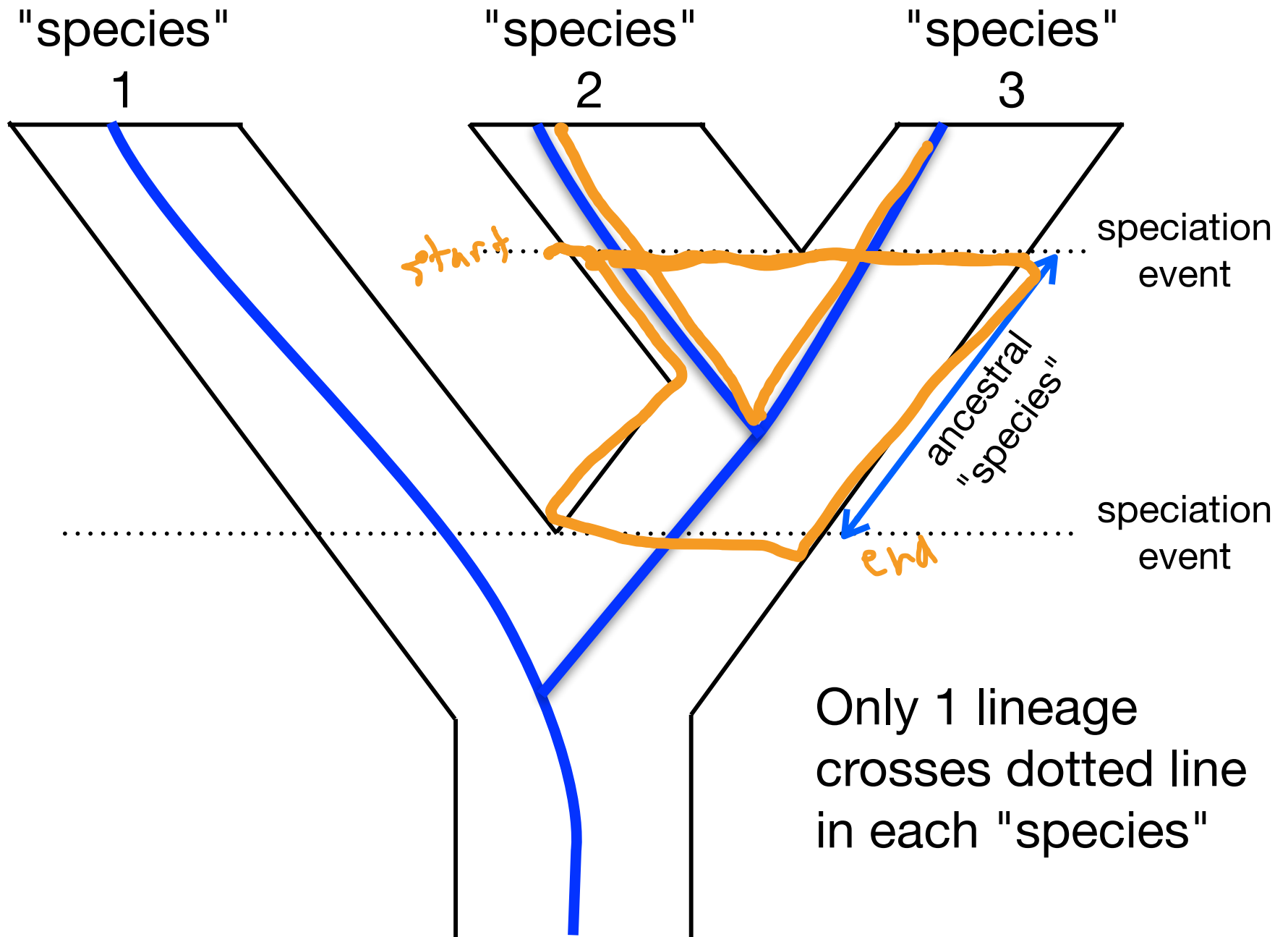


Deep coalescence and ILS

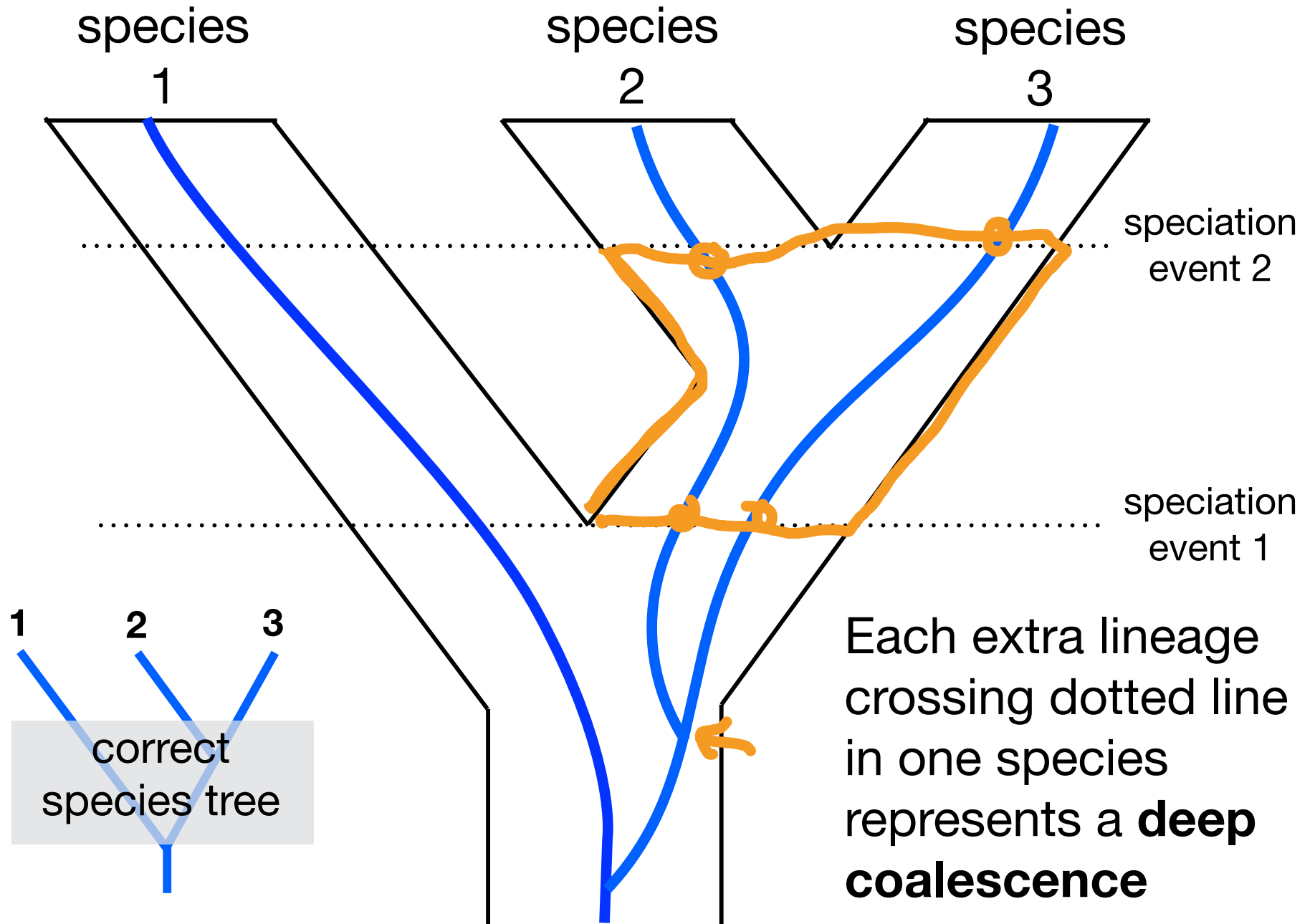
Multispecies coalescent

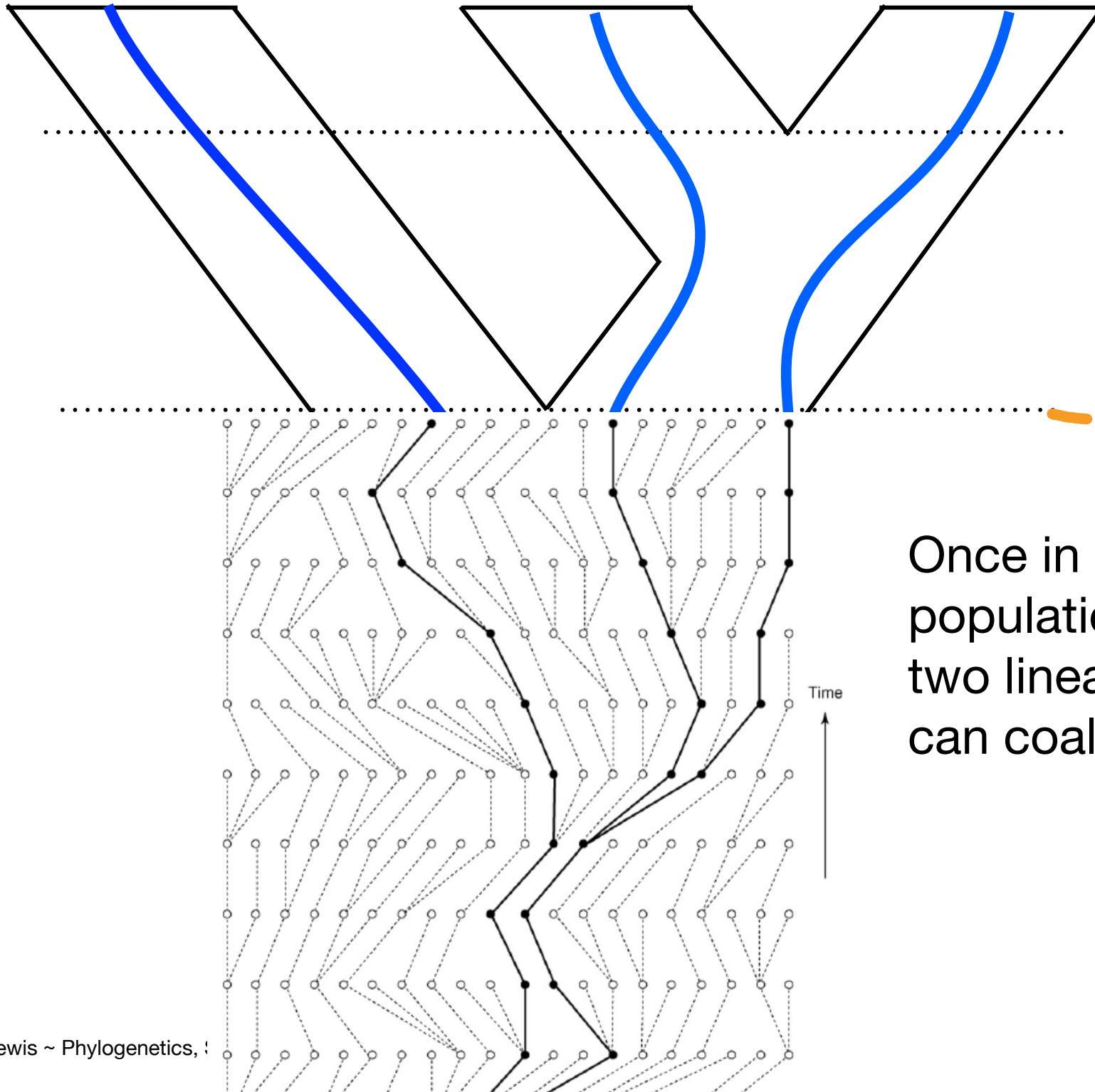


Shallow coalescence



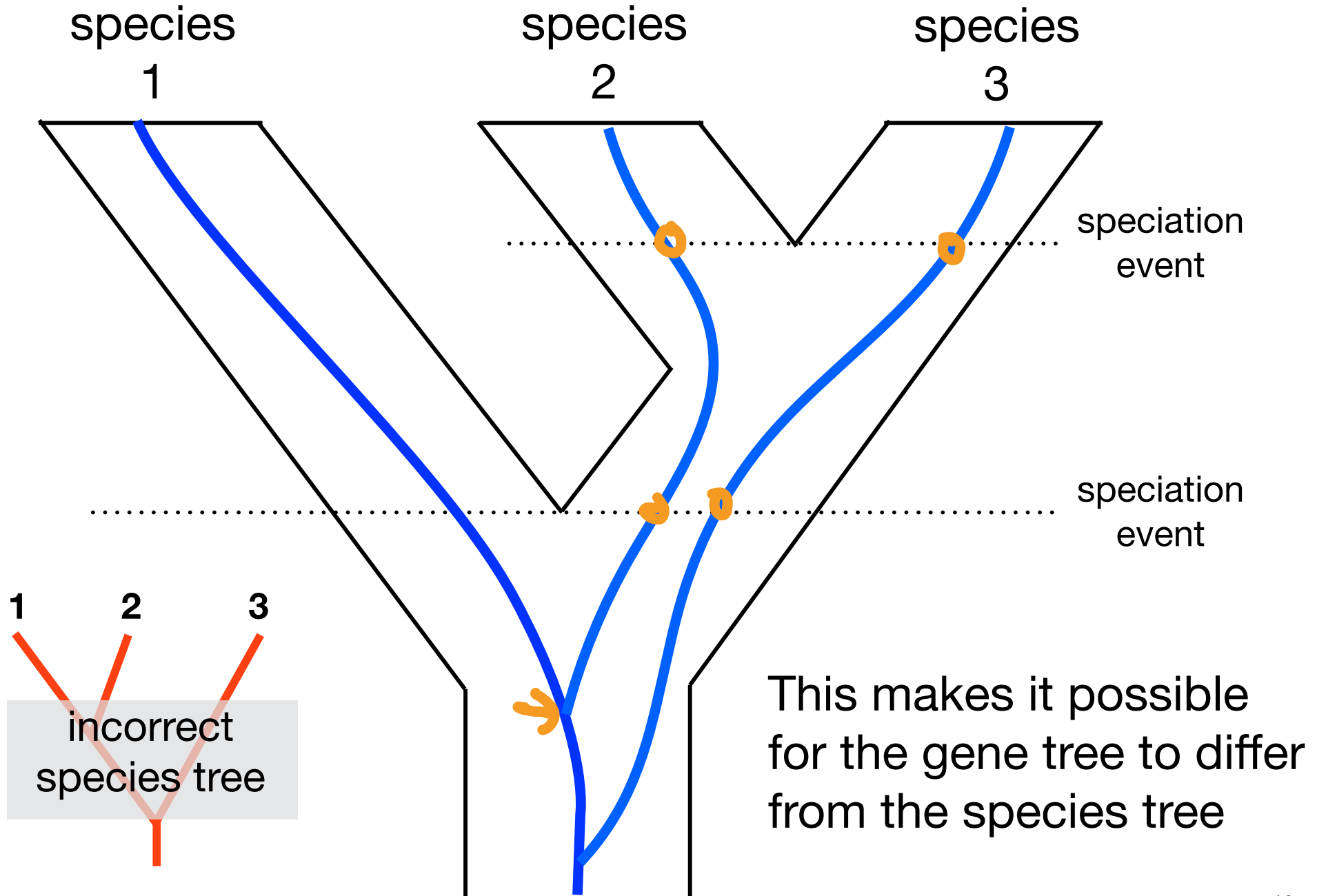
Deep coalescence



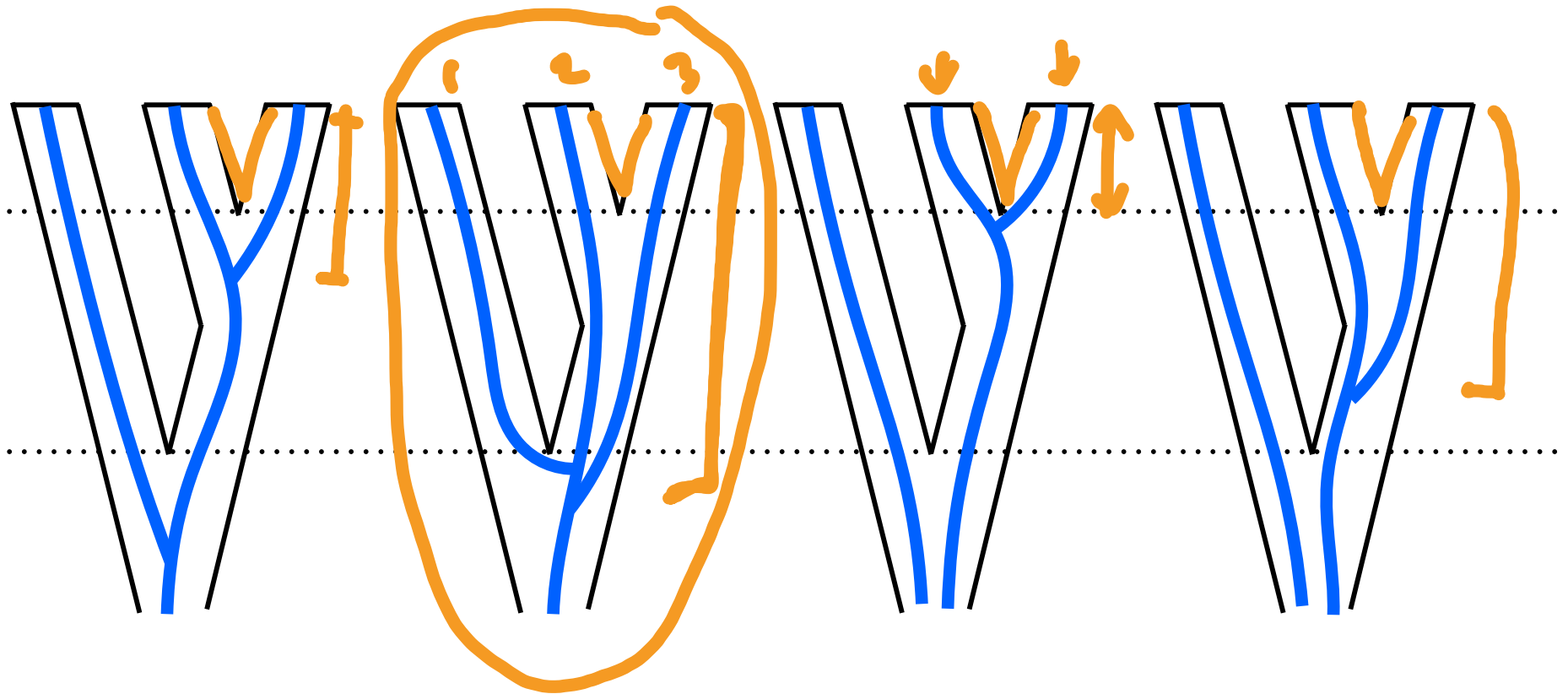


Once in a single population, any two lineages can coalesce...

