

Phylogenetics

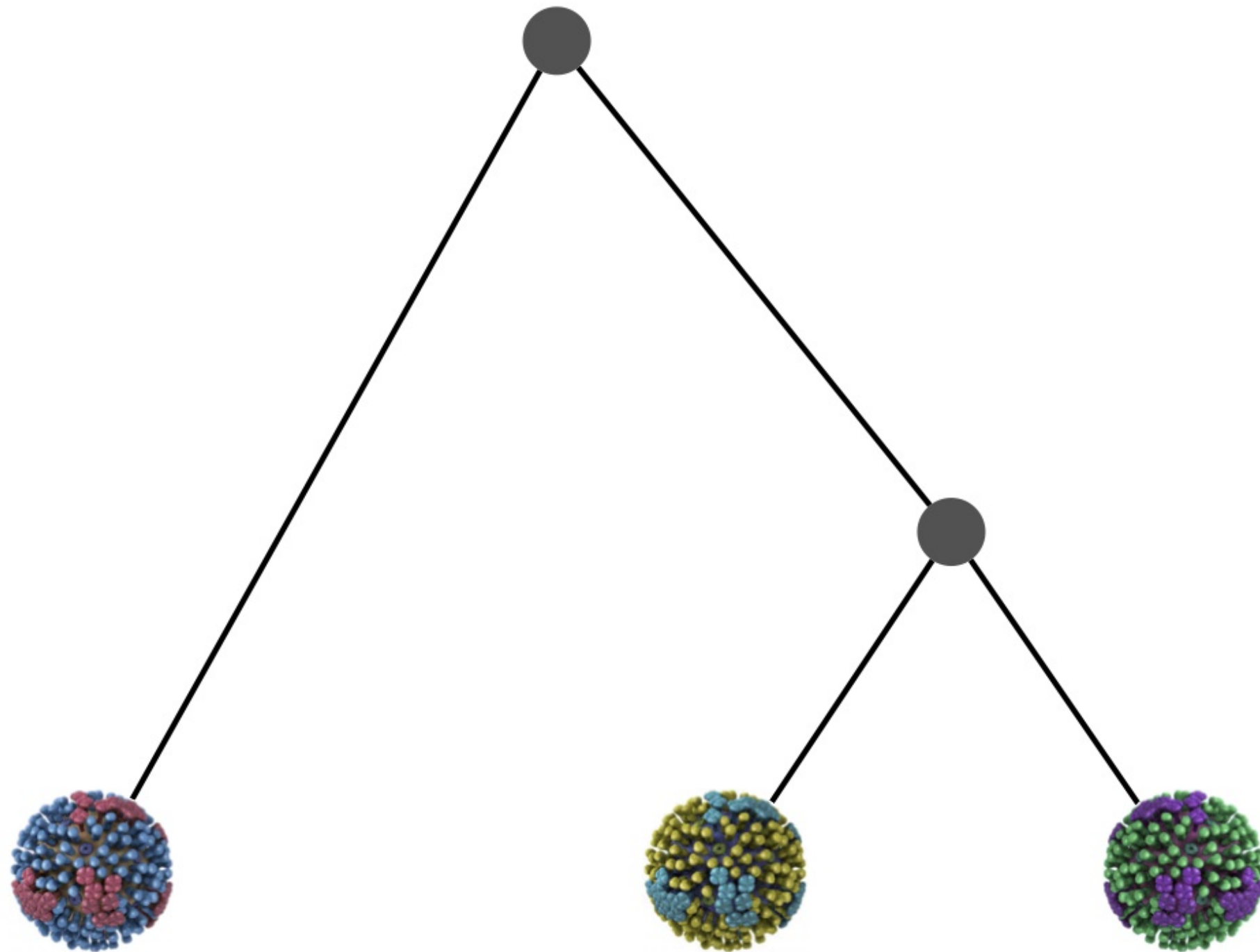
I think



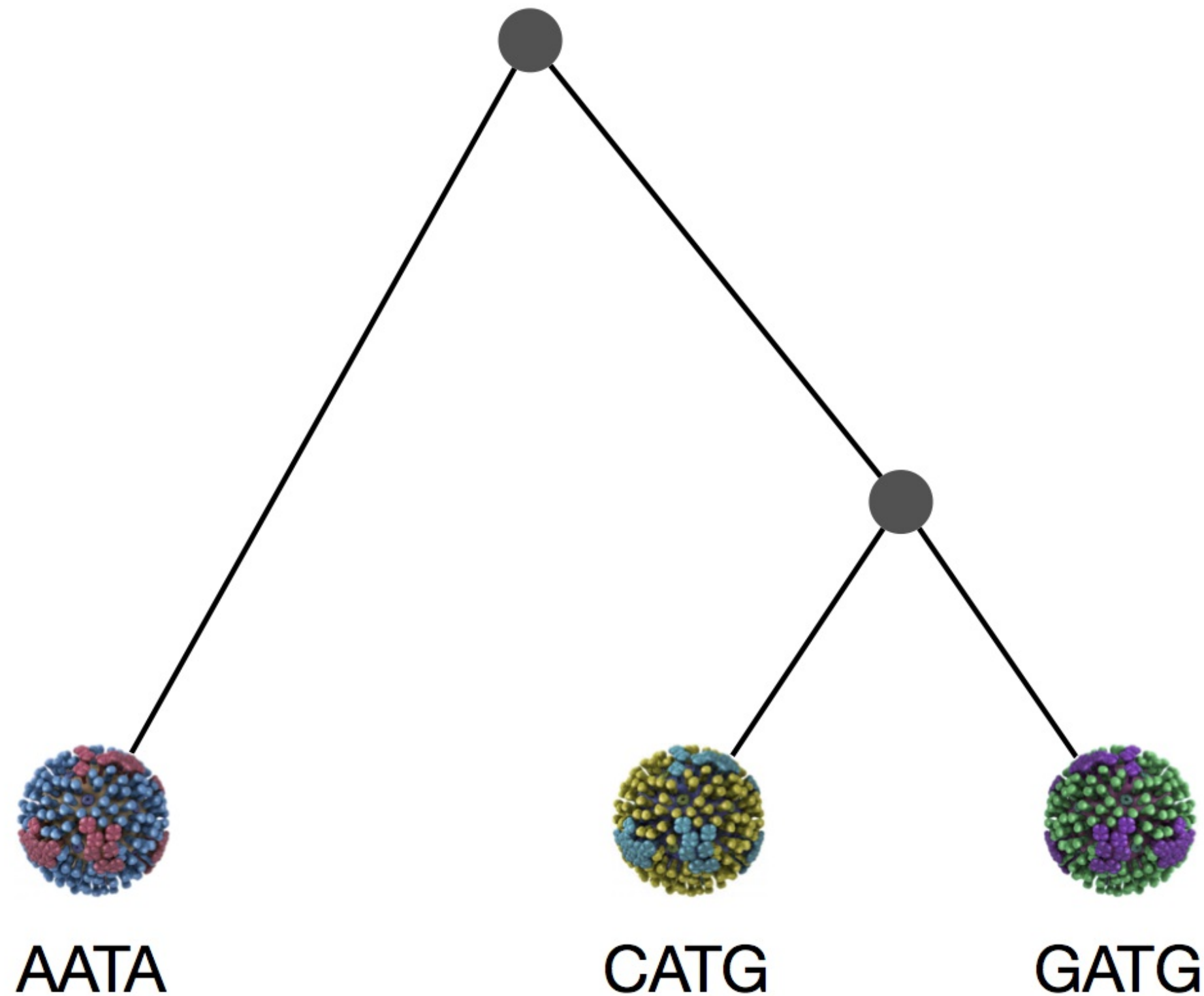
can never be that one
 can never be that one
 living in
 in
 Do do this & have many species
 in the same (as is) - repeated
 in the same.



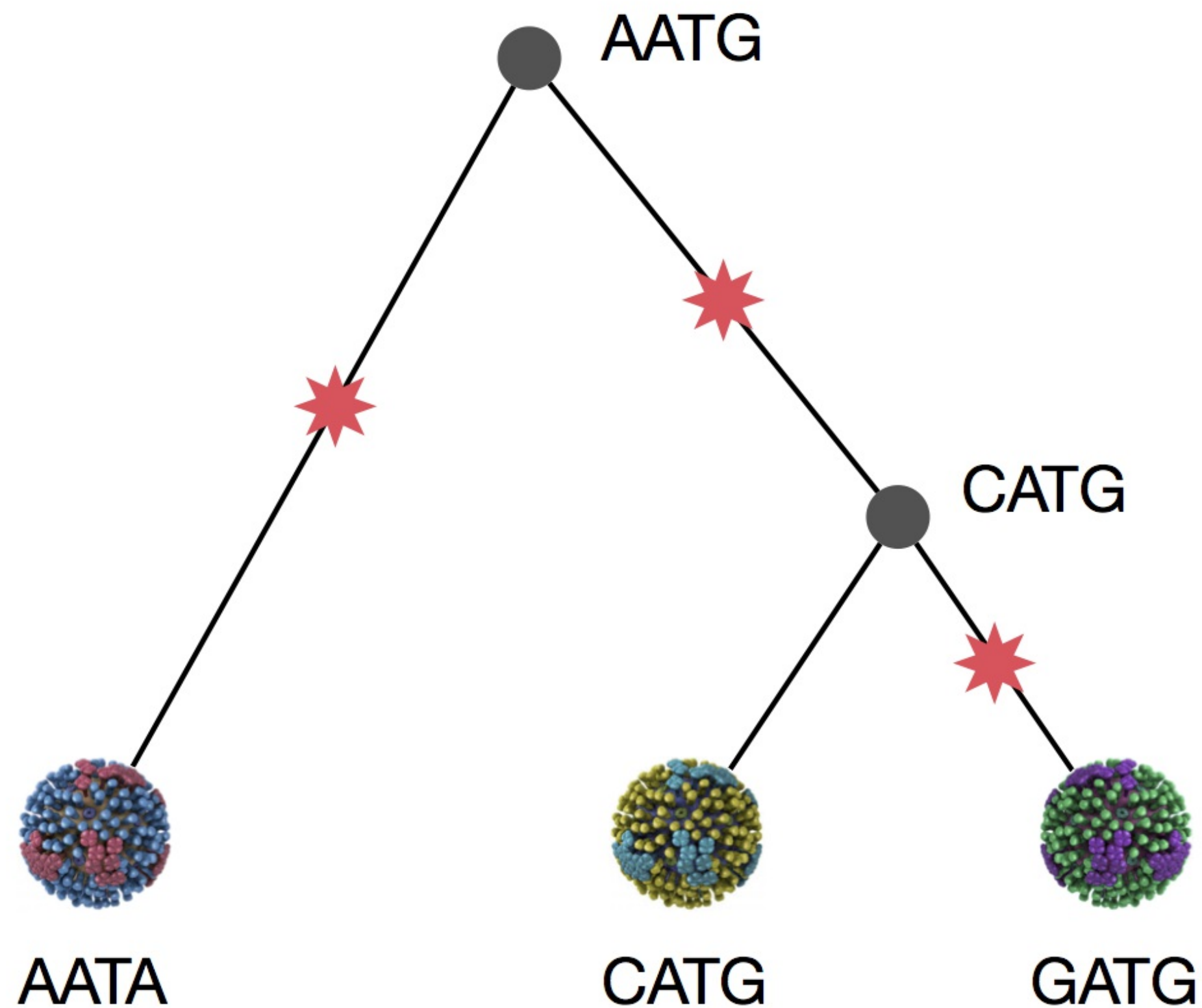
Phylogeny describes evolutionary relationships



