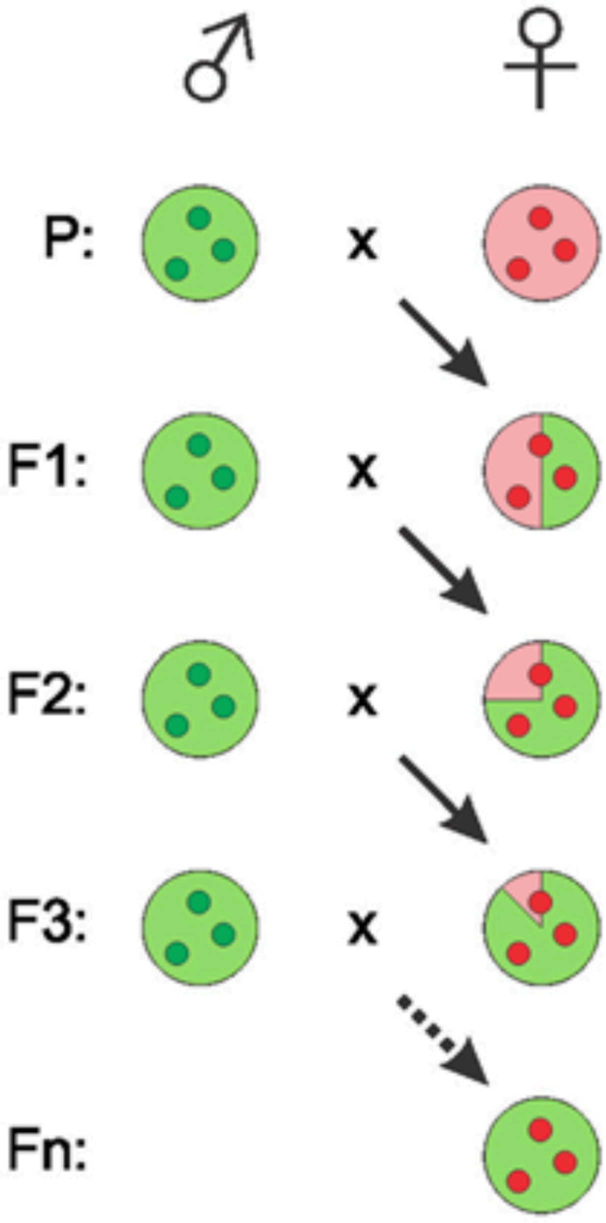
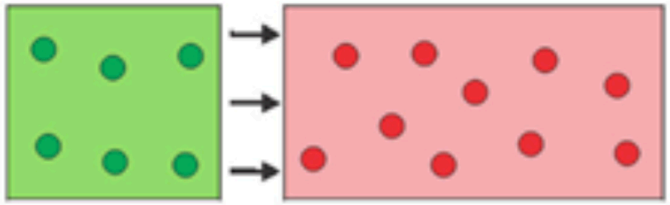
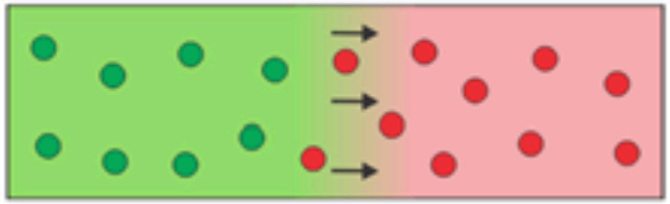
# Introgression, hybridization, and phylogenetic networks

# Mitochondrial introgression

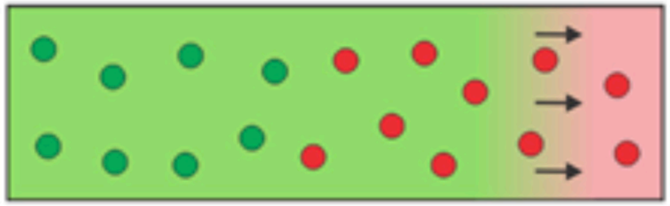




Green expands its range towards red



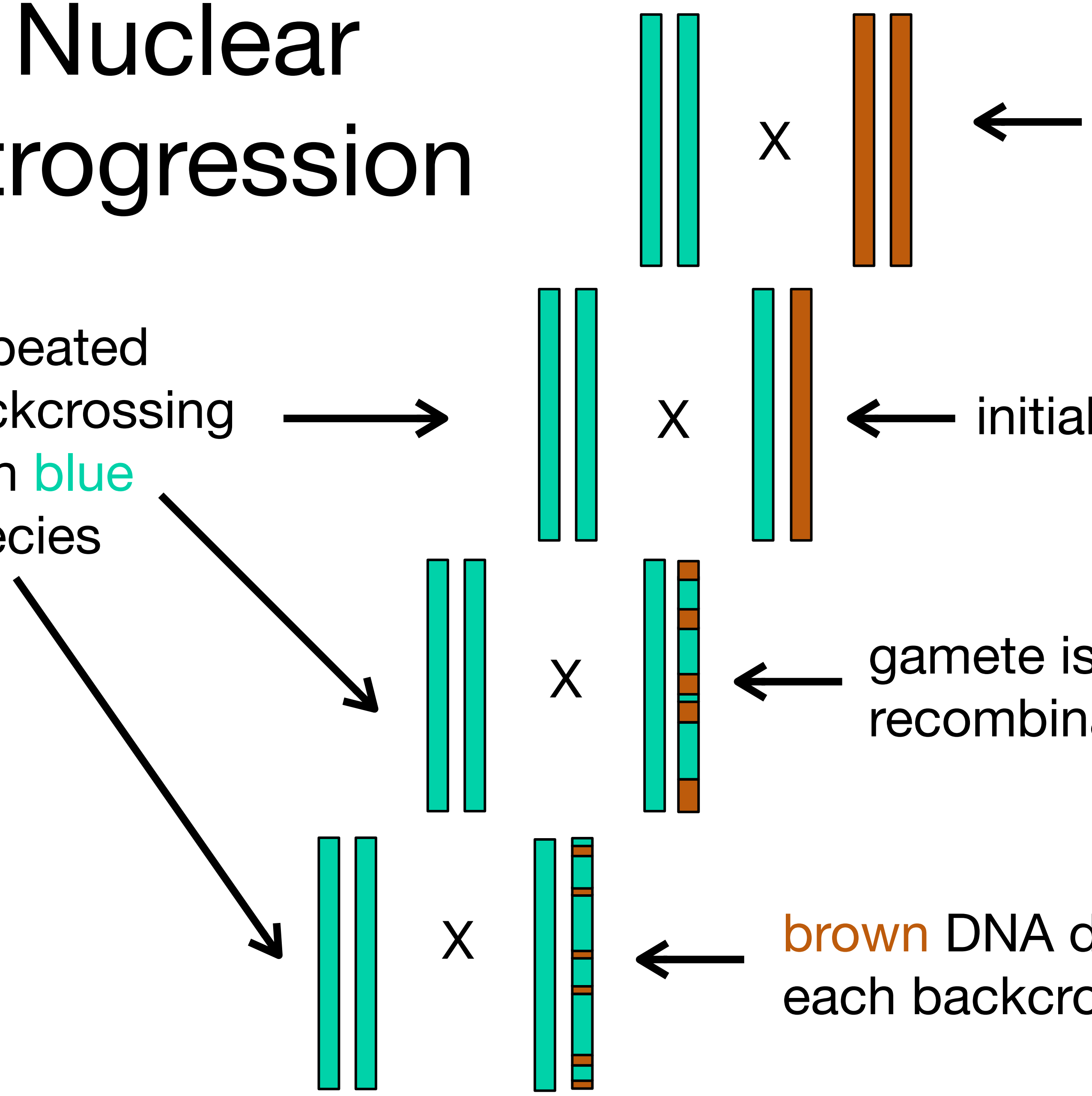
Red mtDNA introgresses into green



Red mtDNA surfs green expansion wave

Wielstra and Arntzen (2012)

# Nuclear introgression



blue and brown species hybridize

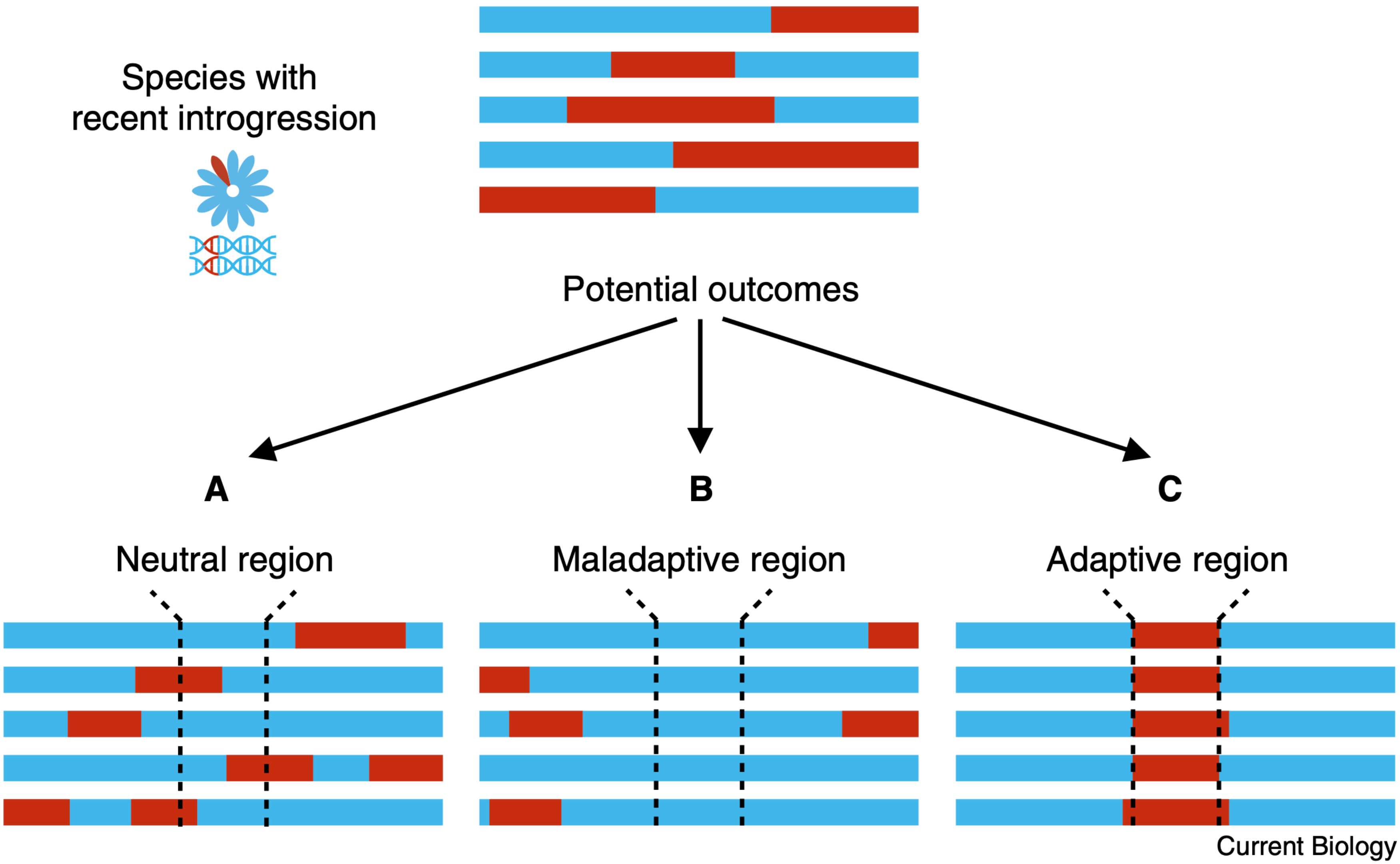
Repeated backcrossing with blue species

initial hybrid (50:50)