Deep coalescence



The power of many genes



"...species tree[s] are not the result of a consensus process - but of a 'reverse auction' - where the lowest bidder sets the limit."
-- Heled and Drummond, 2010 (MBE 27:577)

*BEAST multispecies coalescent

Can estimate a species tree (S) given sequence data (D) from multiple genes:

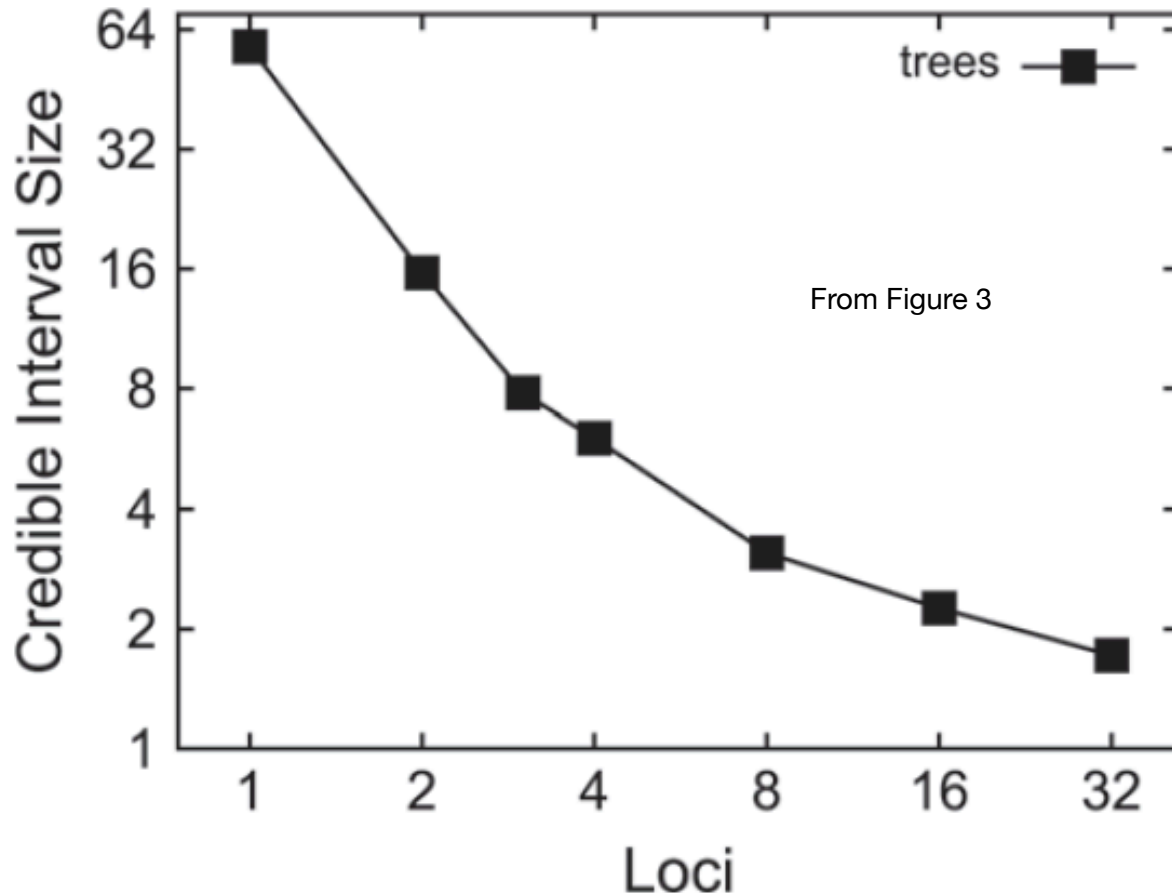
$$p(S, \mathbf{G}, \theta | \mathbf{D}) \propto p(\mathbf{D} | \mathbf{G}) p(\mathbf{G} | S, \theta) p(\theta) p(S)$$

parameters *posterior* *data* *likelihood* *prior*

$$\frac{p(\theta | \mathbf{D})}{p(\mathbf{D})} = \frac{p(\mathbf{D} | \theta) p(\theta)}{p(\mathbf{D})}$$