Phylogeny is usually a hypothesis based on characteristics of sampled taxa



Phylogeny implies a series of mutational events leading to observed tip states

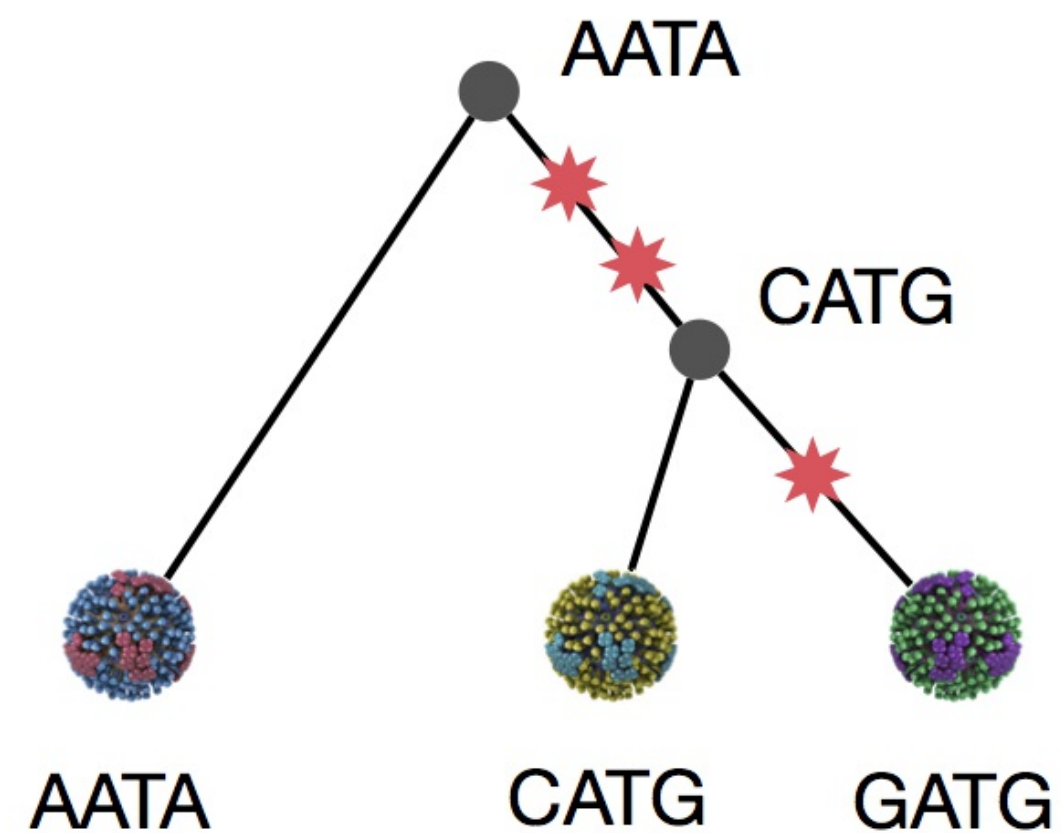
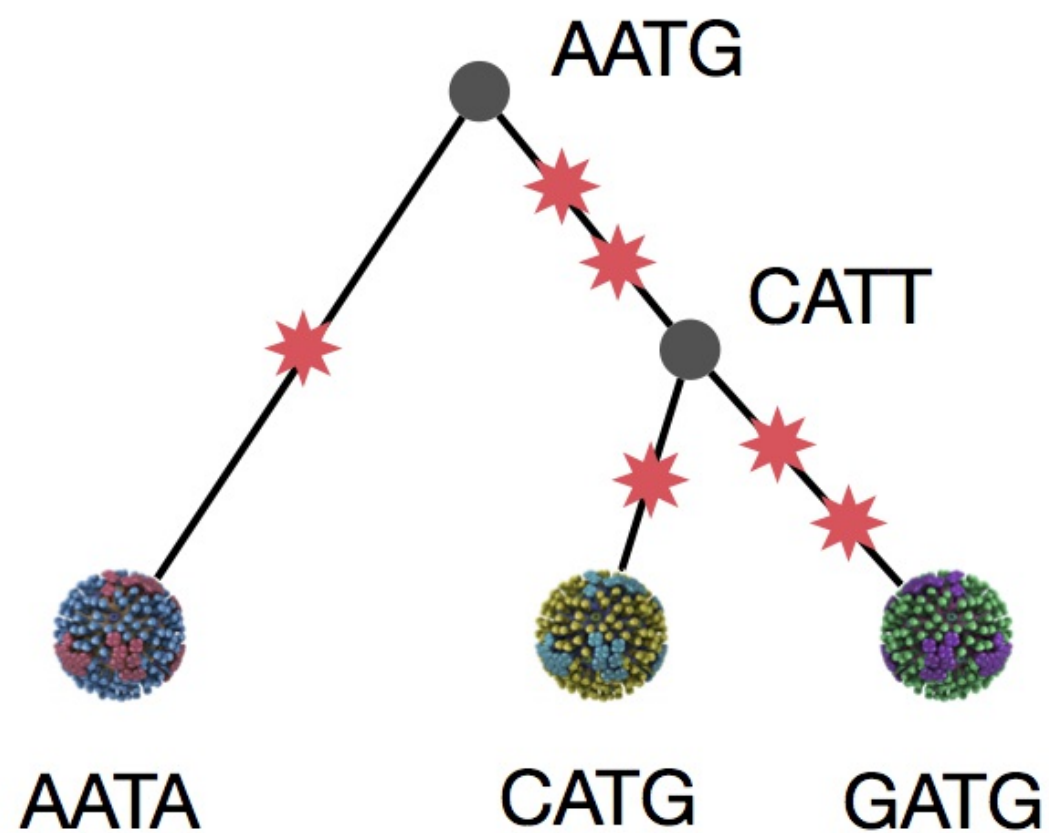
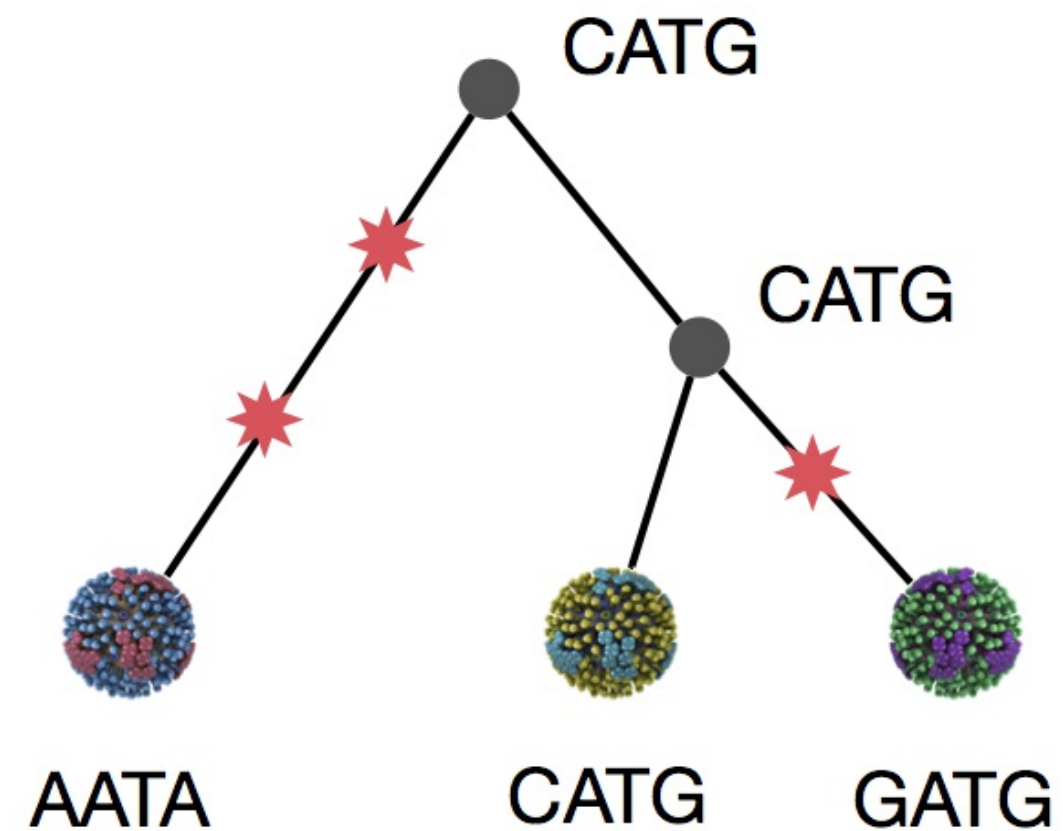
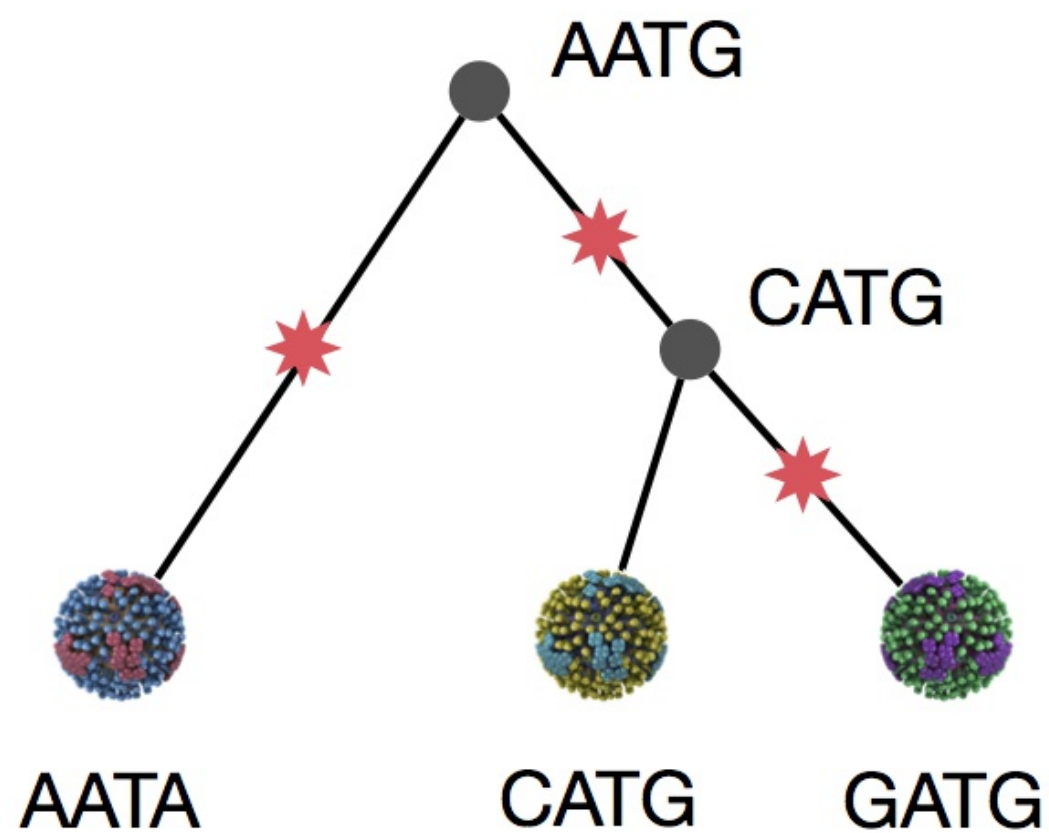