gamete is the result of recombination

brown DNA decreases with each backcrossing if *neutral*

# Nuclear introgression



Species with recent introgression

Potential outcomes

**A**

Neutral region

**B**

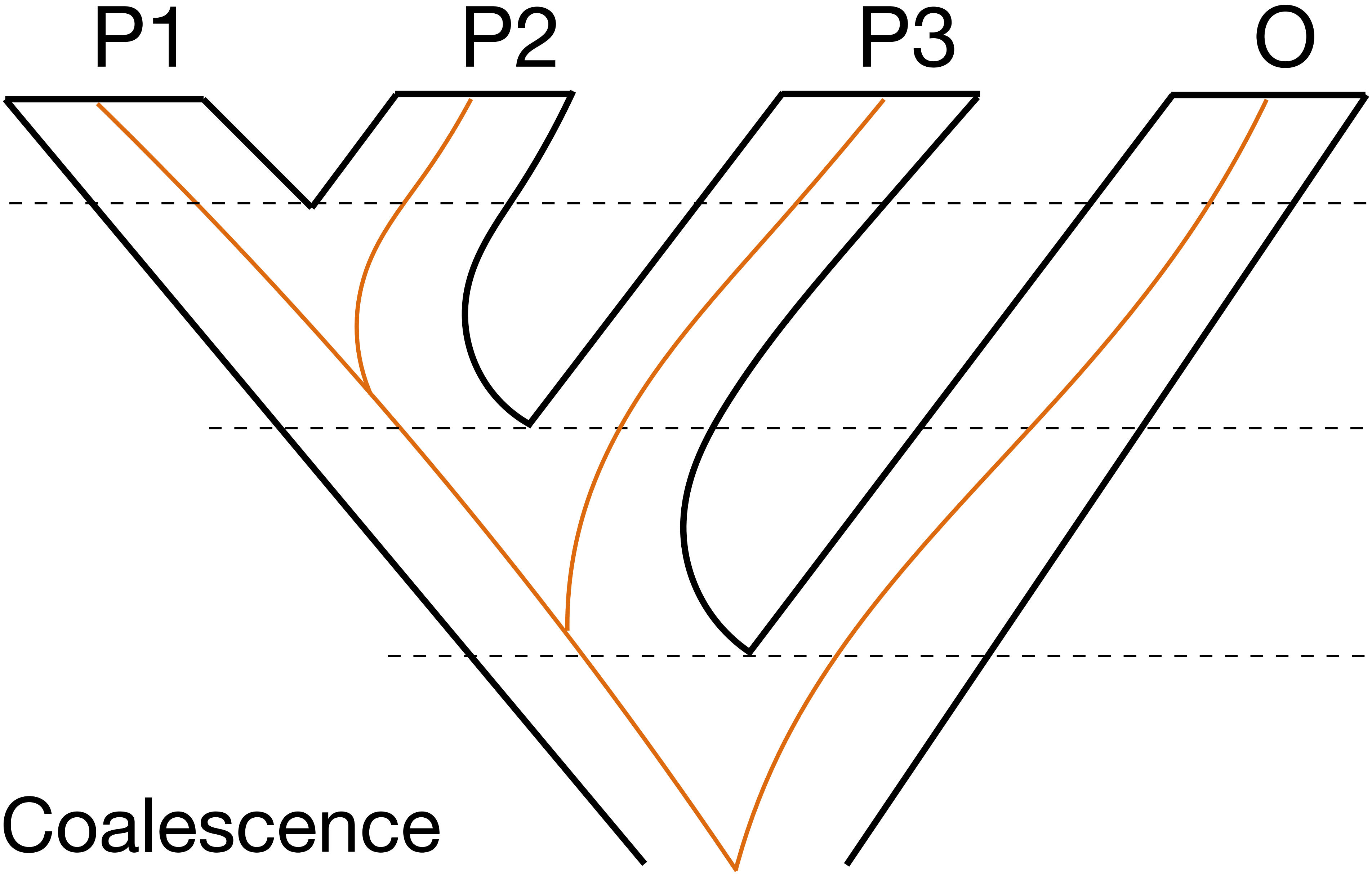
Maladaptive region

**C**

Adaptive region

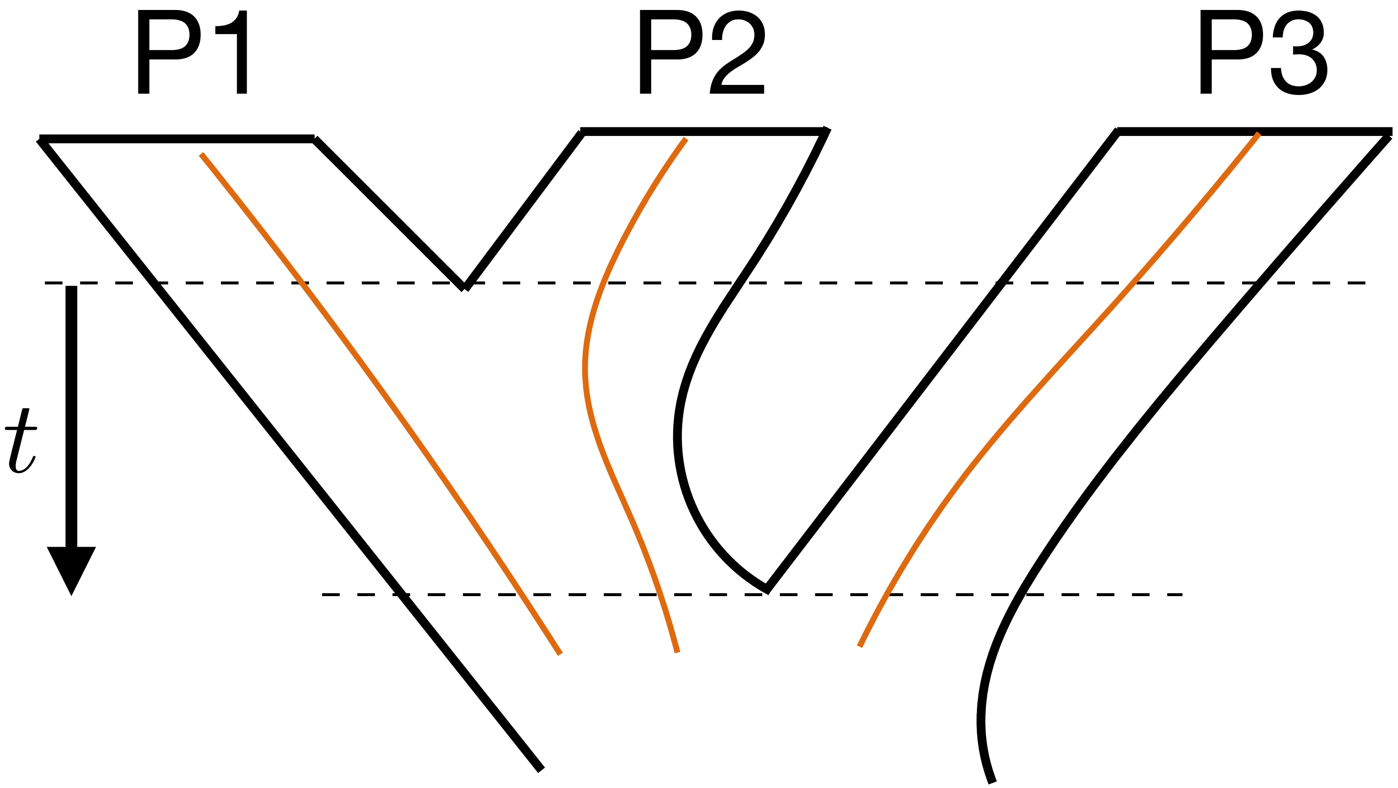
Current Biology

Aguillon et al. (2022)