posterior *likelihood* *prior*

*BEAST multispecies coalescent

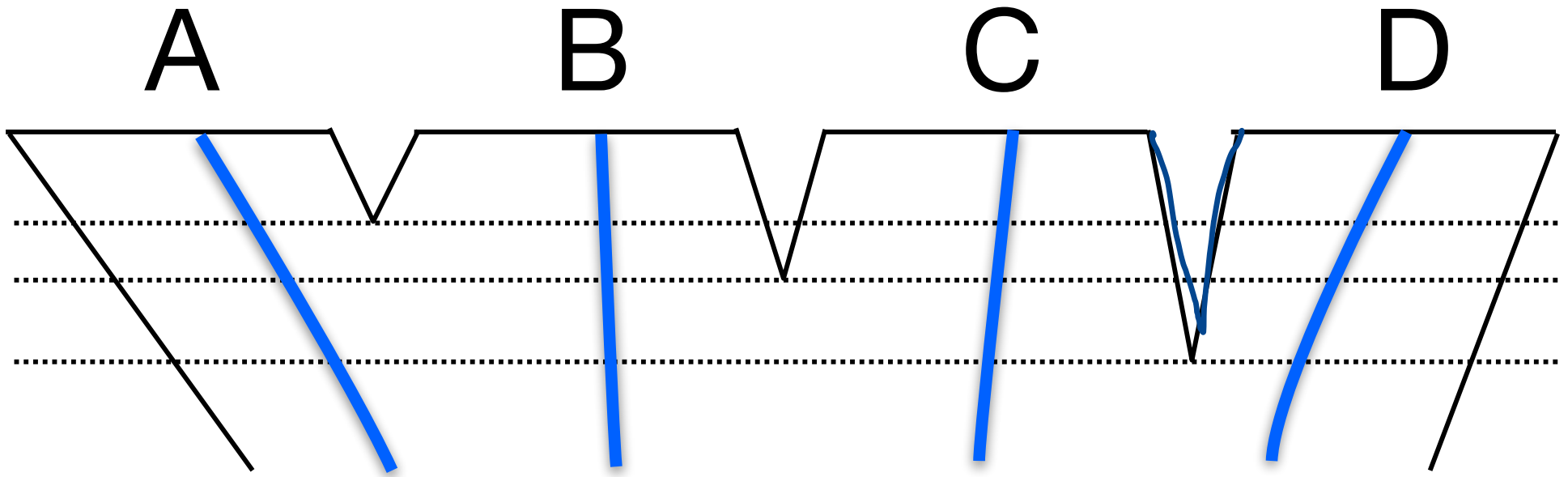


*BEAST vs concatenation

Heled and Drummond (2010) compared their (*BEAST) method with concatenation, which fails when there is a lot of deep coalescence.

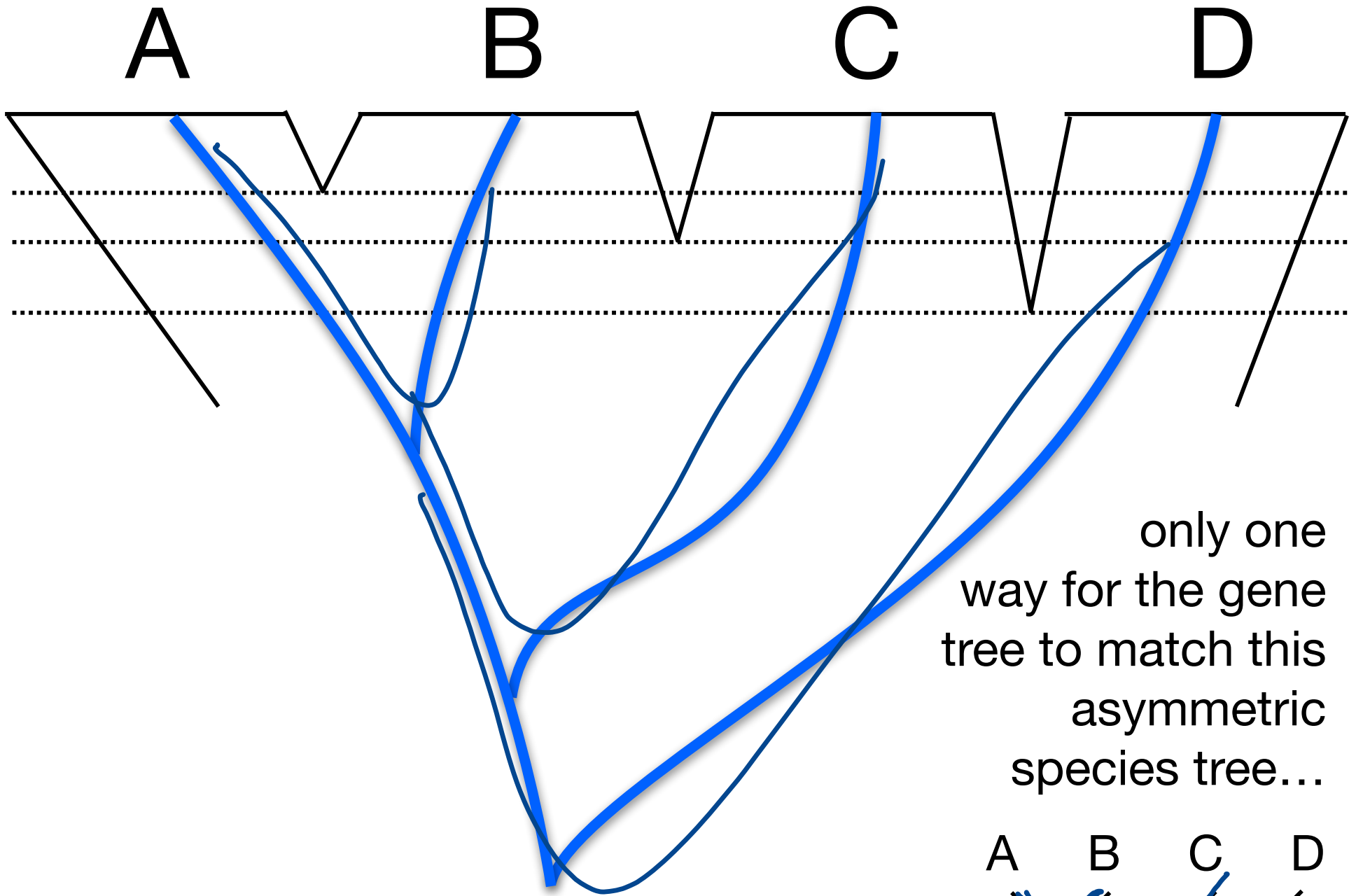
From Table 2	Topology inside 95%	Topology 95% size	Speciation time inside 95%	Speciation time 95% size
*BEAST	97	11.8	96	1.5
<u>Concatenation</u>	9	1.4	0.7	5.3