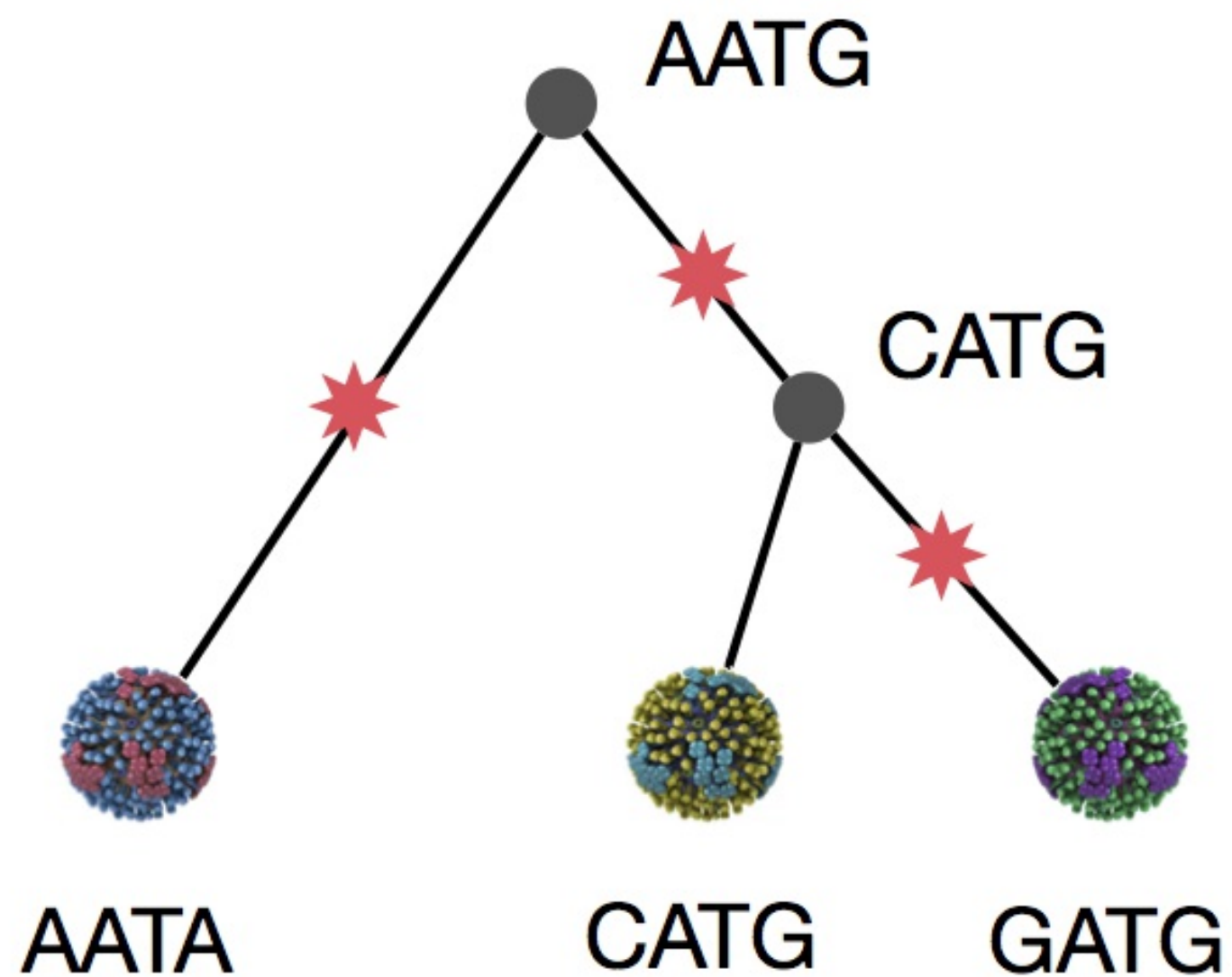
Parsimony is based on Occam's razor

Among competing hypotheses that predict equally well, the one with the fewest assumptions should be selected.

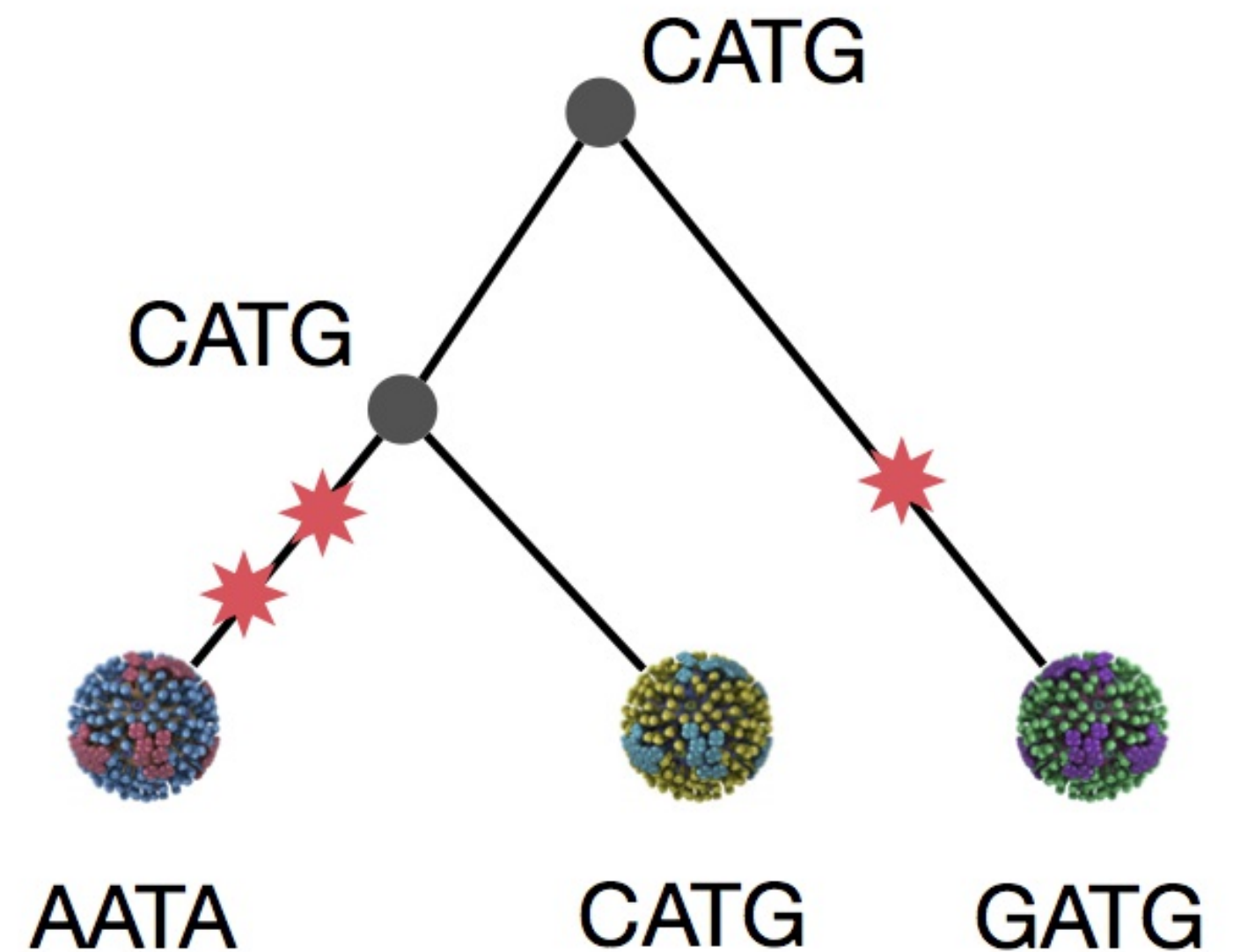
Parsimony suggests this topology requires
3 mutations at minimum