Coalescence

# Probability of coalescence

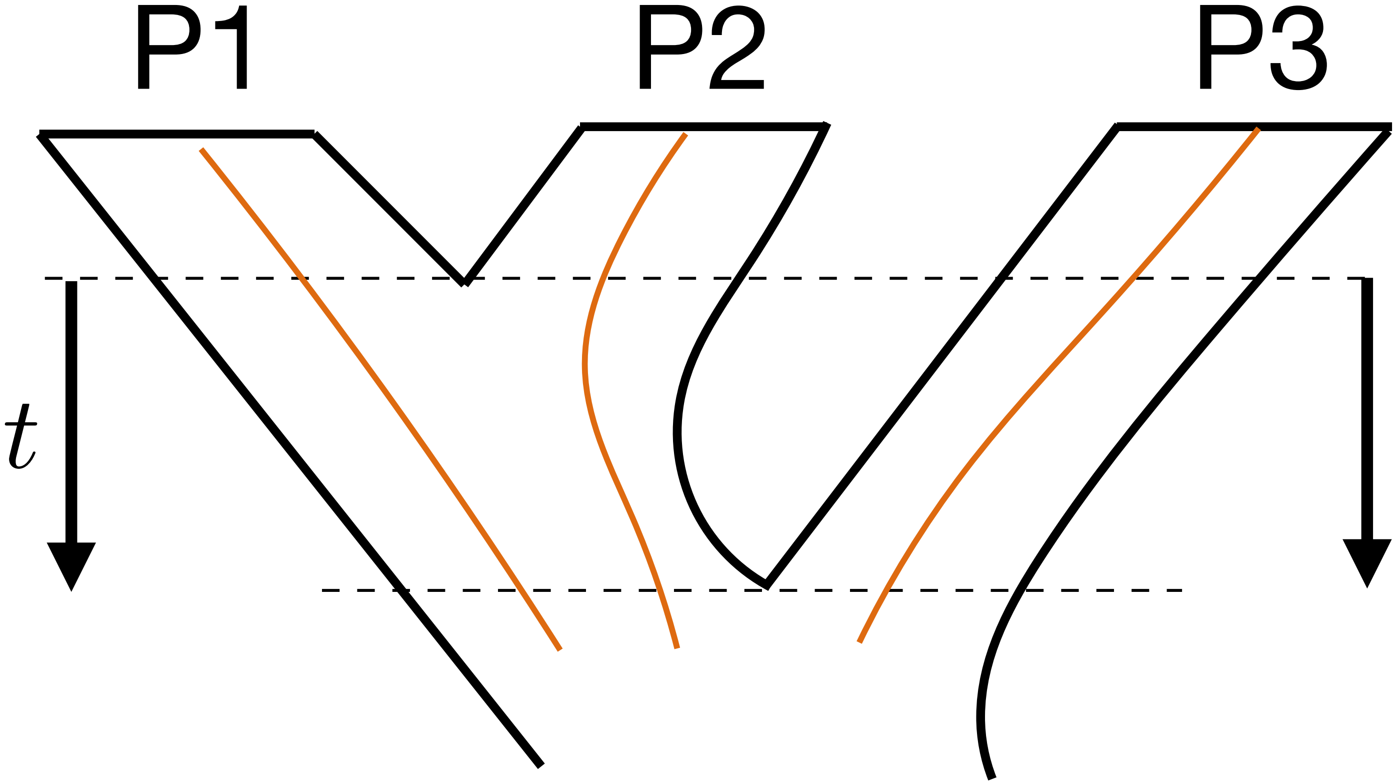


$$r = \frac{\binom{n}{2}}{2N_e}$$

$$p(\text{coal. by time } t) = 1 - e^{-rt}$$

$$p(\text{no coal. by time } t) = e^{-rt}$$

# Probability of coalescence

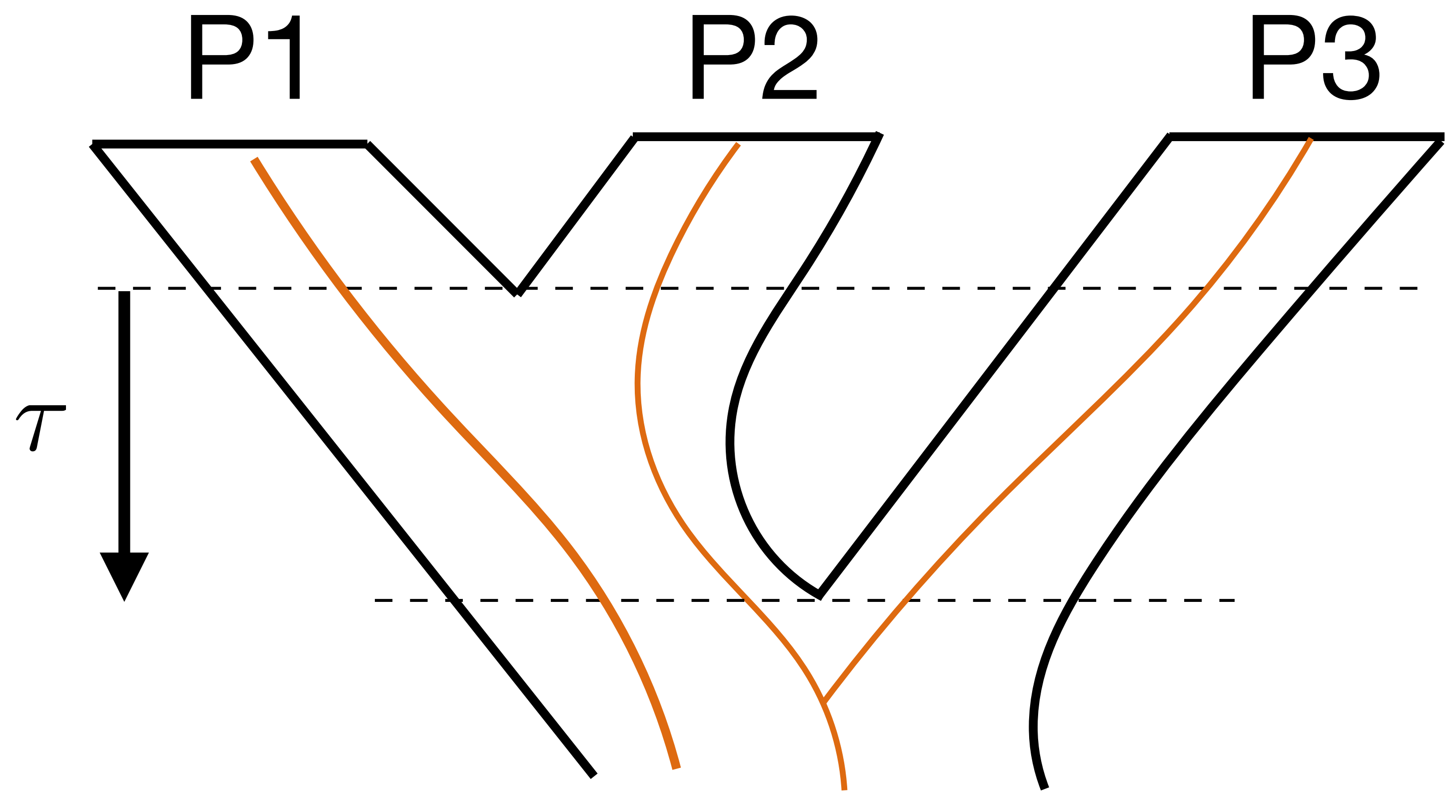


$$\tau = \frac{t}{2N_e}$$

$$p(\text{no coal. by time } t) = e^{-\frac{\binom{n}{2}}{2N_e}t}$$

$$p(\text{no coal. by time } \tau) = e^{-\tau}$$

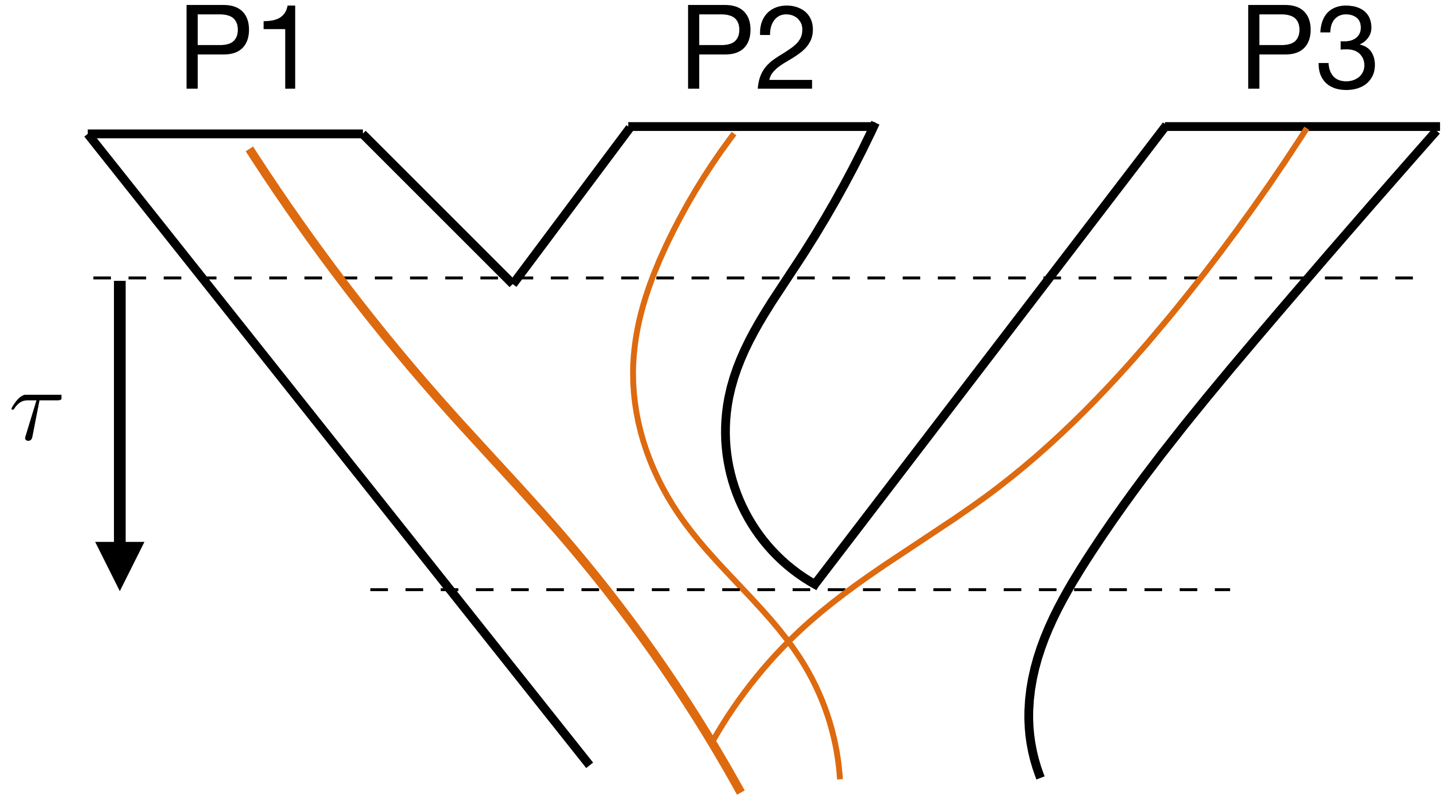
# Probability of **conflicting** topology