The anomaly zone

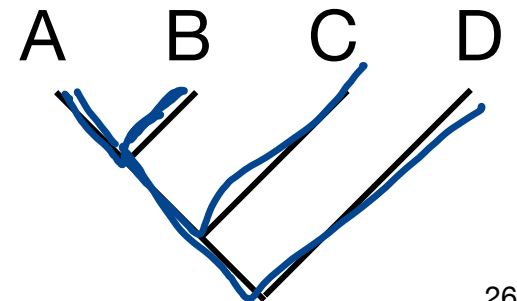


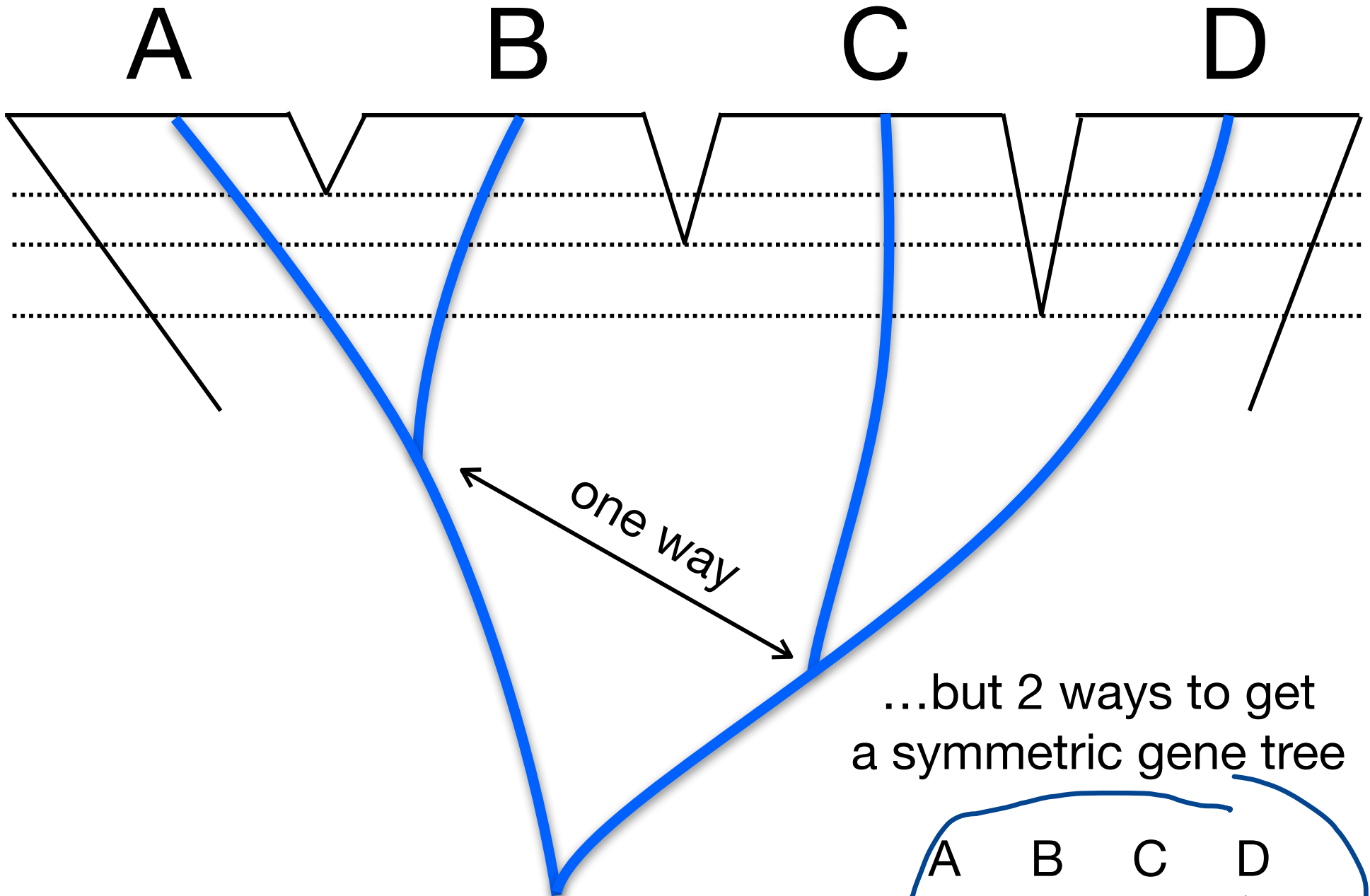
Anything can happen now!

The Anomaly Zone

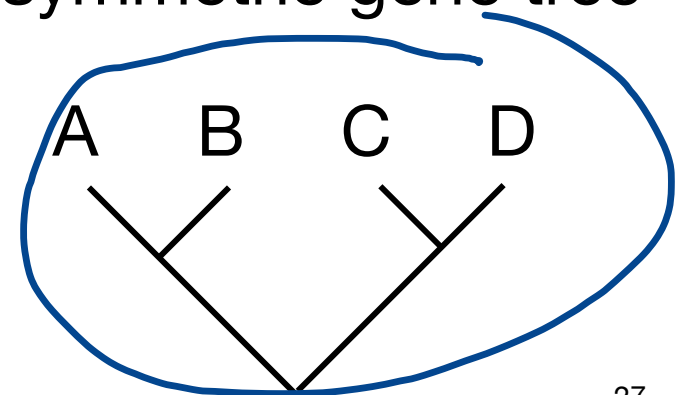


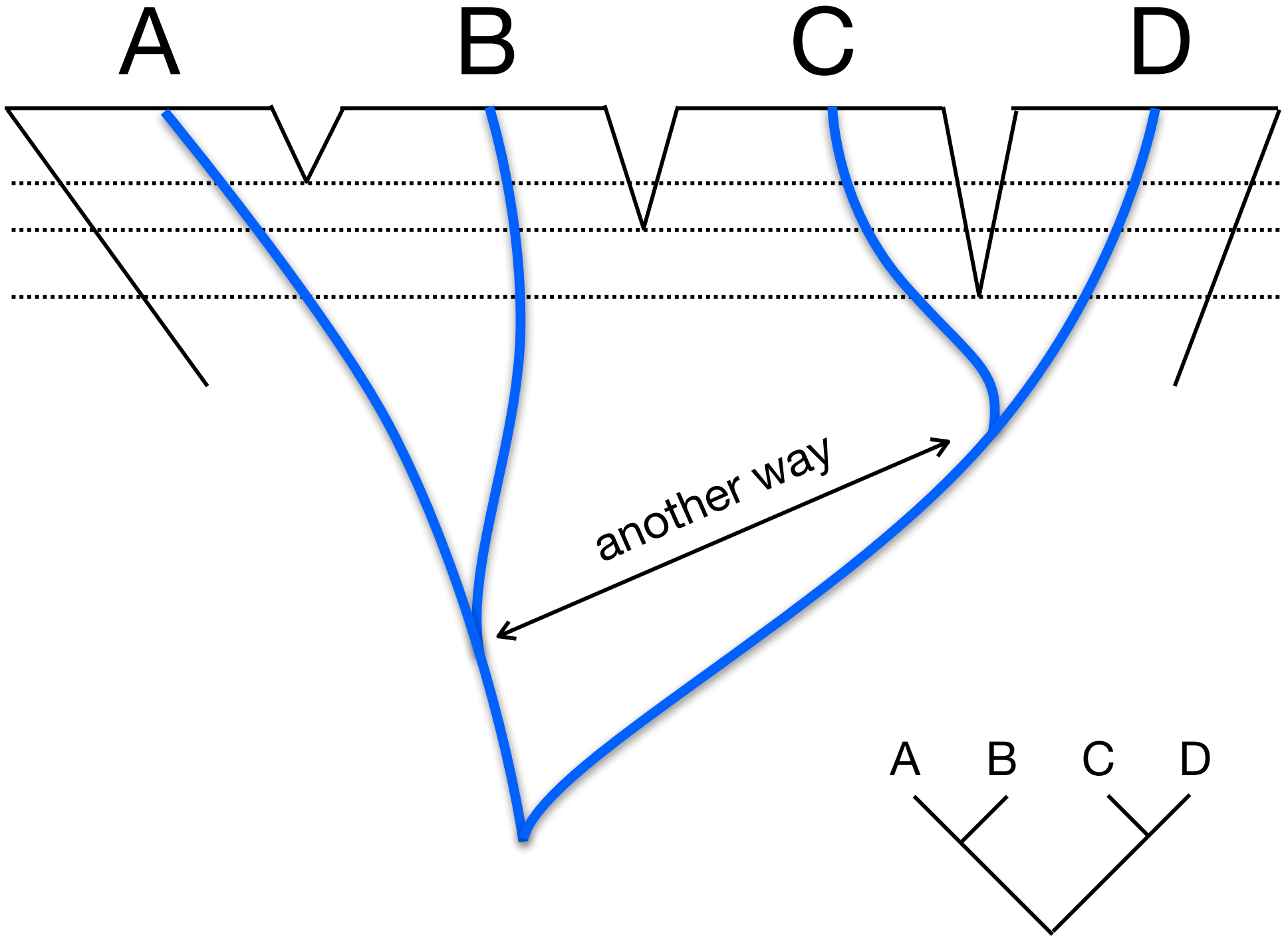
only one way for the gene tree to match this asymmetric species tree...