Parsimony suggests both topologies
equally tenable



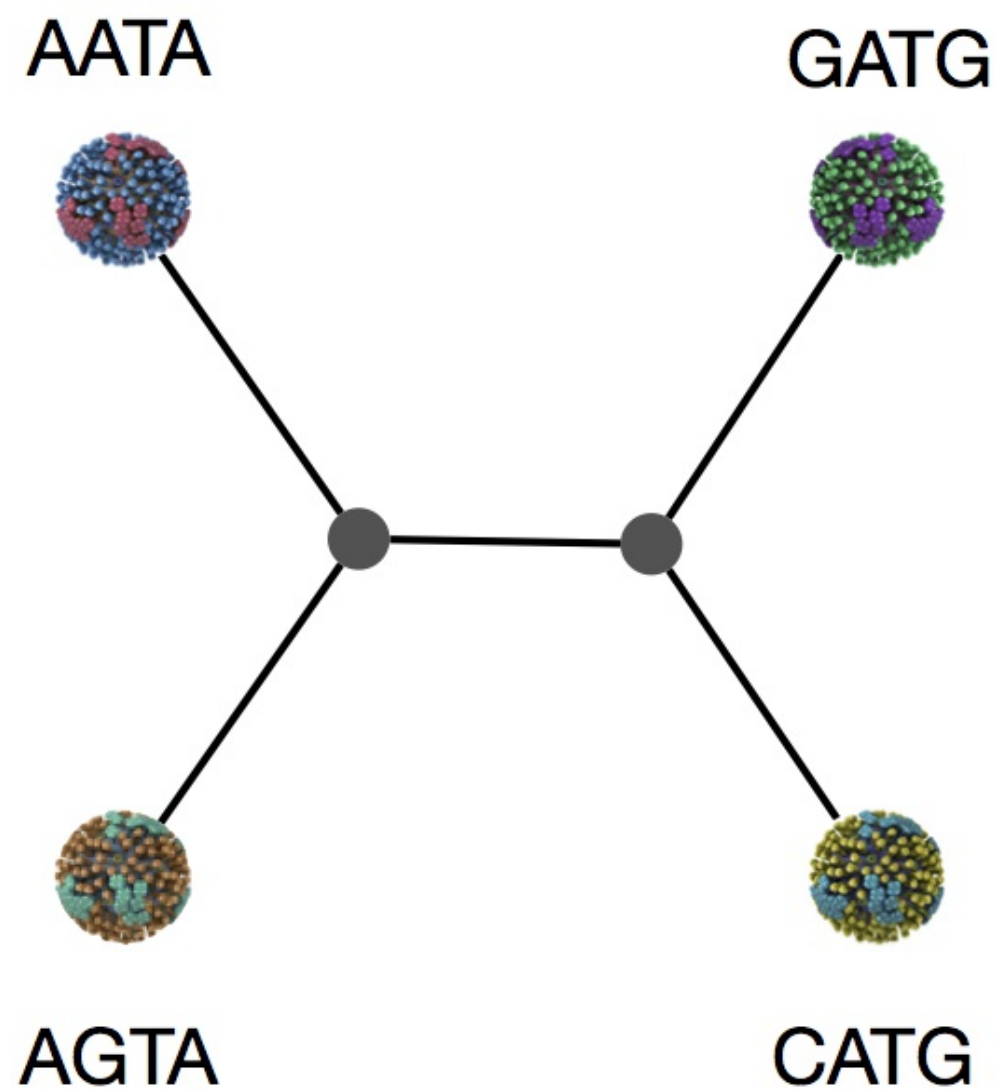
3 mutations minimum



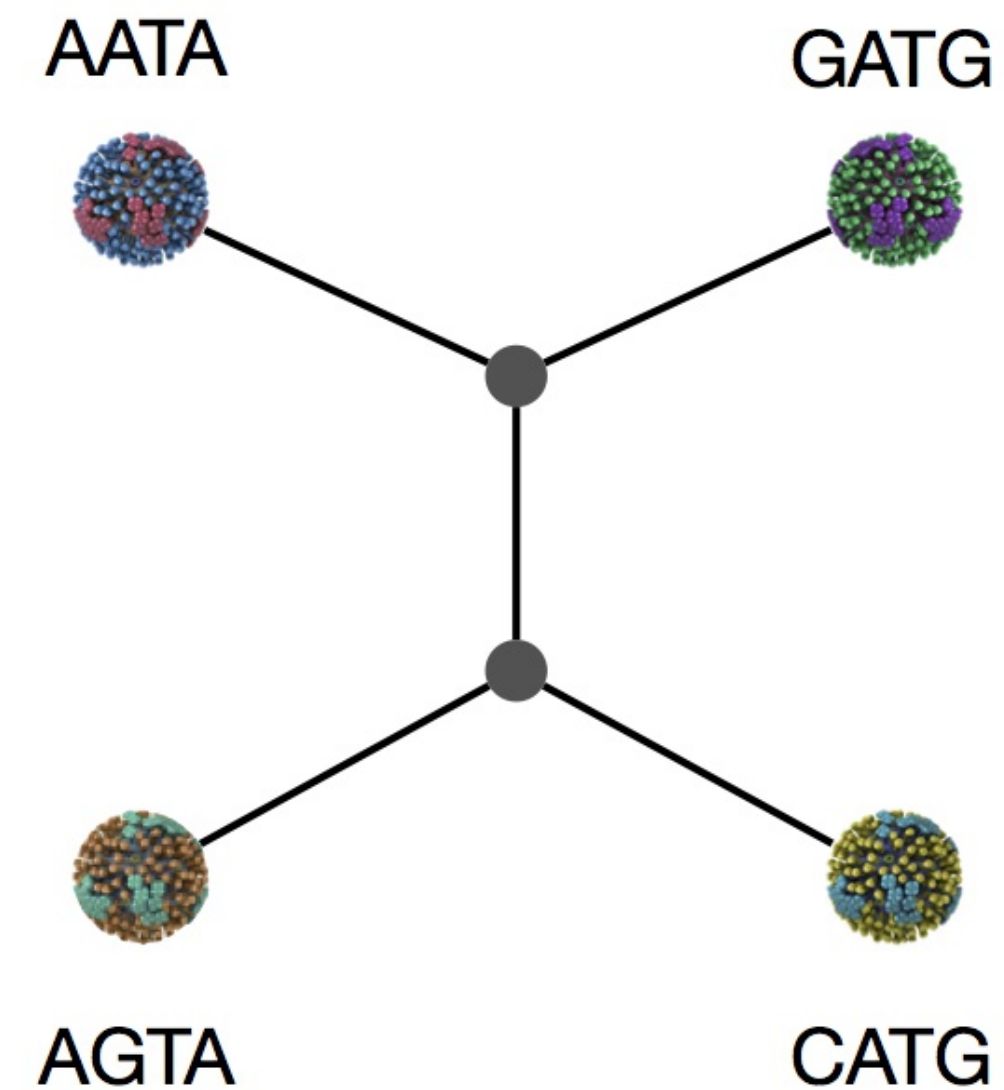
3 mutations minimum

Exercise: which topology is more likely under parsimony?