Prob(no coalescence AND P2 joins with P3)

$$= \frac{1}{3}e^{-\tau}$$

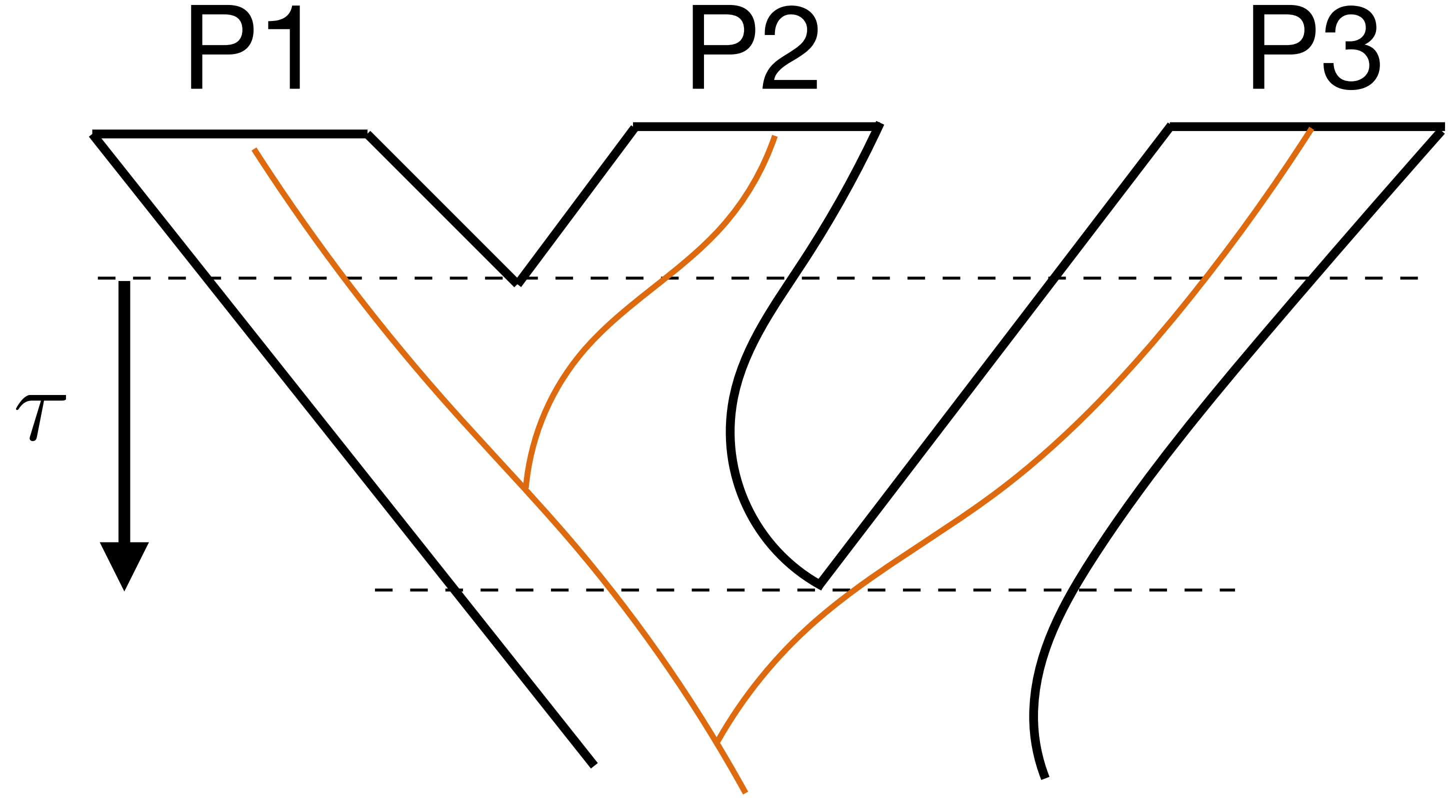
# Probability of **conflicting** topology



Prob(no coalescence AND P1 joins with P3)

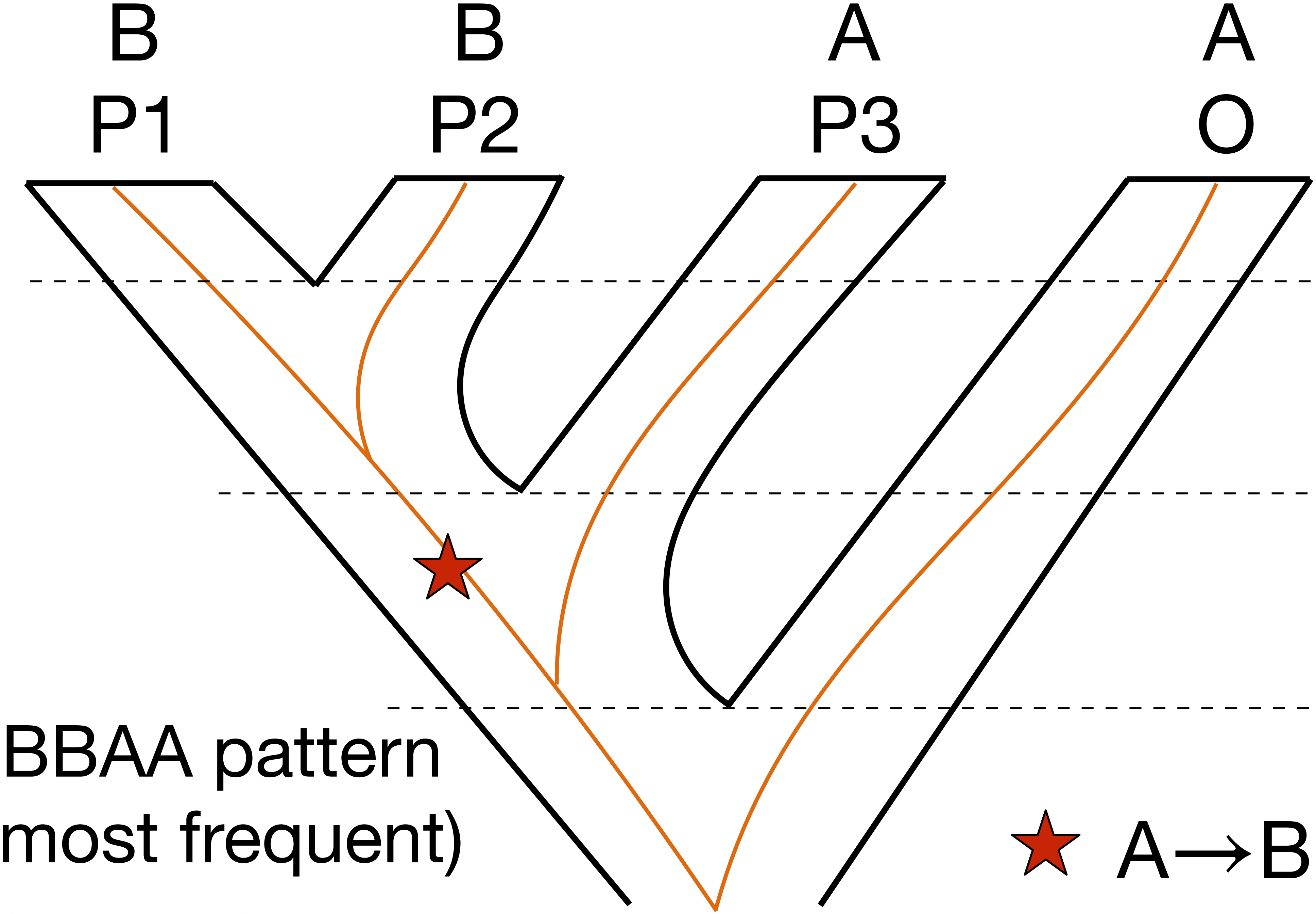
$$= \frac{1}{3} e^{-\tau}$$

# Probability of **compatible** topology

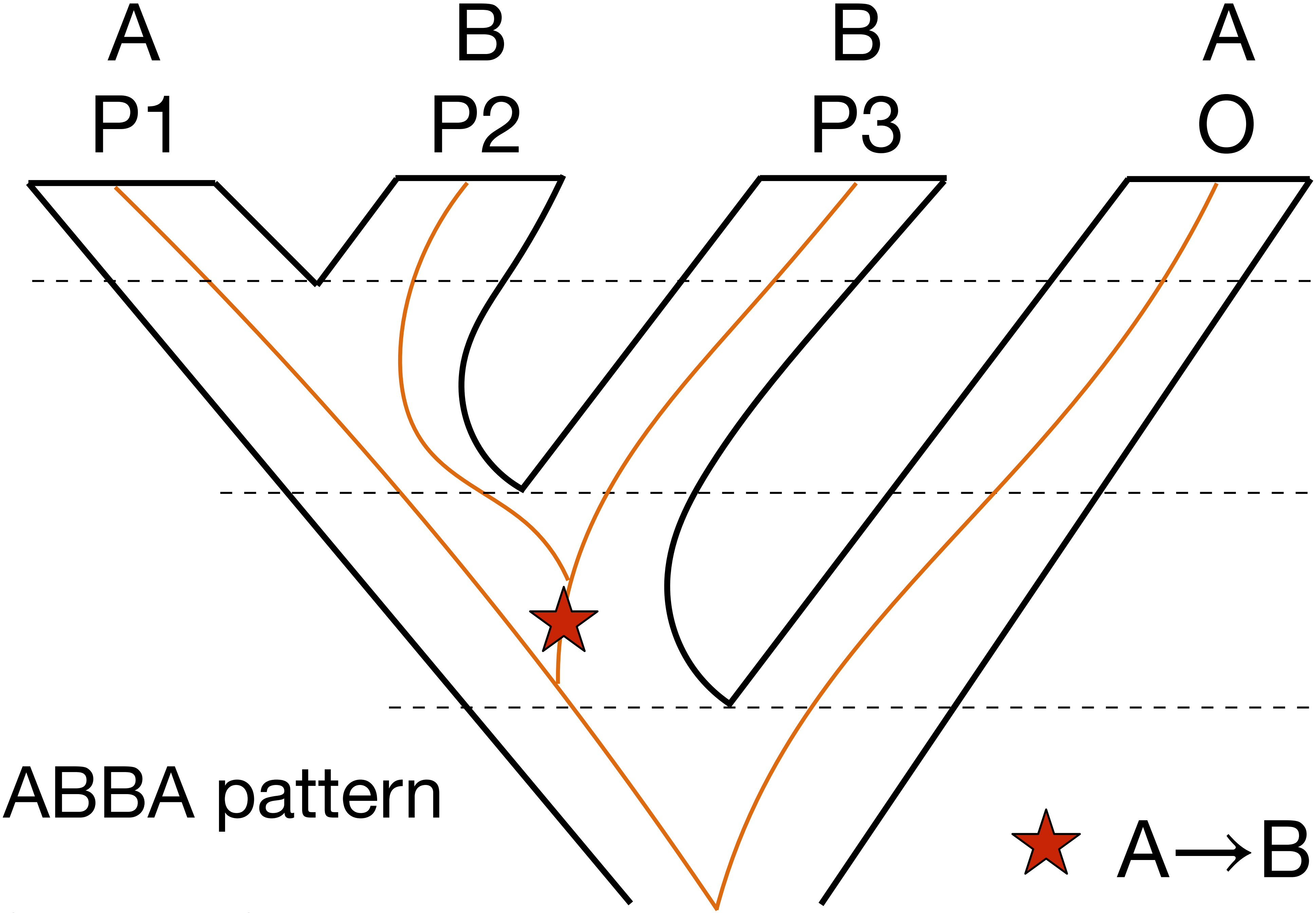


Prob(coalescence OR P1 joins with P2 anyway)

$$= \left(1 - e^{-\tau}\right) + \frac{1}{3}e^{-\tau} = 1 - \frac{2}{3}e^{-\tau}$$