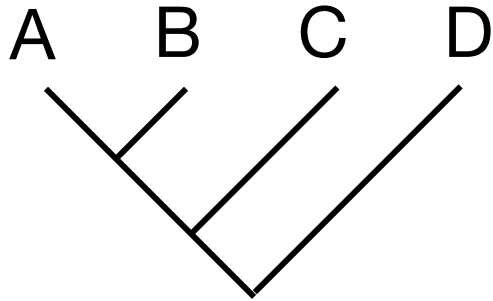


...but 2 ways to get a symmetric gene tree

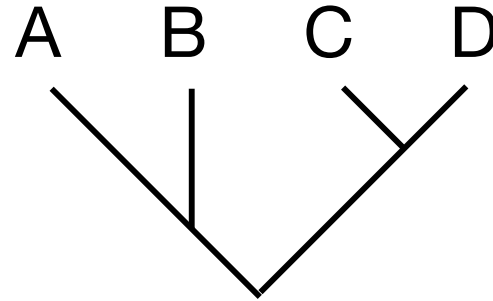




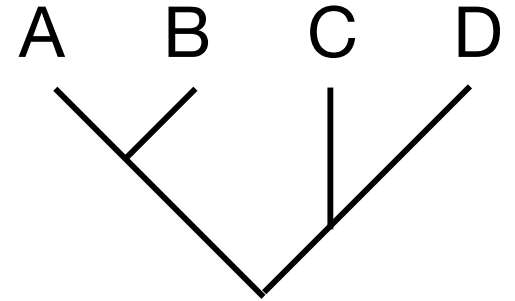
The Anomaly Zone



1 way to get gene tree that matches species tree



2 ways to get symmetric gene tree that does not match species tree



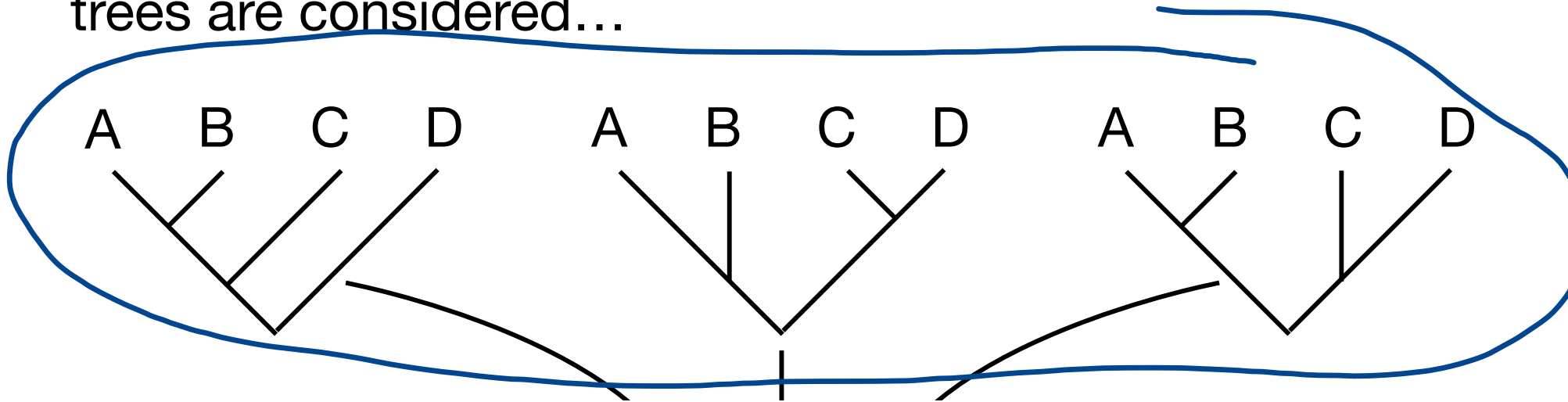
Anomalous gene trees (AGTs) are gene trees that are more probable than the species tree

Degnan and Rosenberg (2006)

Quartet-based species tree methods

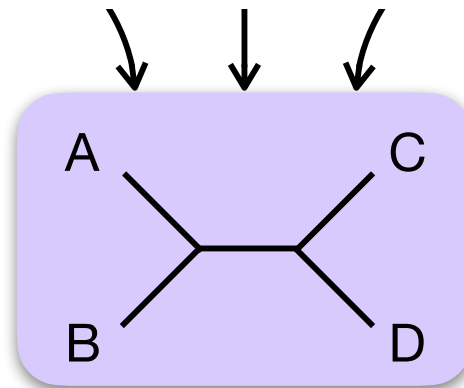
ASTRAL

For 4 taxa, anomalous gene trees only exist if *rooted* trees are considered...



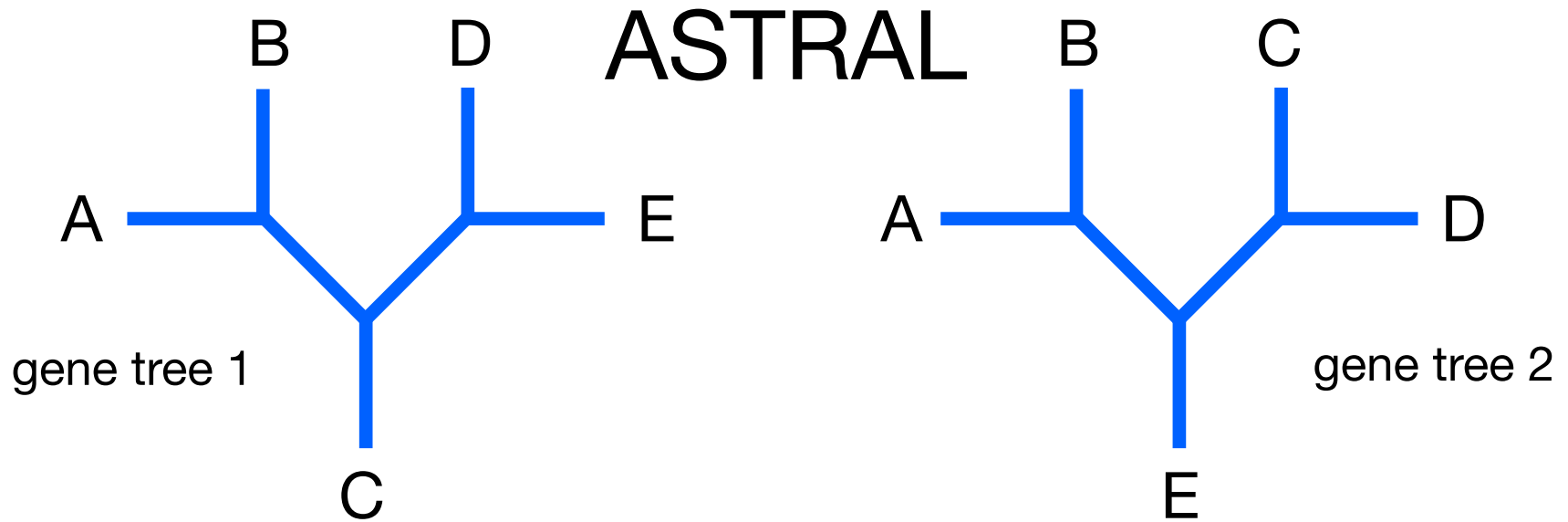
These are all the same unrooted tree

ASTRAL takes advantage of this fact, building a complete species tree from unrooted **quartet subtrees** nested within gene trees