Tree #1

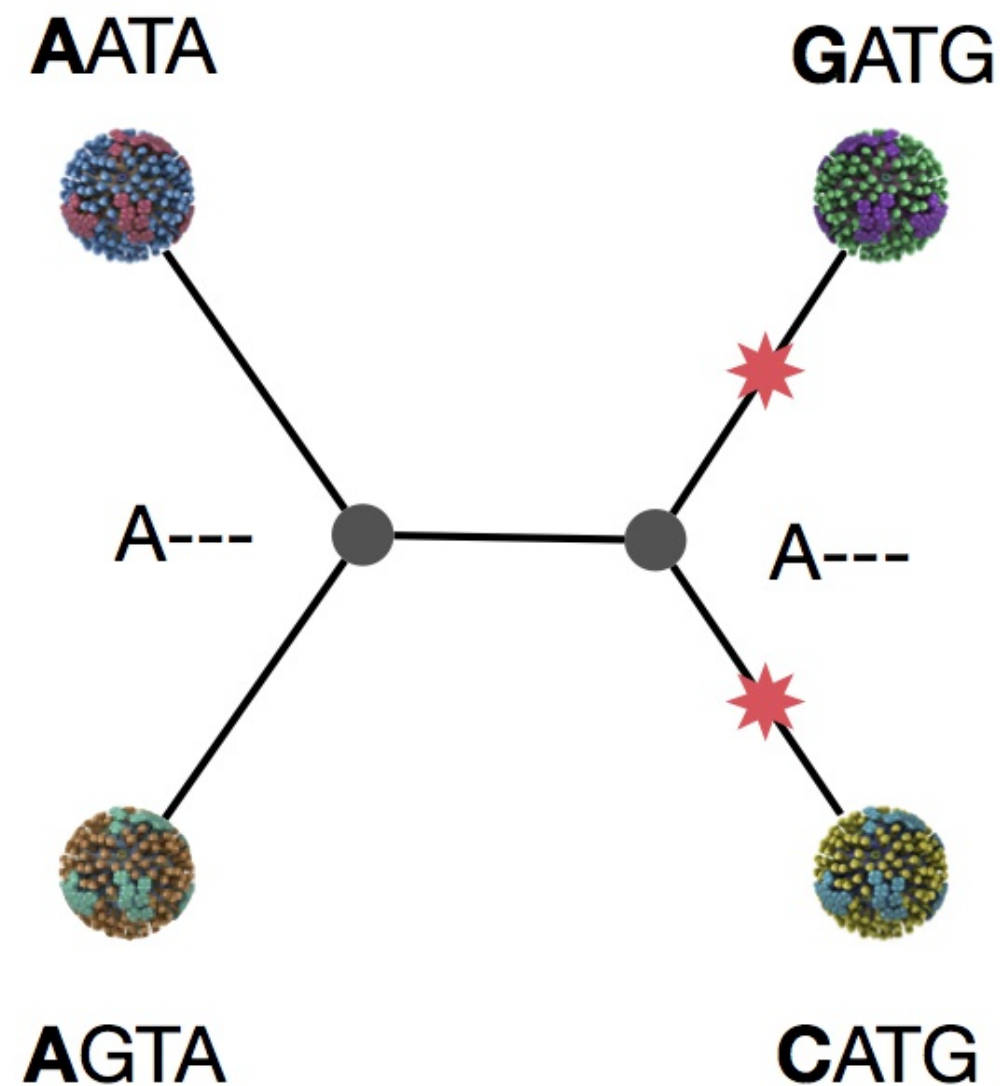


Tree #2



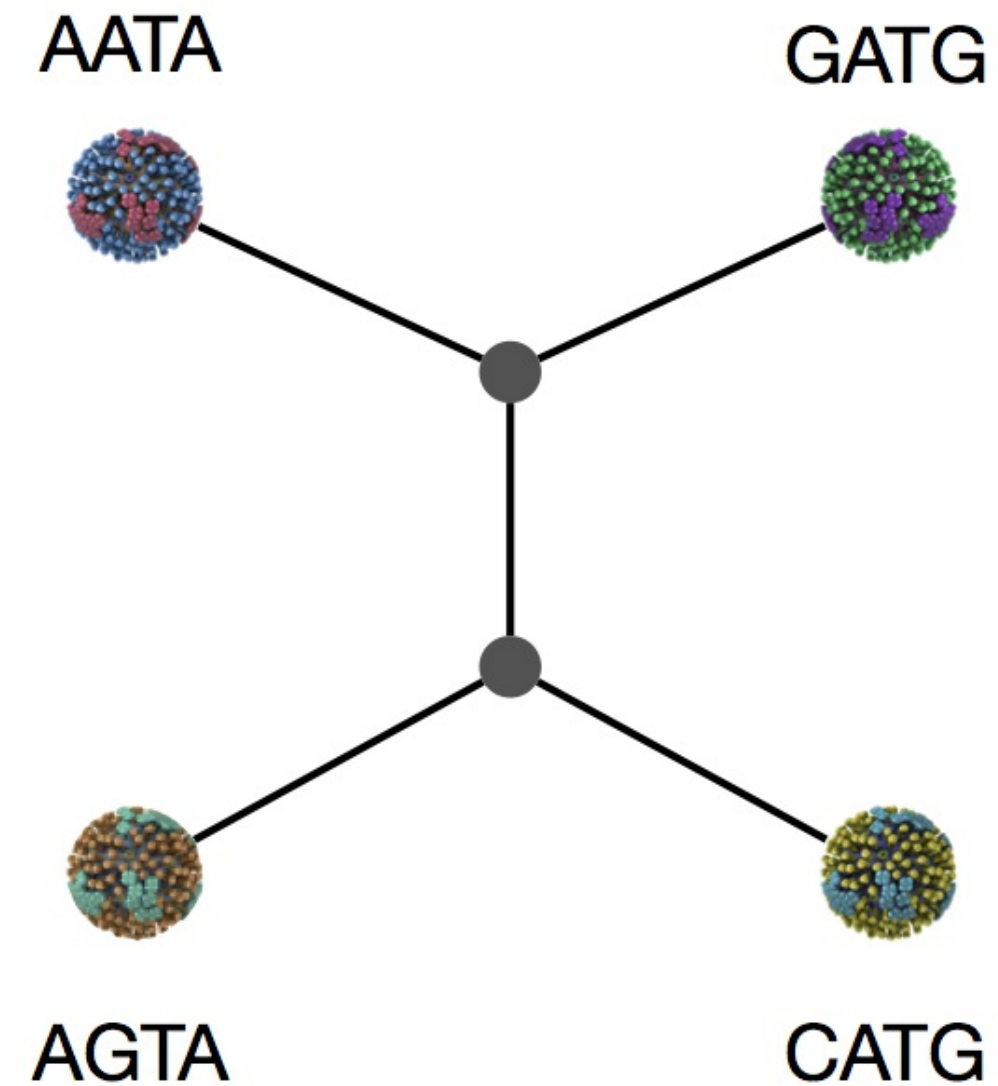
Exercise: which topology is more likely under parsimony?