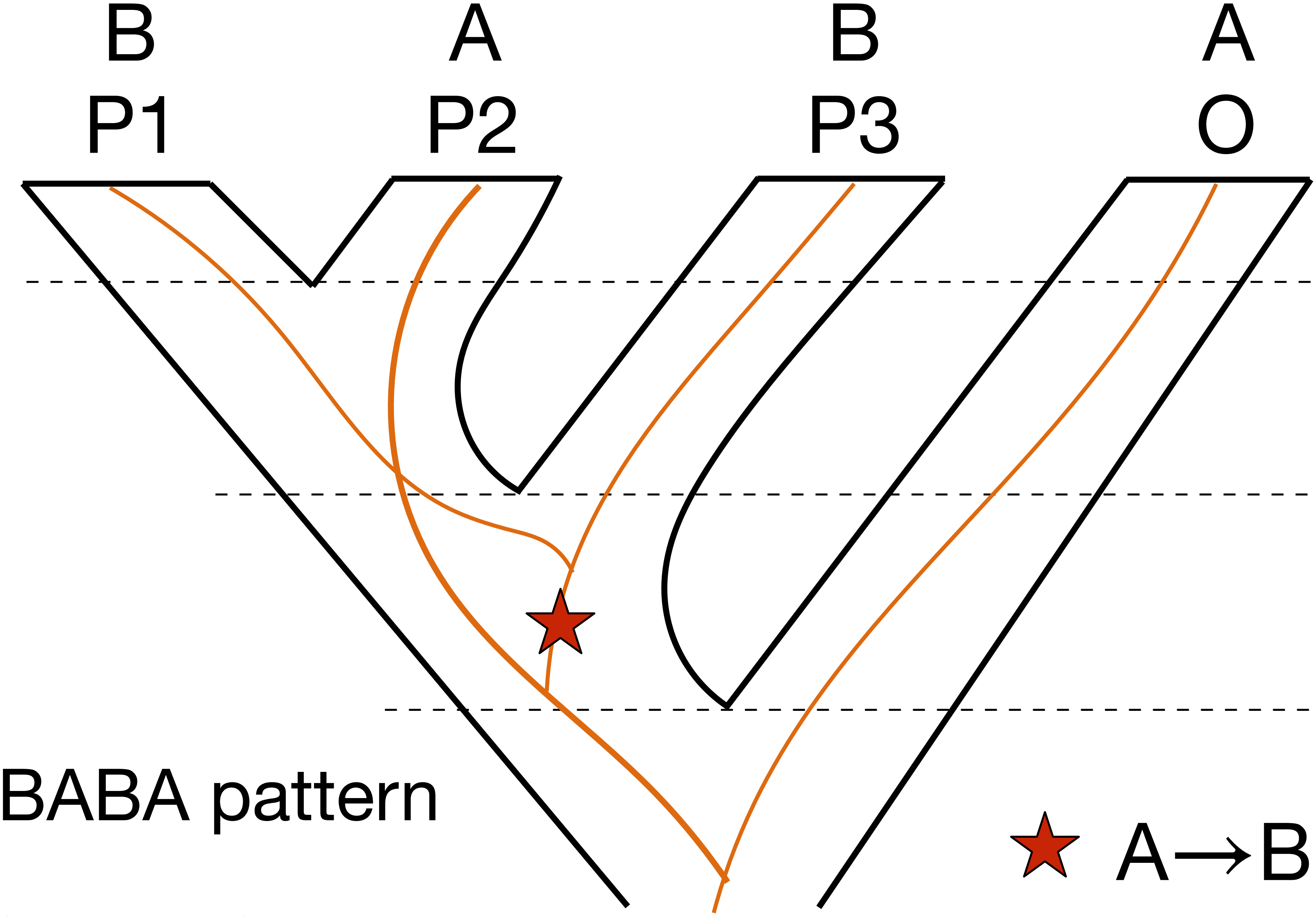


# BBAA pattern (most frequent)



ABBA pattern

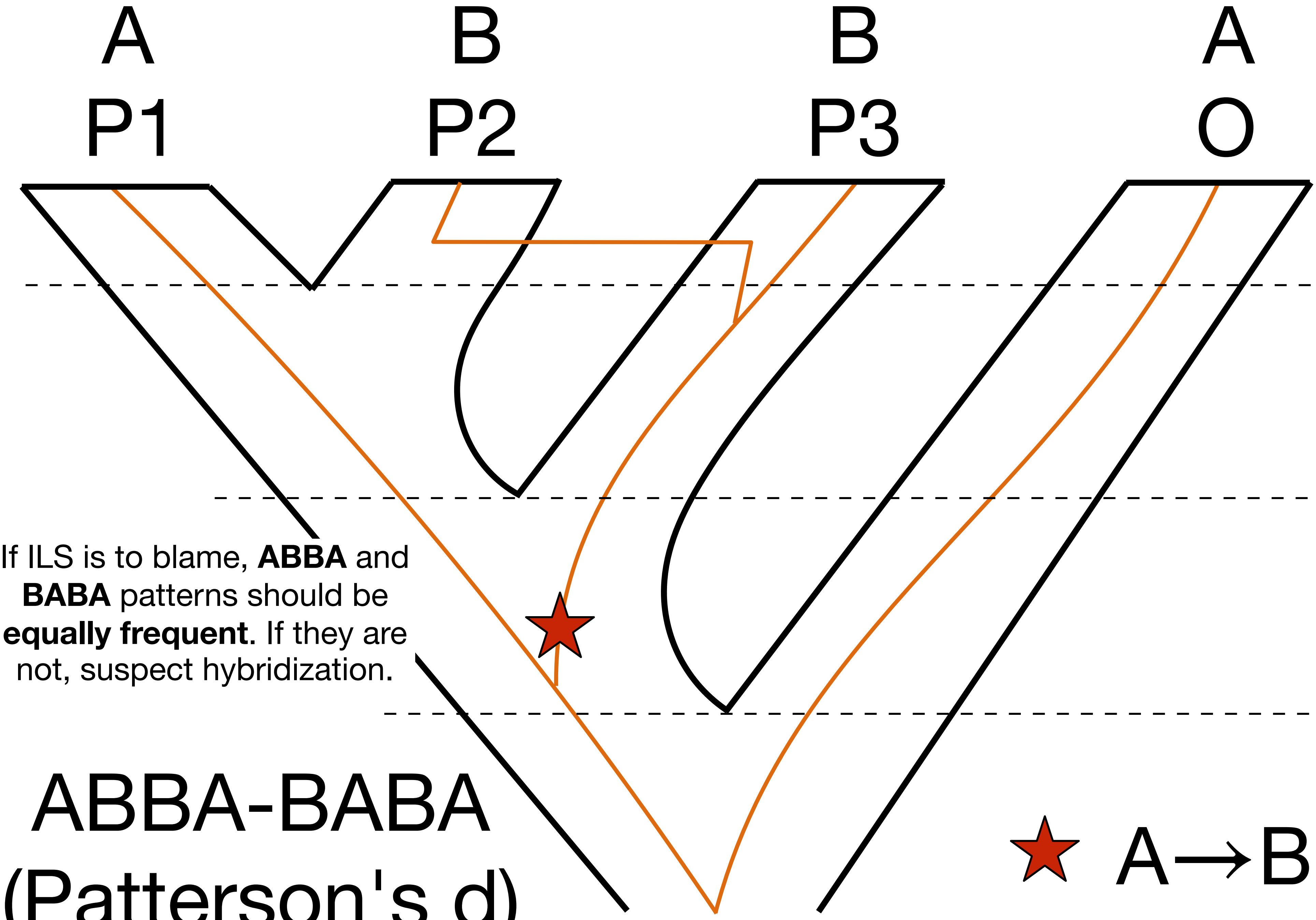
★ A→B



## BABA pattern

Paul O. Lewis ~ Phylogenetics, Spring 2026

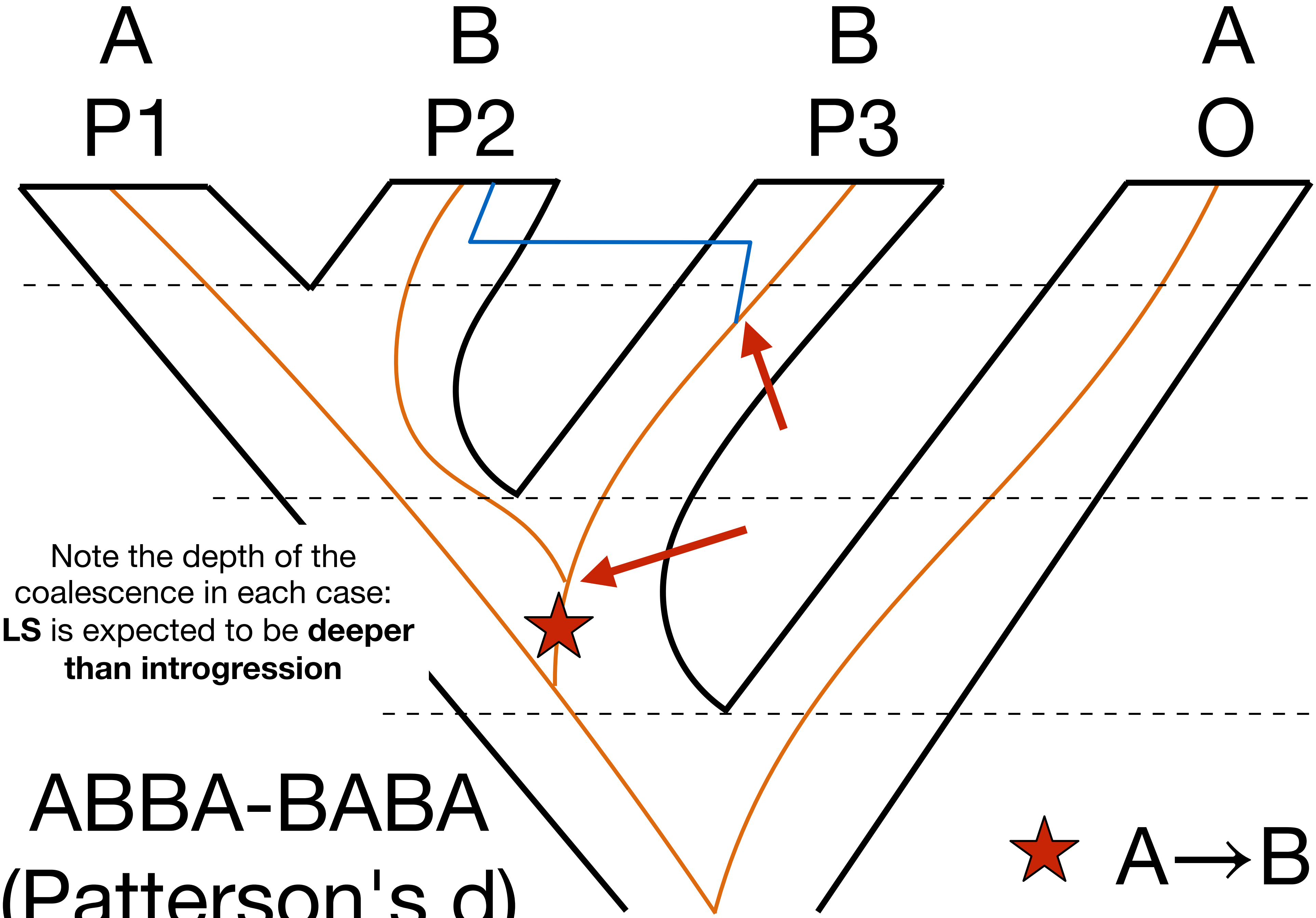
# ABBA-BABA (Patterson's d)



If ILS is to blame, **ABBA** and **BABA** patterns should be **equally frequent.** If they are not, suspect hybridization.