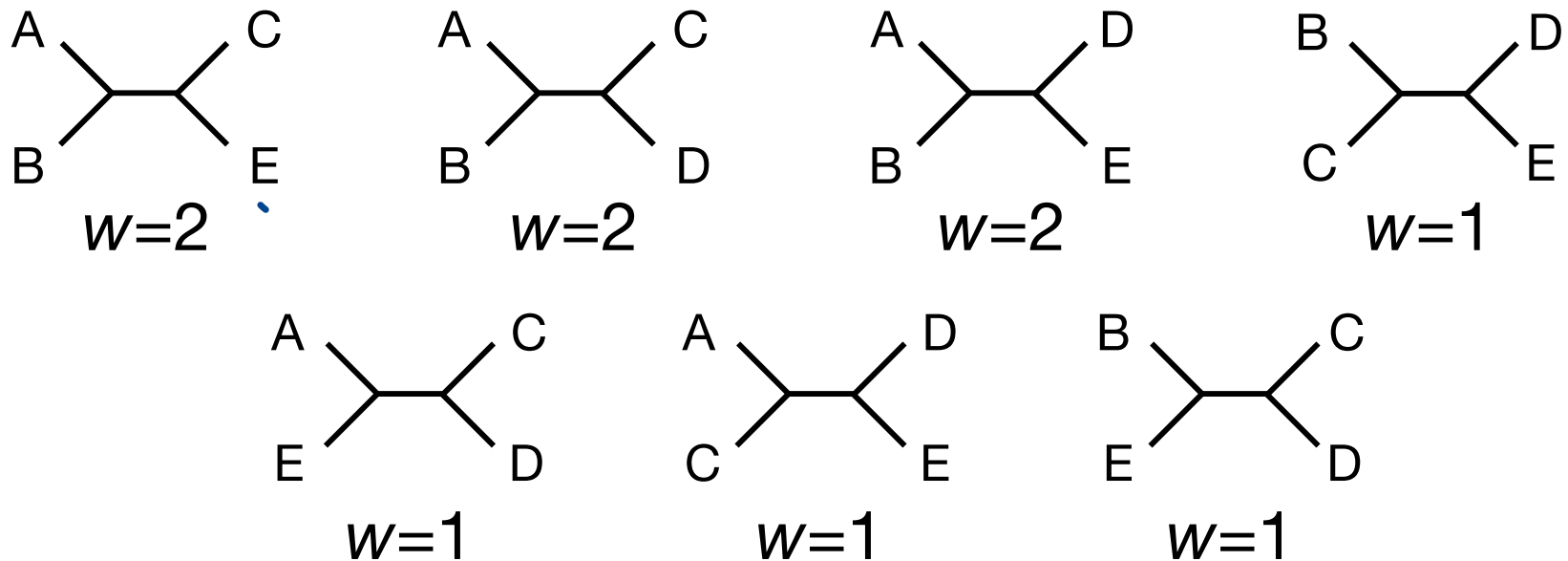


ASTRAL is a **supertree method** that is statistically consistent in the face of substantial incomplete lineage sorting

Mirarab and Warnow (2015)

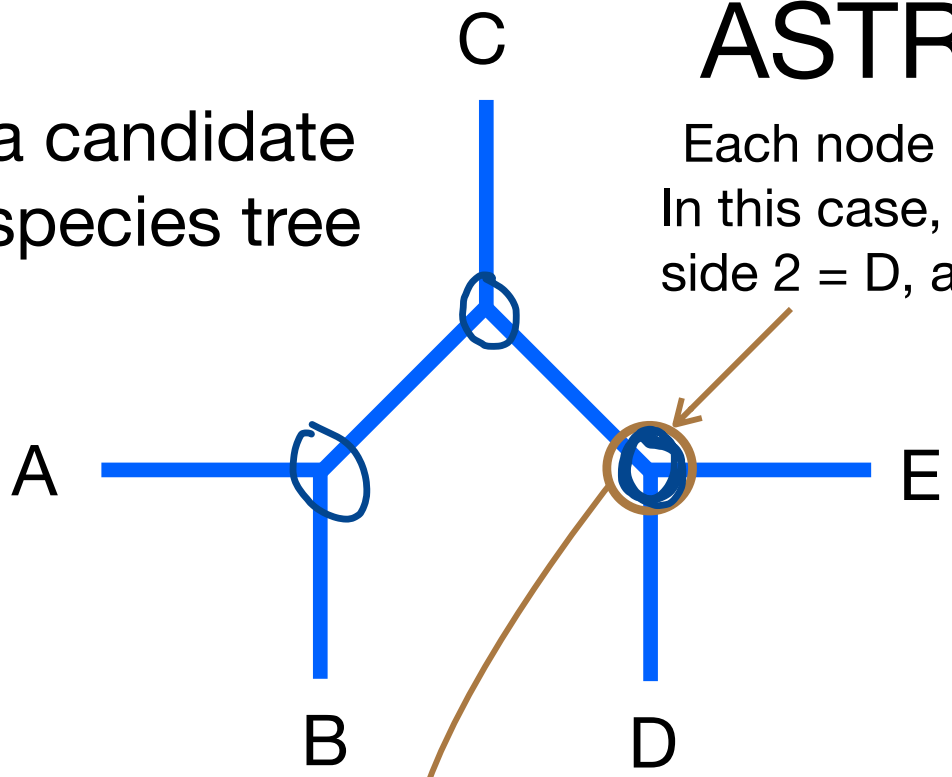


weight (w) = number of gene trees in which quartet is found



ASTRAL

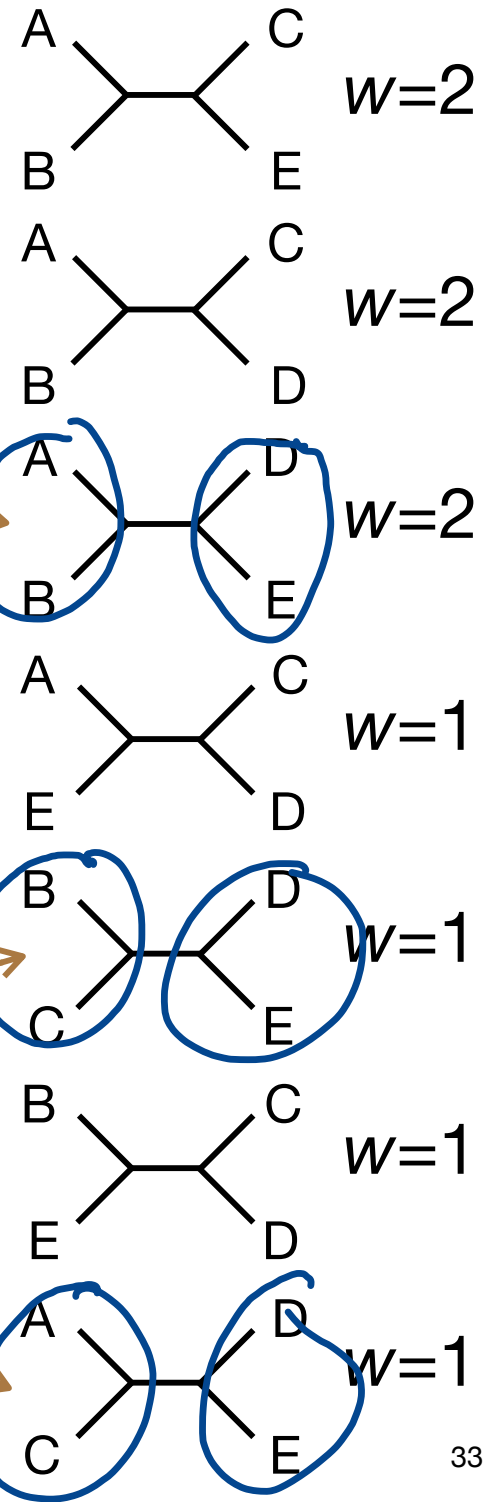
a candidate species tree



Each node has 3 "sides"
 In this case, side 1 = ABC,
 side 2 = D, and side 3 = E)