Tree #1



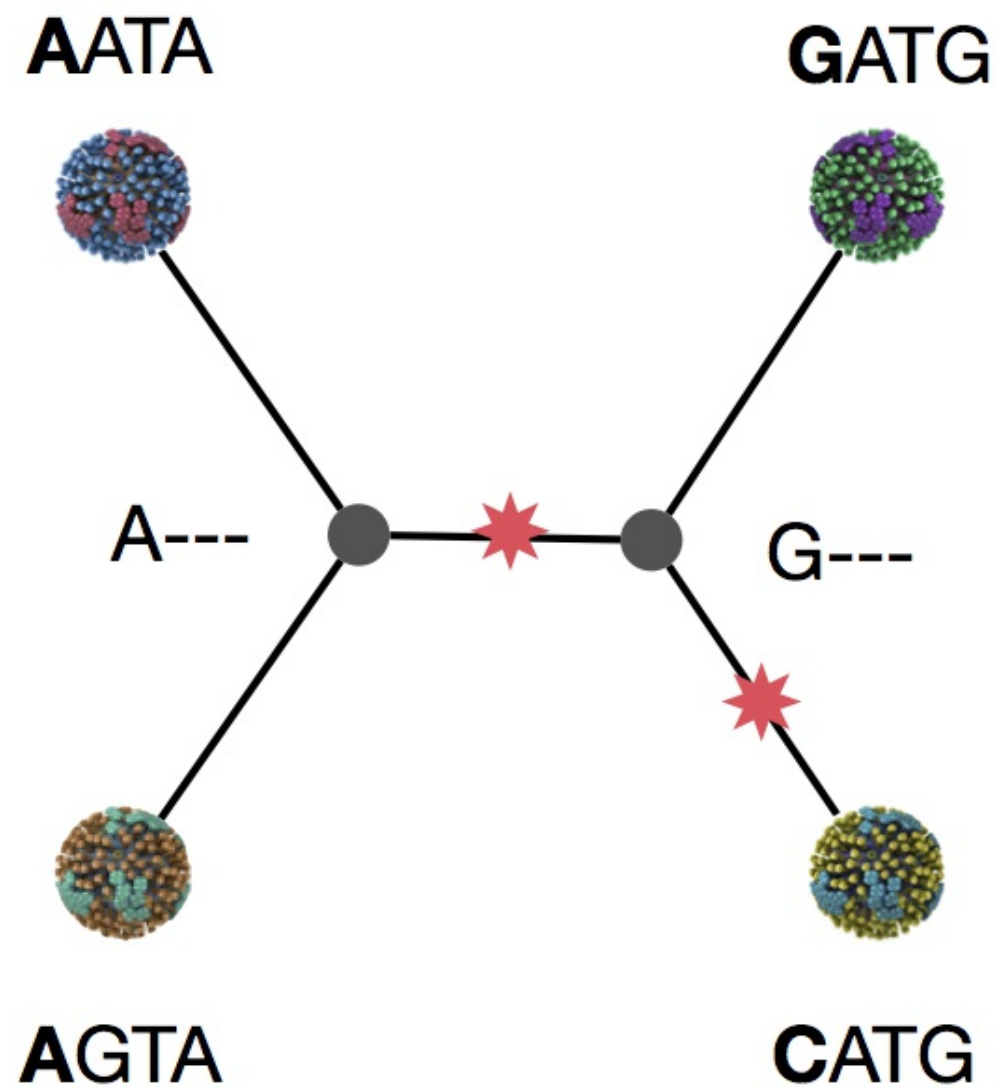
2+?+?+? mutations

Tree #2



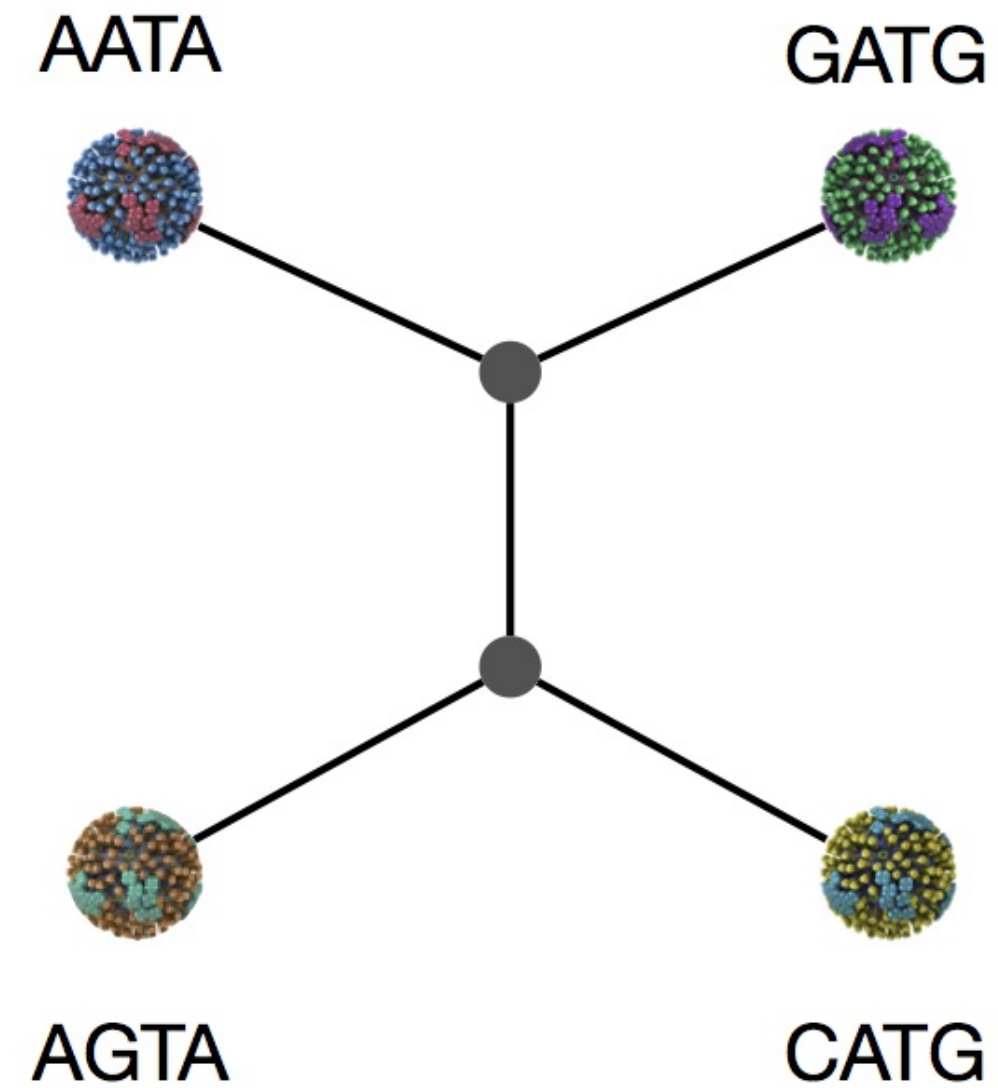
Exercise: which topology is more likely under parsimony?

Tree #1



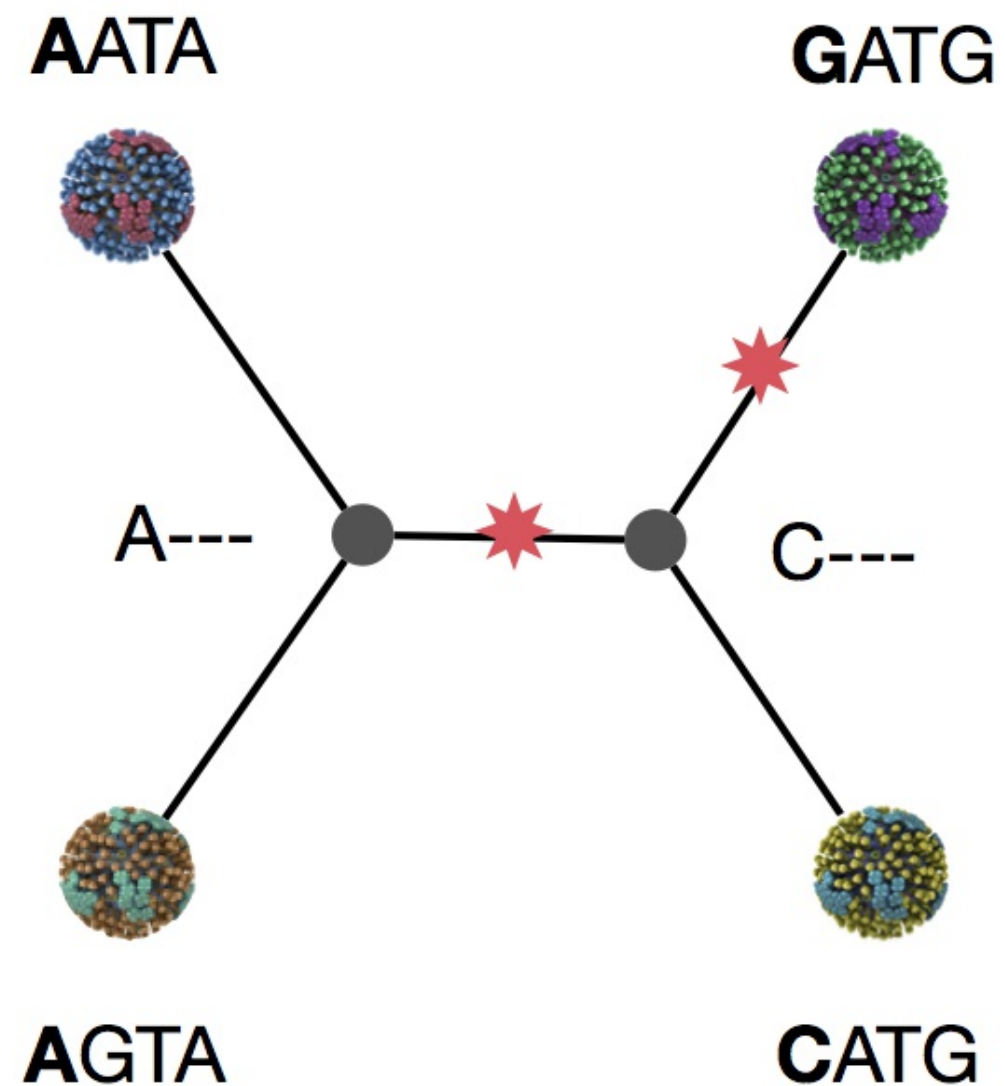
2+?+?+? mutations

Tree #2



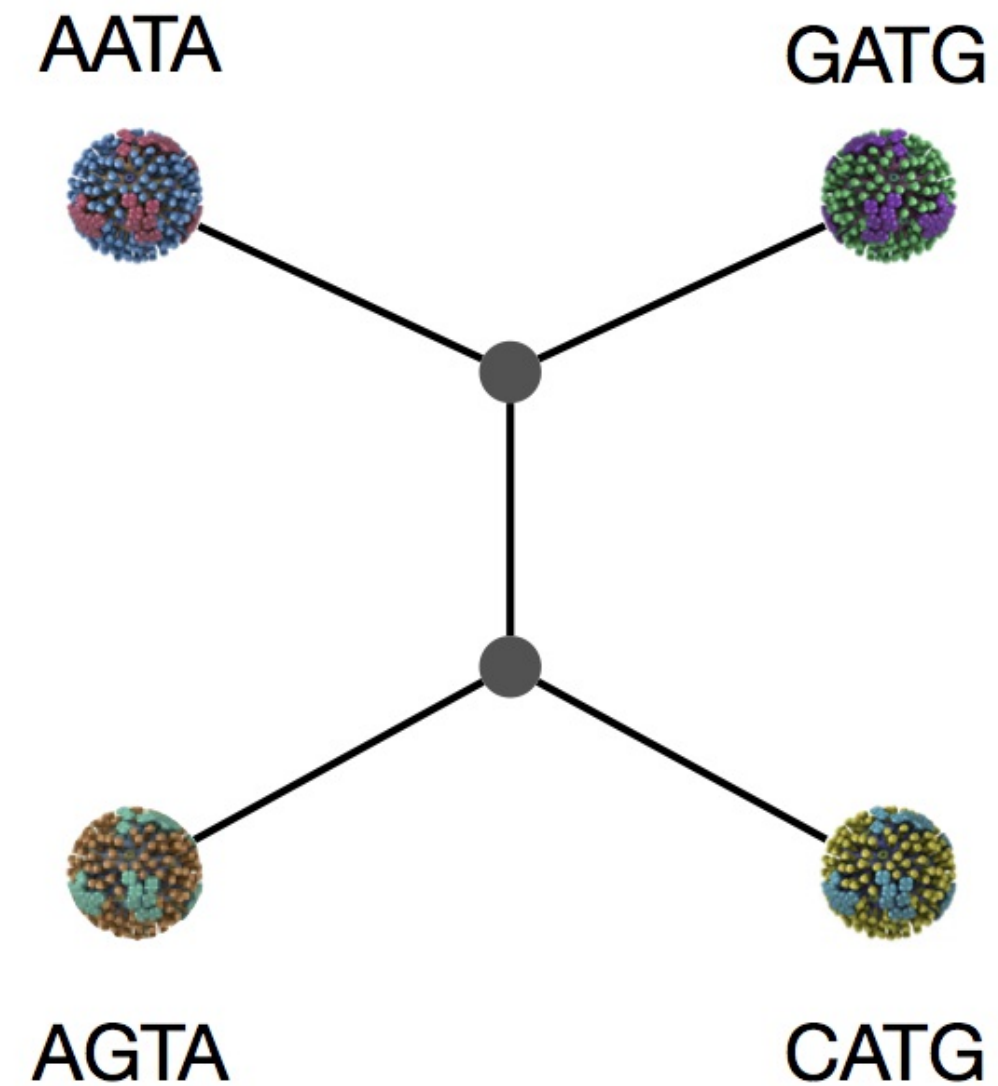
Exercise: which topology is more likely under parsimony?