Durand et al. 2011; Patterson et al. 2012

# ABBA-BABA (Patterson's d)



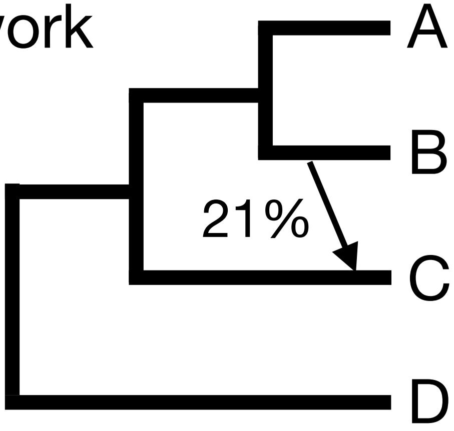
Note the depth of the coalescence in each case: **ILS** is expected to be **deeper than introgression**

★ A→B

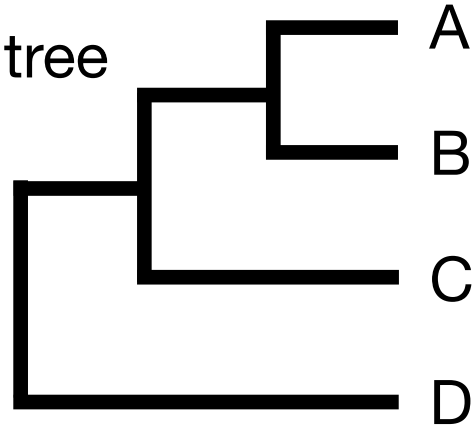
Durand et al. 2011; Patterson et al. 2012