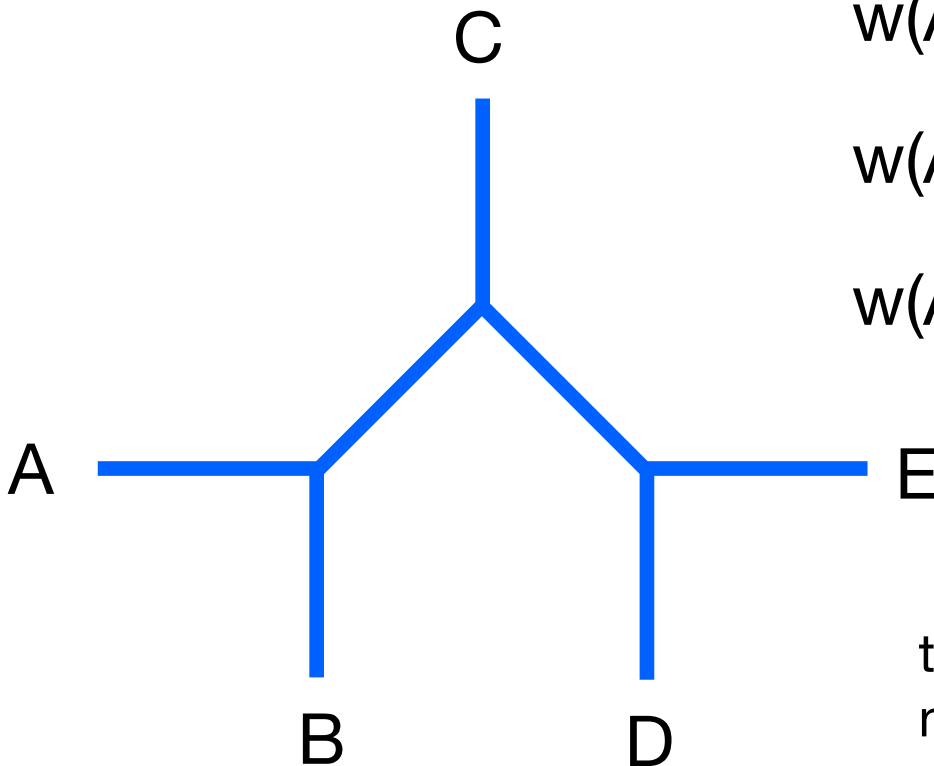
weight for each node in candidate tree = sum of weights for all quartets that map to the node

$$w(ABC|D|E) = 2+1+1$$



ASTRAL

Total support for candidate species tree
 $= (4+6+6)/2 = 8$



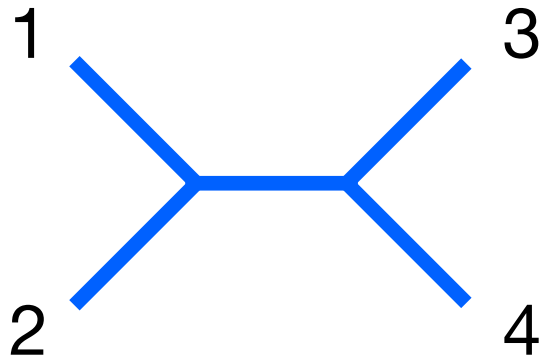
$$w(ABC|D|E) = 2+1+1 = 4$$

$$w(AB|C|DE) = 2+2+1+1 = 6$$

$$w(A|B|CDE) = 2+2+2 = 6$$

total weight is half the sum of the
node weights because each
quartet gets counted twice

STUPPED HERE 2024-03-19



SVDQuartets

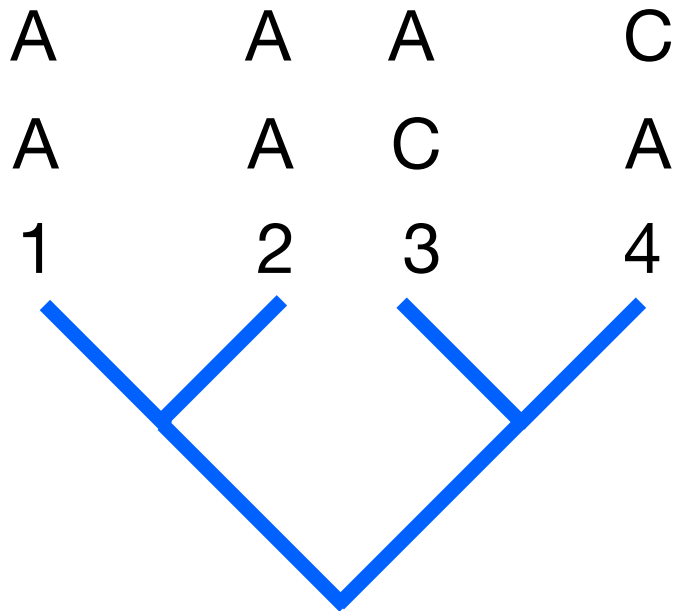
Quartet method that consults the original sequences rather than depending on gene trees being correctly estimated

```

taxon1 AAAGATTACAGGTTGACTTATTACACCCCGGAG...
taxon2 AAAGATTATCGACTGACTTATTACACCCCGAA...
taxon3 AAAGATTACAGATTA ACTTATTATACTCCTGAA...
taxon4 AAAGATTATAAATTGACTTACTACACCCCGGAG...
  
```

$$\text{flat}_{12|34} = \begin{matrix} & \text{AA} & \text{AC} & \text{AG} & \text{AT} & \text{CA} & & \\ \text{AA} & p_{AAAA} & p_{AAAC} & p_{AAAG} & p_{AAAT} & p_{AACA} & \cdots & \\ \text{AC} & p_{ACAA} & p_{ACAC} & p_{ACAG} & p_{ACAT} & p_{ACCA} & \cdots & \\ \text{AG} & p_{AGAA} & p_{AGAC} & p_{AGAG} & p_{AGAT} & p_{AGCA} & \cdots & \\ \text{AT} & p_{ATAA} & p_{ATAC} & p_{ATAG} & p_{ATAT} & p_{ATCA} & \cdots & \\ \text{CA} & p_{CAAA} & p_{CAAC} & p_{CAAG} & p_{CAAT} & p_{CACA} & \cdots & \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \end{matrix} \quad \begin{matrix} \text{The flat matrix for} \\ \text{any quartet tree} \\ \text{stores counts of sites} \\ \text{with patterns that} \\ \text{match the quartet} \\ \text{tree} \end{matrix}$$

SVDQuartets



Symmetries among columns in the flat matrix are expected, even if there is ILS

For example p_{AAAC} should approximately equal p_{AACA} (and the same is true for every pair in these two columns)

$$\text{flat}_{12|34} = \begin{matrix} & \text{AA} & \text{AC} & \text{AG} & \text{AT} & \text{CA} & \cdots \\ \text{AA} & p_{AAAA} & p_{AAAC} & p_{AAAAG} & p_{AAAAT} & p_{AAACA} & \cdots \\ \text{AC} & p_{ACAA} & p_{ACAC} & p_{ACAG} & p_{ACAT} & p_{ACCA} & \cdots \\ \text{AG} & p_{AGAA} & p_{AGAC} & p_{AGAG} & p_{AGAT} & p_{AGCA} & \cdots \\ \text{AT} & p_{ATAA} & p_{ATAC} & p_{ATAG} & p_{ATAT} & p_{ATCA} & \cdots \\ \text{CA} & p_{CAAA} & p_{CAAC} & p_{CAAG} & p_{CAAT} & p_{CACA} & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

Such symmetries reduce the **rank** of the flat matrix, which is a measure of how many columns (rows) are independent

SVDQuartets

- Flat matrix should have rank
 - 16 for quartets **not** in the species tree
 - ≤ 10 for quartets that **are** in the species tree under coalescent model
 - ≤ 4 for quartets that are in the true tree if true tree is common to all sites
- Evaluate all quartets (or a sample)
- Construct tree from quartets that have rank ≤ 10
- To use version recently added to PAUP*:
 - load concatenated data matrix
 - use svdquartets command