Tree #1



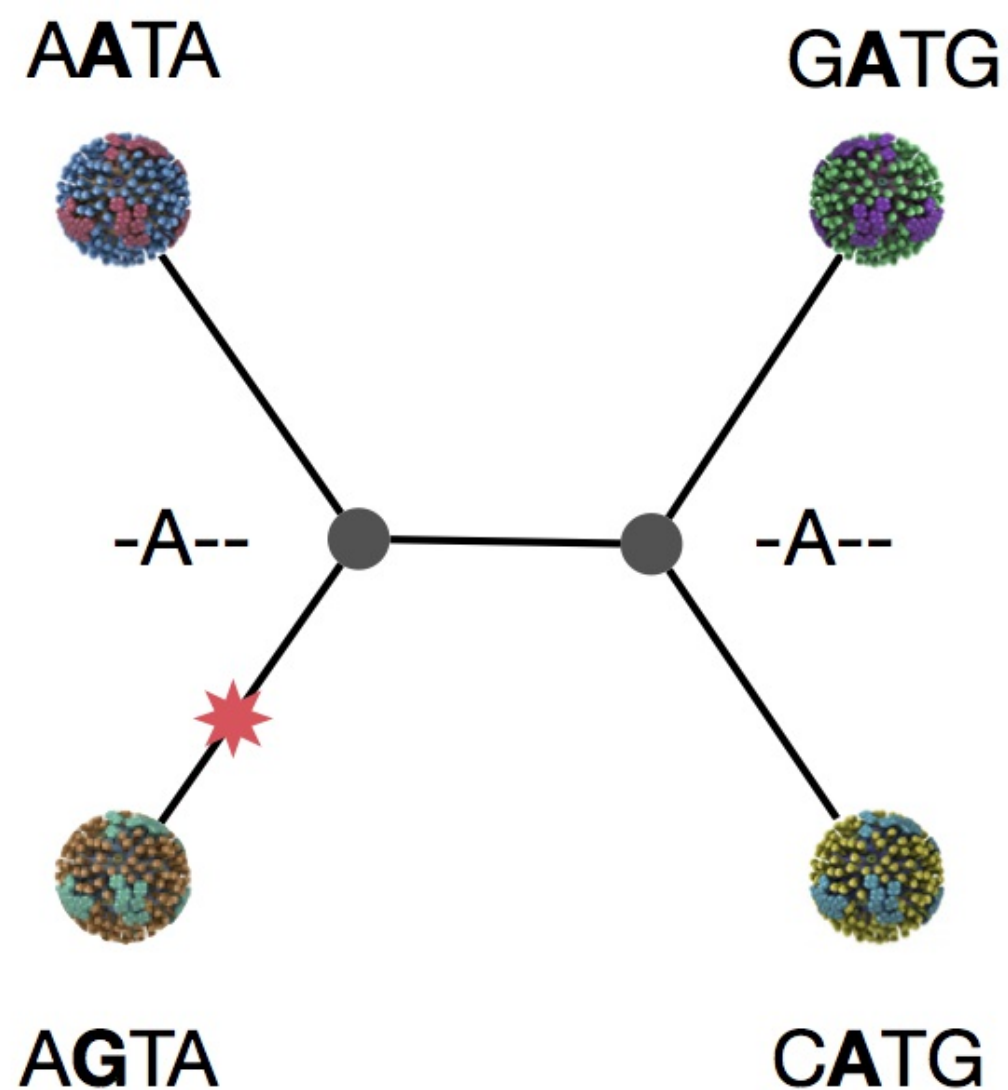
2+?+?+? mutations

Tree #2



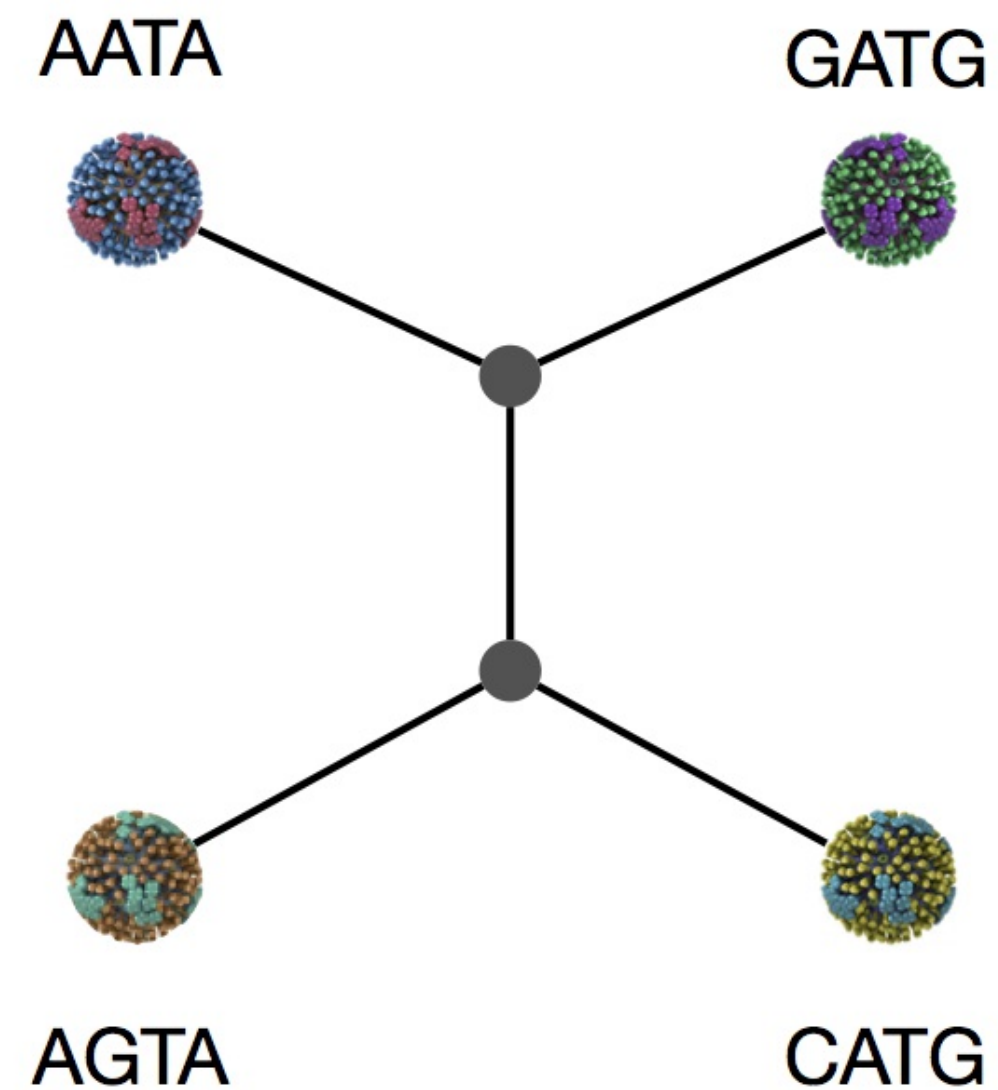
Exercise: which topology is more likely under parsimony?