# Major and minor trees

network

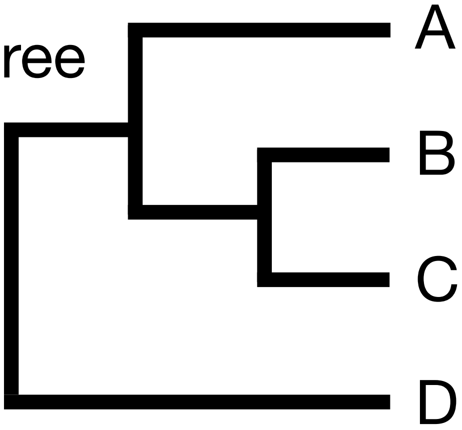


major tree



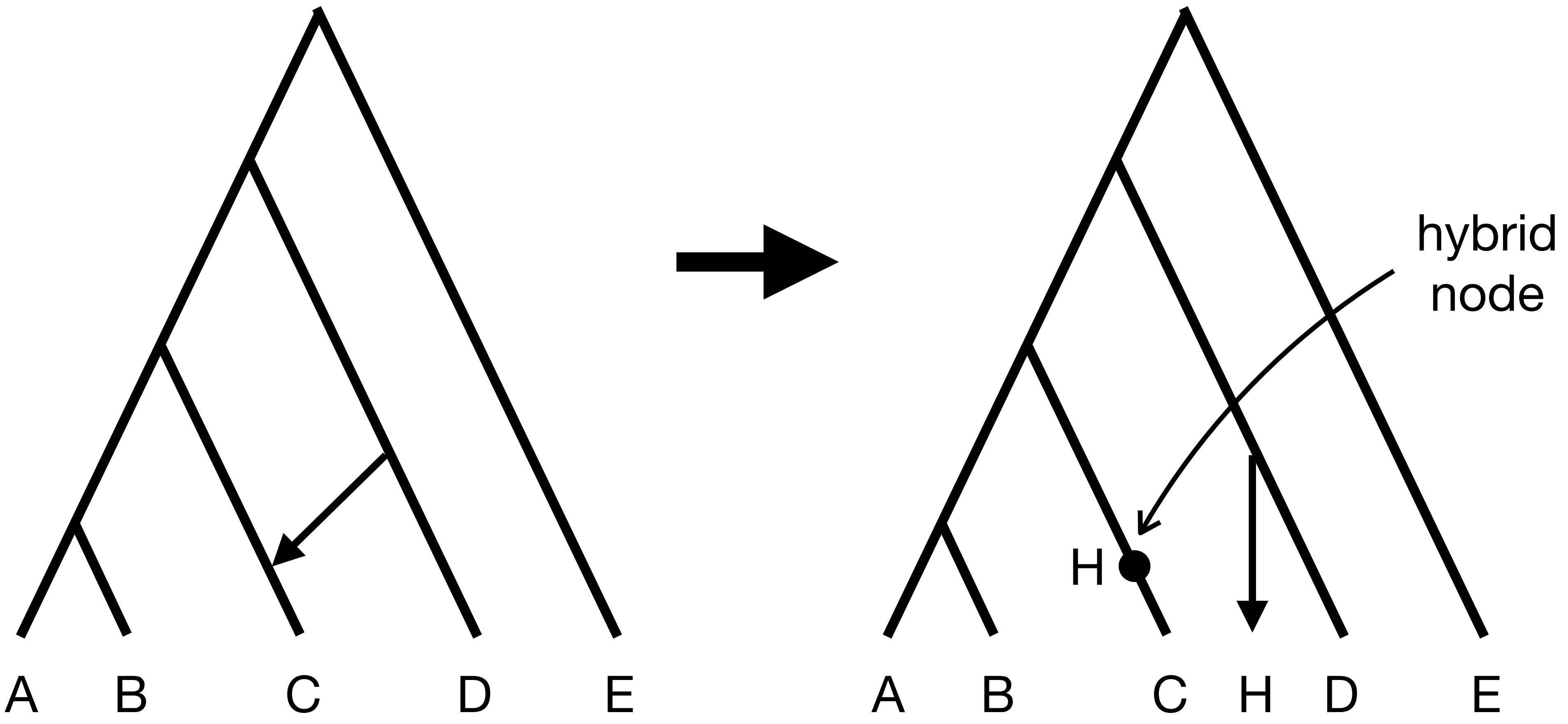
79% of loci

minor tree



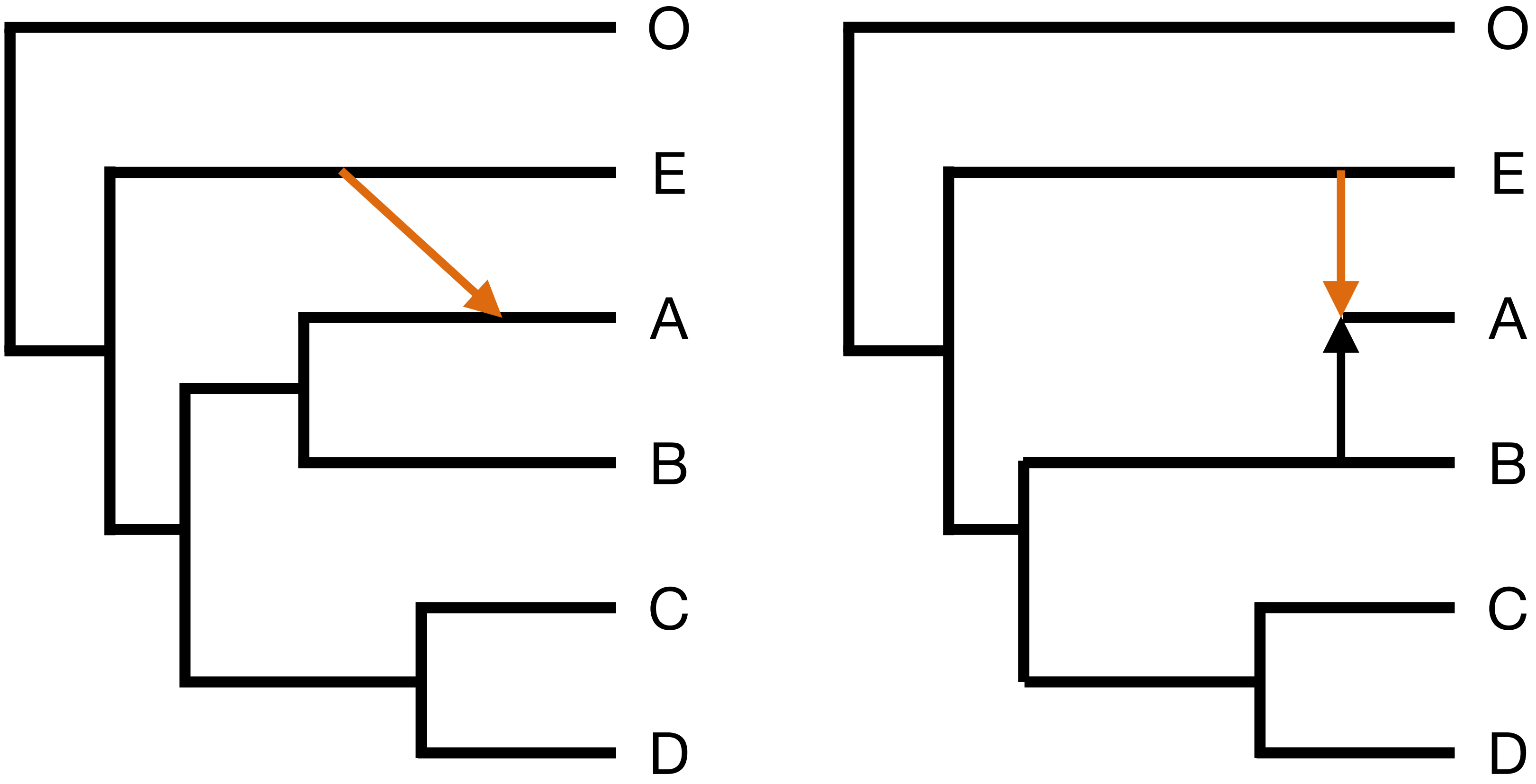
21% of loci

# Network newicks



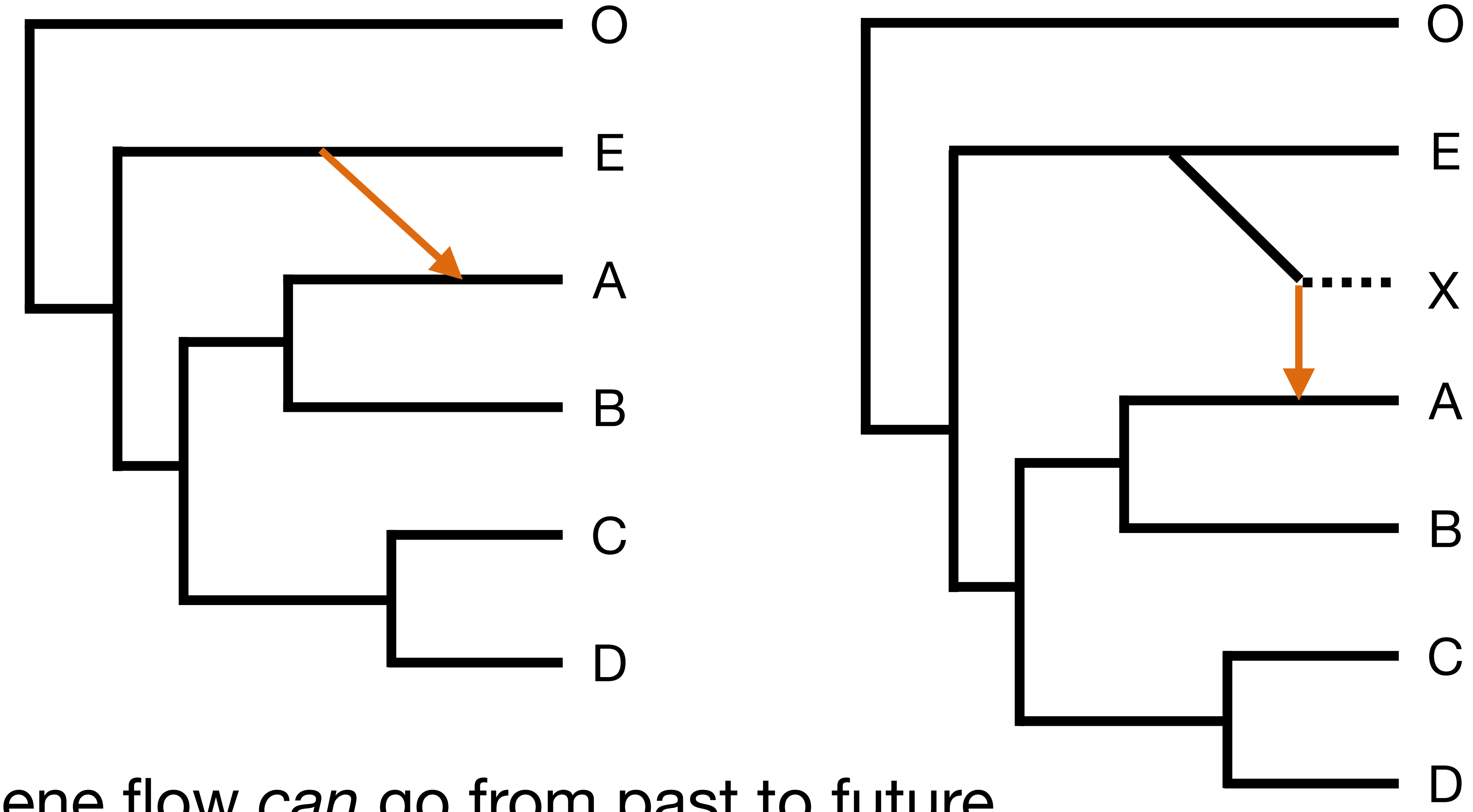
(((A,B),(C)#H),(D,#H)),E)

# Artistic license



Same network

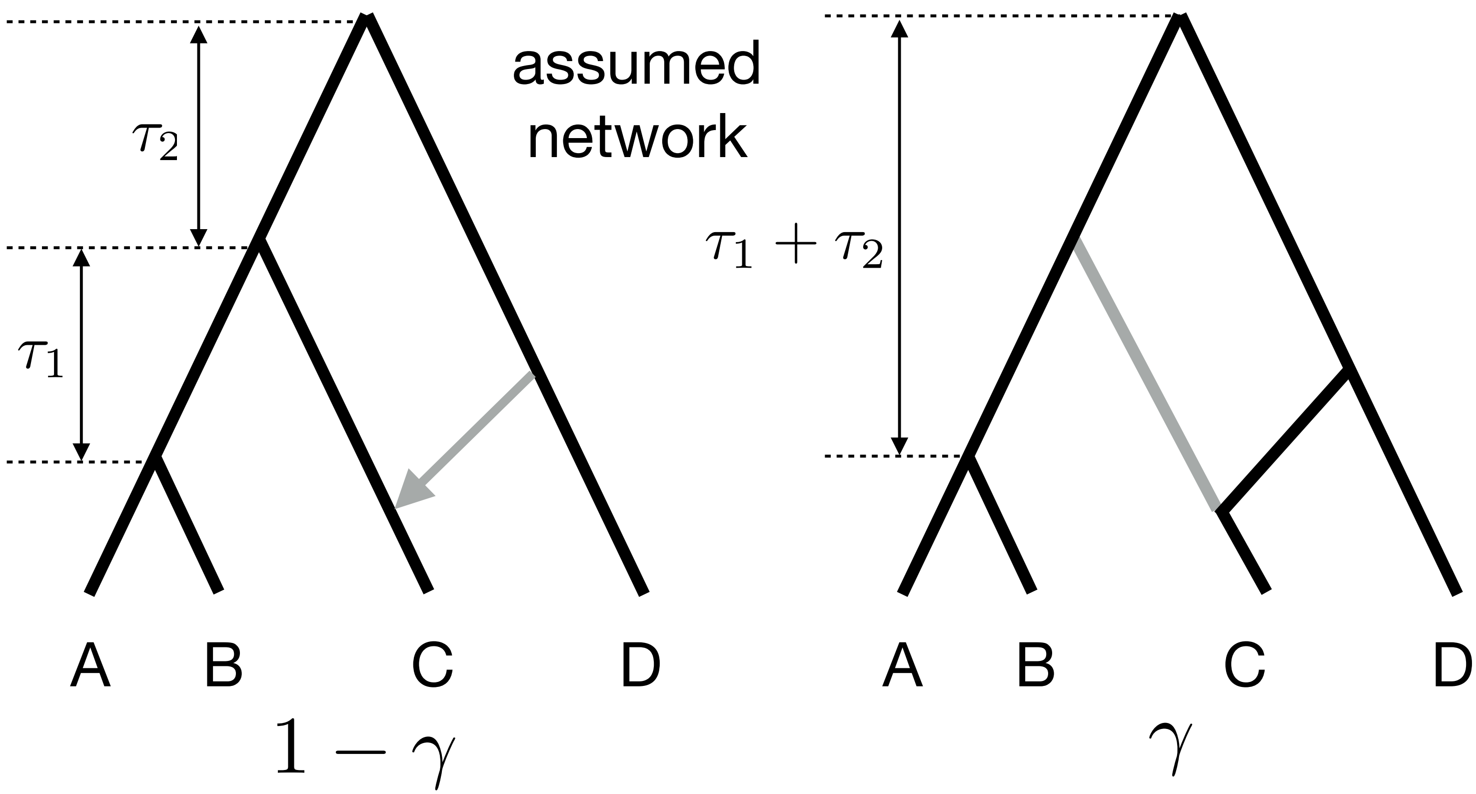
# Ghosts



Gene flow *can* go from past to future

# SNaQ

(Species Networks applying Quartets)



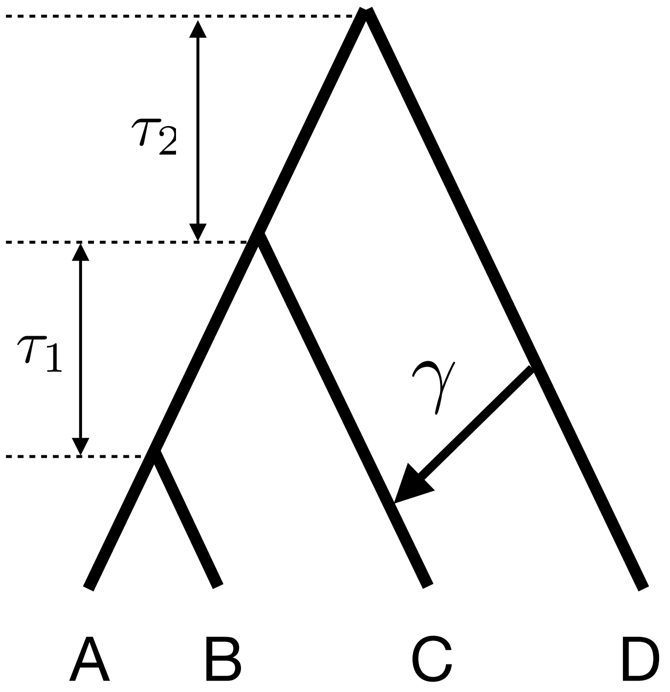
unrooted gene tree probabilities:

$$\Pr(\mathrm{AB}|\mathrm{CD}) = (1-\gamma)\left(1 - \tfrac{2}{3}e^{-\tau_1}\right) + \gamma\left(1 - \tfrac{2}{3}e^{-(\tau_1+\tau_2)}\right)$$

$$\Pr(\mathrm{AC}|\mathrm{BD}) = (1-\gamma)\left(\tfrac{1}{3}e^{-\tau_1}\right) + \gamma\left(\tfrac{1}{3}e^{-(\tau_1+\tau_2)}\right)$$

$$\Pr(\mathrm{AD}|\mathrm{BC}) = (1-\gamma)\left(\tfrac{1}{3}e^{-\tau_1}\right) + \gamma\left(\tfrac{1}{3}e^{-(\tau_1+\tau_2)}\right)$$

Solís-Lemus and Ané (2016)



# SNaQ

Likelihood of this species network is the multinomial probability of observing:

- $n_1$ gene trees with split AB|CD
- $n_2$ gene trees with split AC|BD
- $n_3$ gene trees with split AD|BC

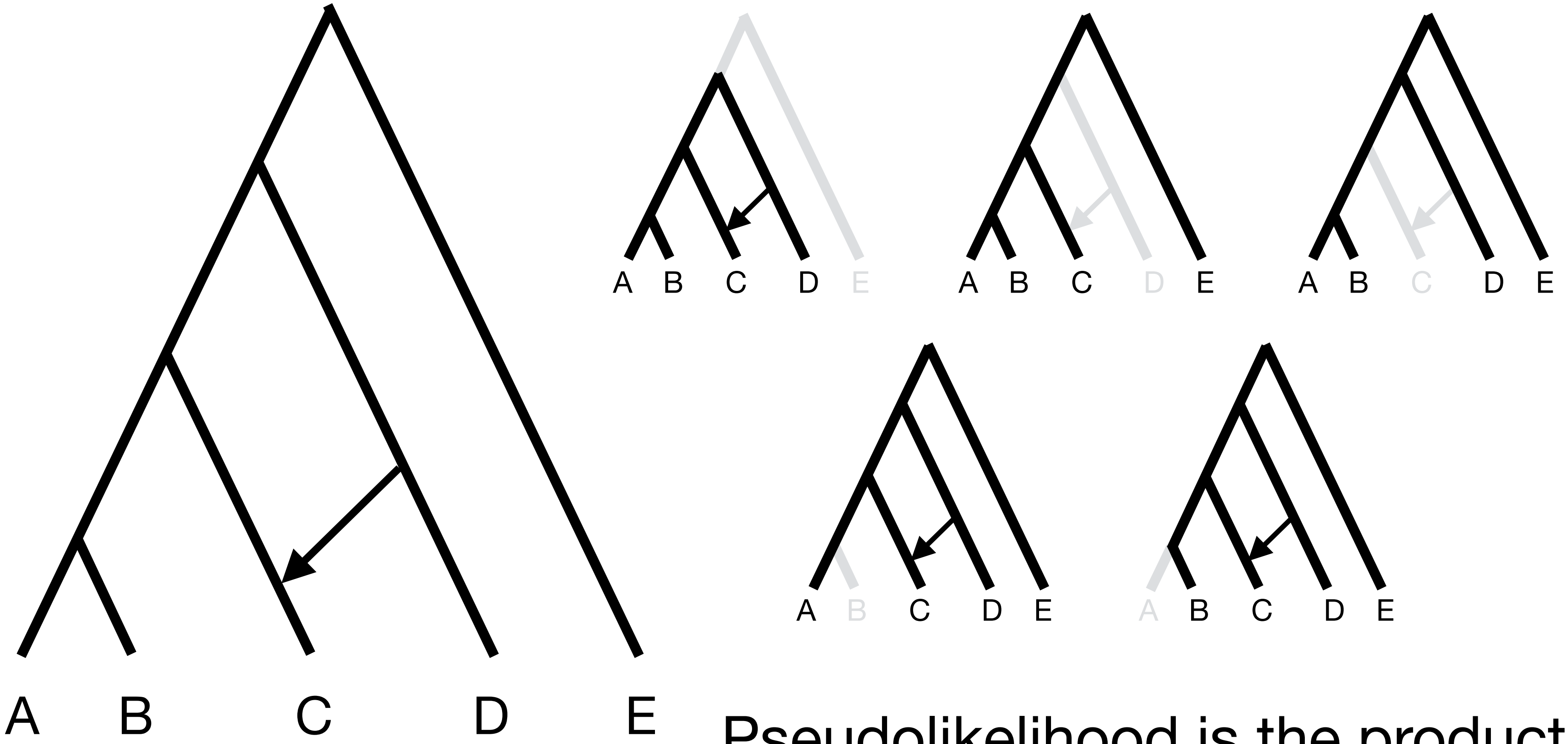
$$L = \Pr(\text{AB}|\text{CD})^{n_1}\ \Pr(\text{AC}|\text{BD})^{n_2}\ \Pr(\text{AD}|\text{BC})^{n_3}$$

The probabilities are conditional on the network topology and particular values for $\tau_1$,$\tau_2$, and $\gamma$, but my notation omits this.

SNaQ searches through networks, $\tau$ values, and $\gamma$ values to find the maximum likelihood species network.

# SNaQ pseudolikelihoods

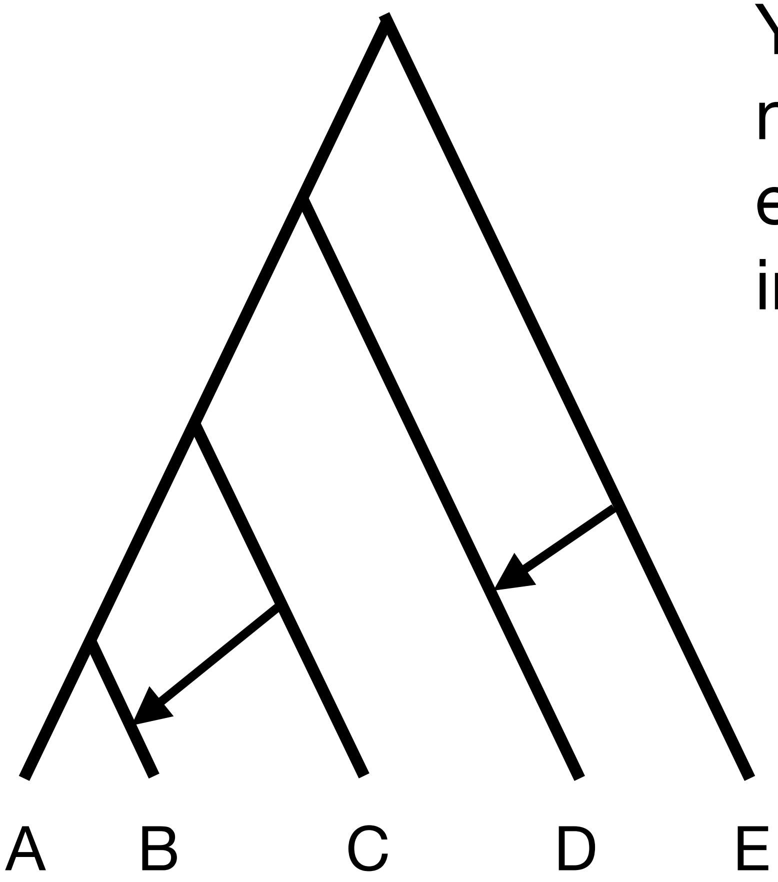
What if there are more than 4 taxa in the network?



Pseudolikelihood is the product of quartet likelihoods.

# SNaQ pseudolikelihoods

What if there is more than one introgression event?

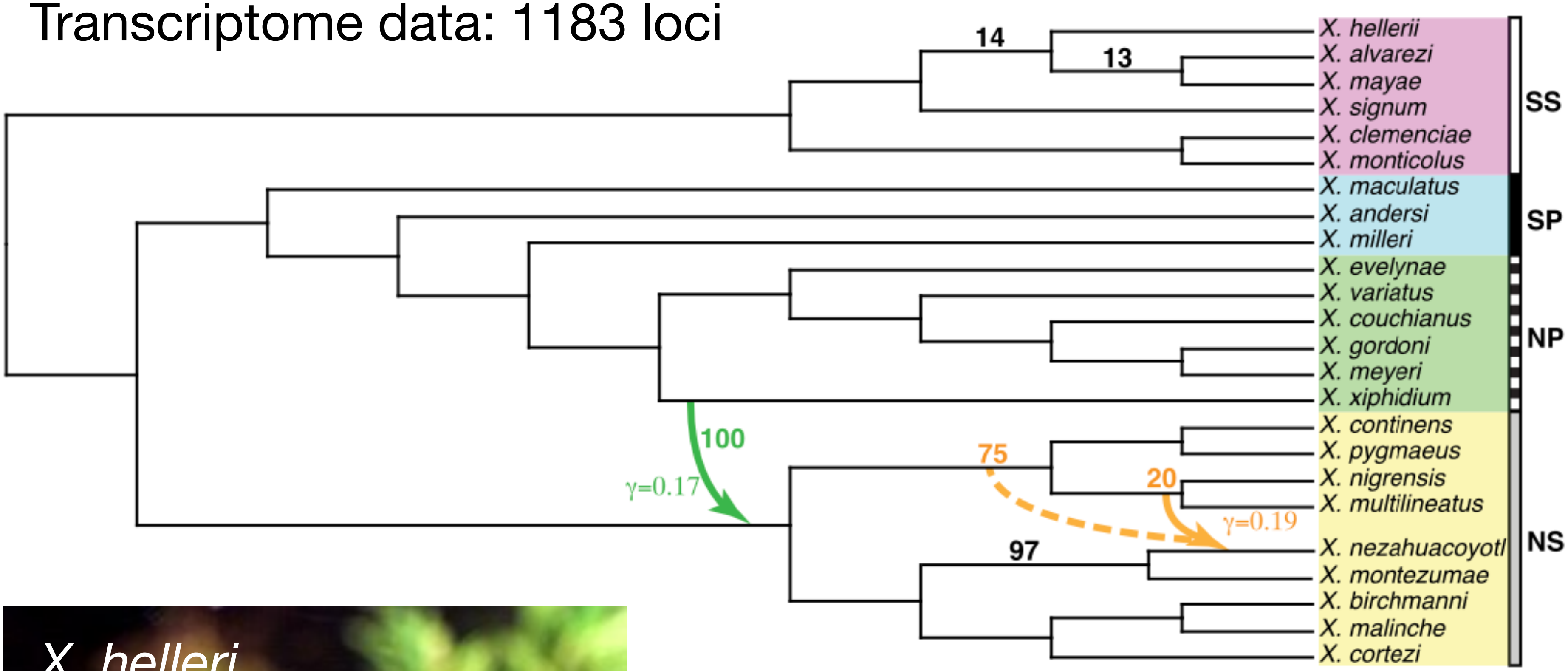


You specify *k*, the maximum number of events, and it will estimate networks with 0 to *k* introgression events.

Can use AIC/BIC to decide which value of *k* provides the best fit.

# *Xiphophorus* fishes

Transcriptome data: 1183 loci



*https://en.wikipedia.org/wiki/Xiphophorus*

SS (southern swordtails); NS (northern swordtails);
SP (southern platyfishes); NP (northern platyfishes)

Solís-Lemus and Ané (2016)

# Network Software

| | |
|---|---|
| SNaQ | https://github.com/JuliaPhylo/PhyloNetworks.jl<br>https://github.com/solislemuslab/snaq-tutorial |
| BEAST2 | https://www.beast2.org |
| PhyloNet | https://phylogenomics.rice.edu |

NANUQ, RF-Net, PhyNEST, HyDe, MSCQuartets,...

# Implicit networks

SplitsTree looks like a network, but is really just a way to visualize conflicting splits