Tree #1



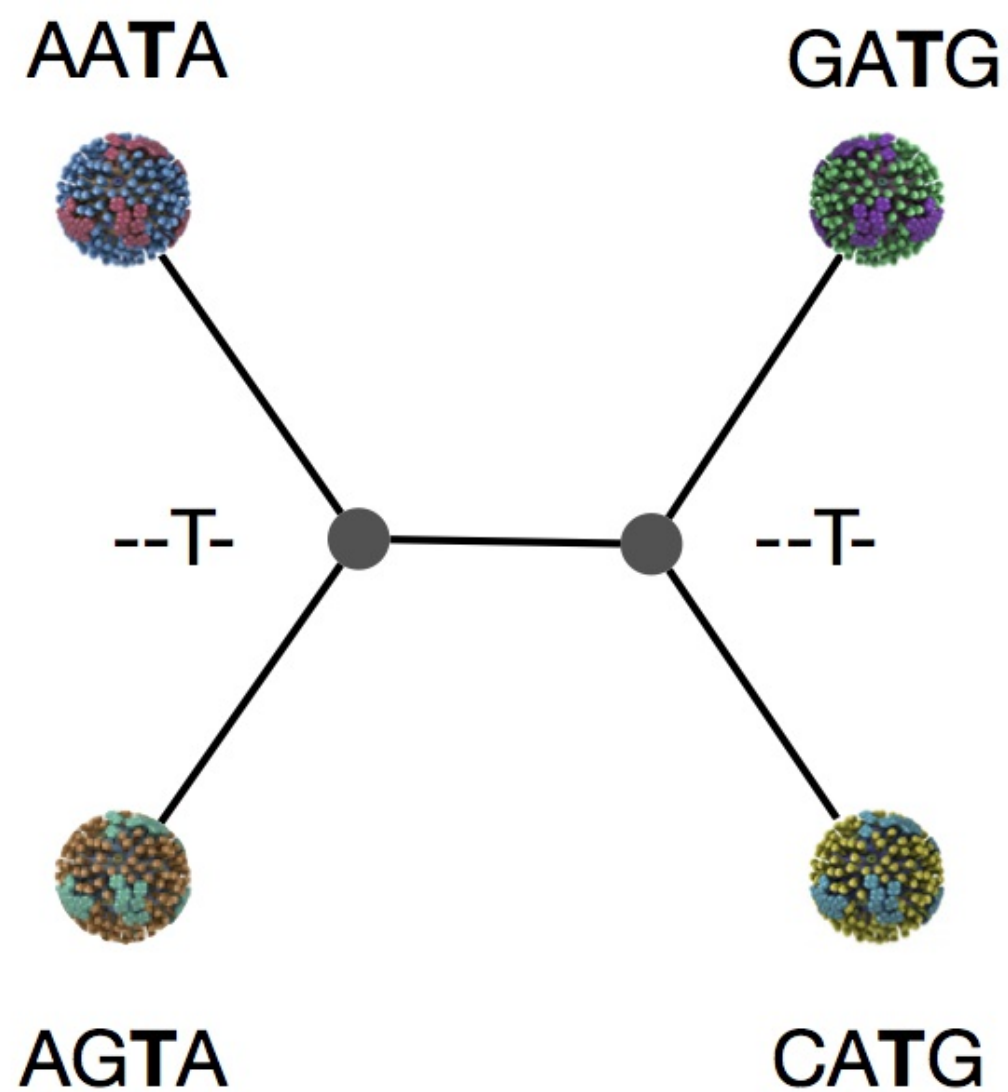
2+1+?+? mutations

Tree #2



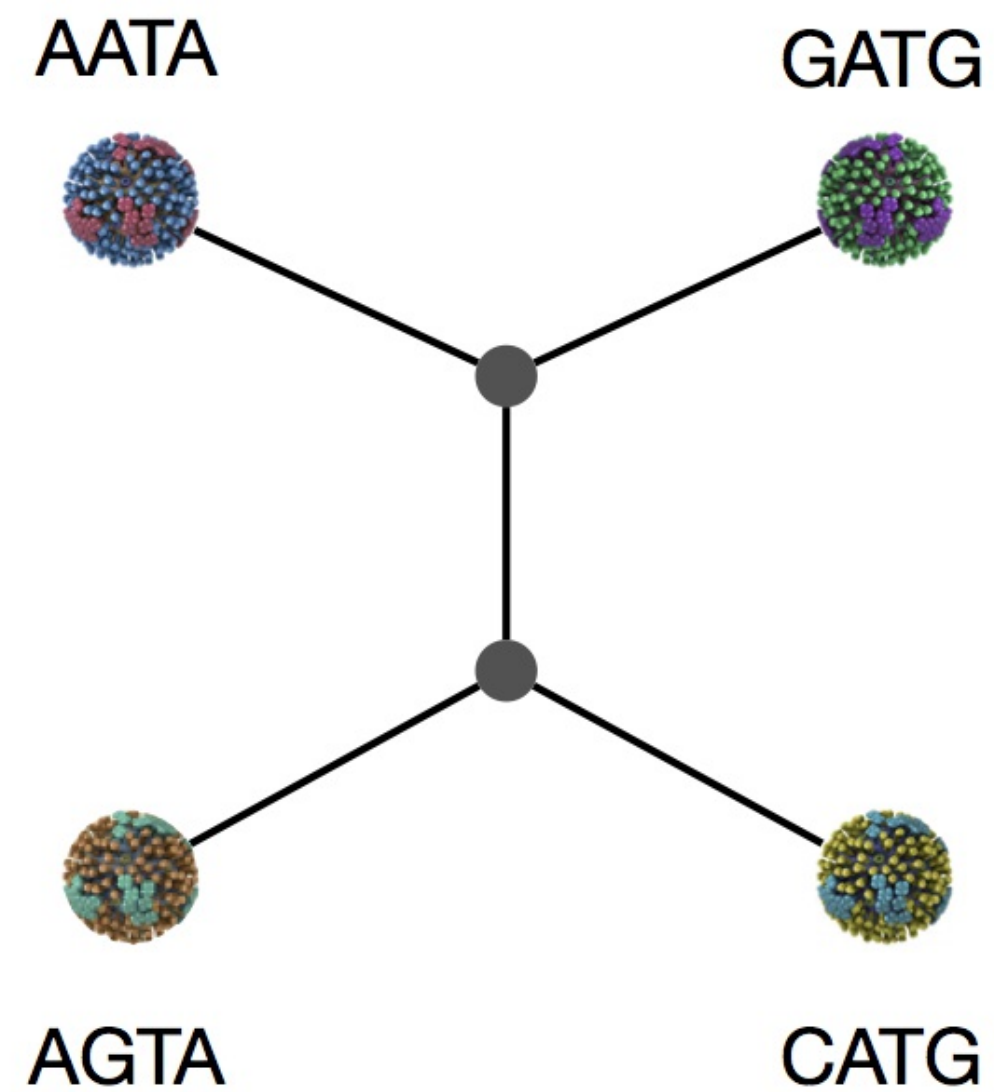
Exercise: which topology is more likely under parsimony?

Tree #1



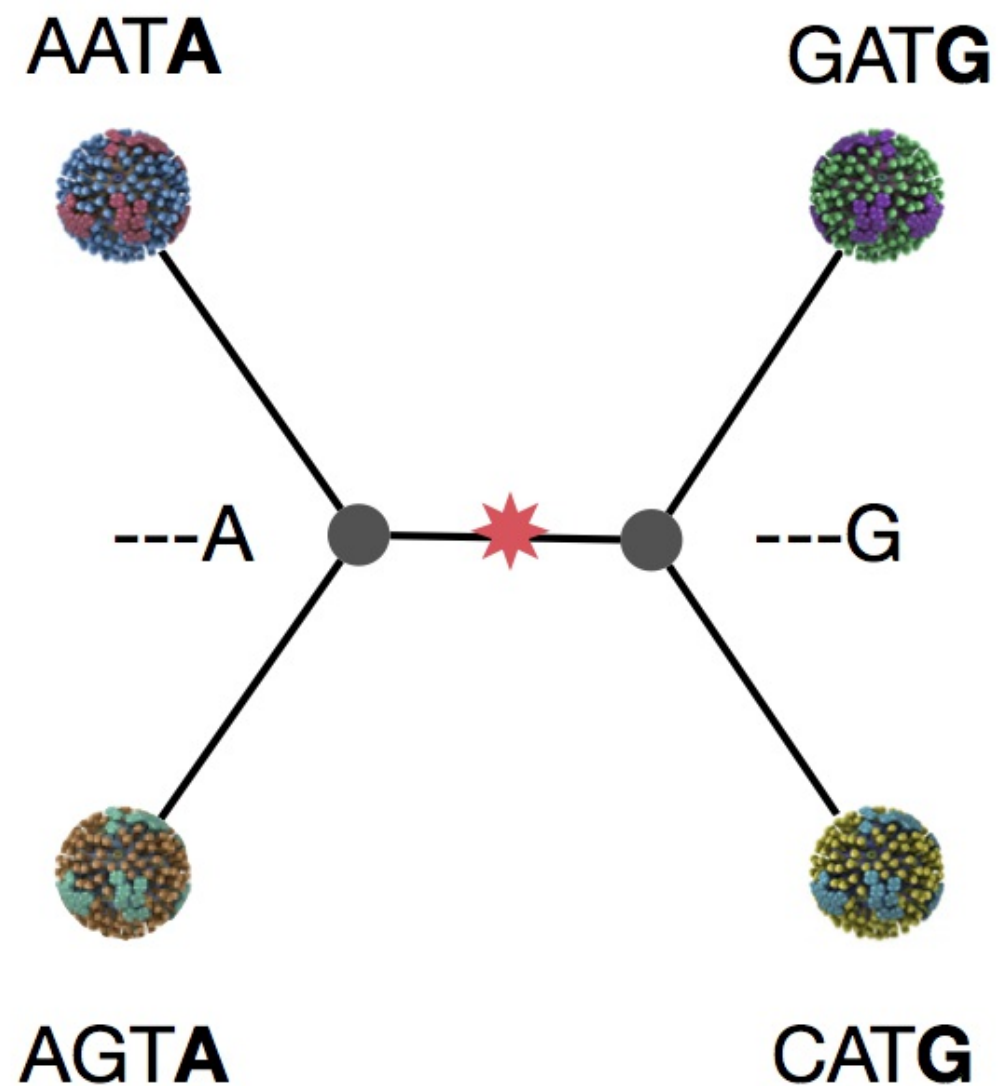
2+1+0+? mutations

Tree #2



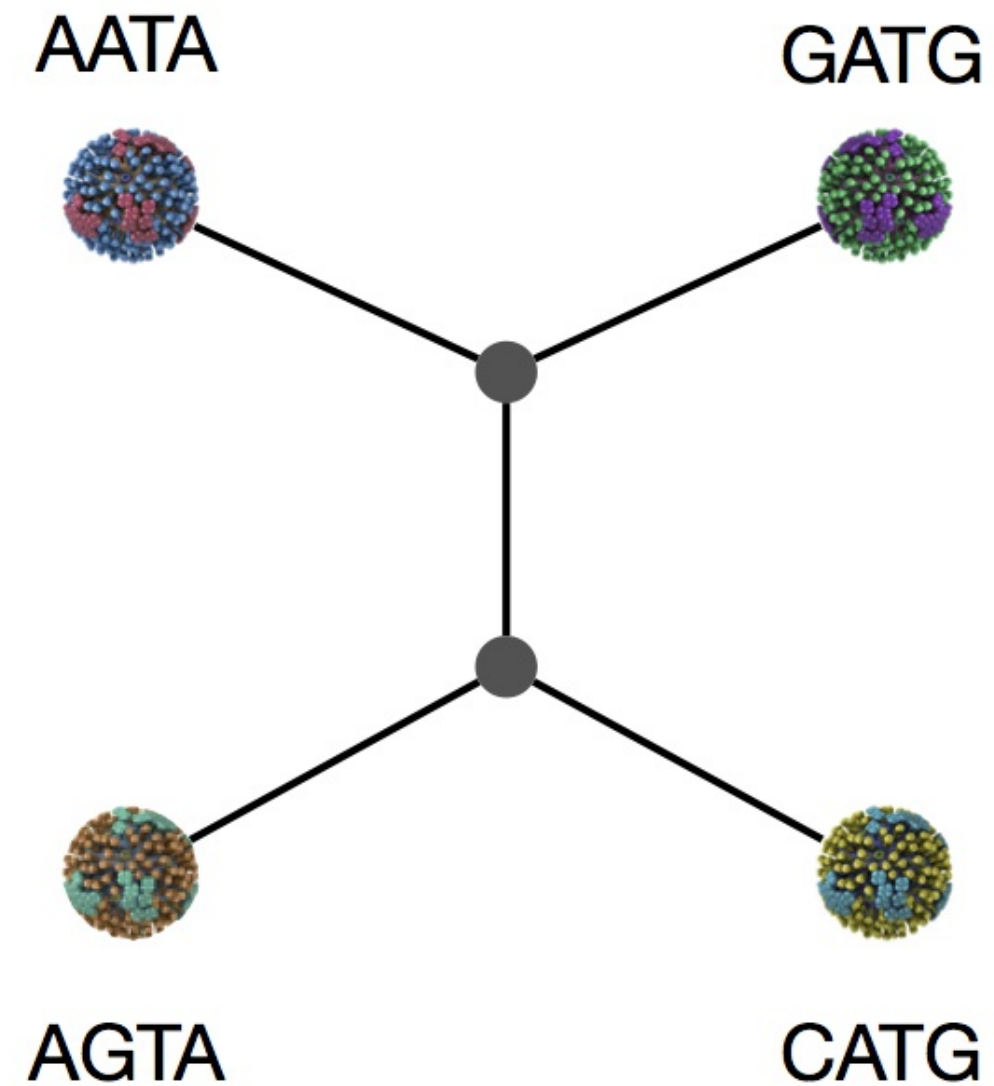
Exercise: which topology is more likely under parsimony?

Tree #1



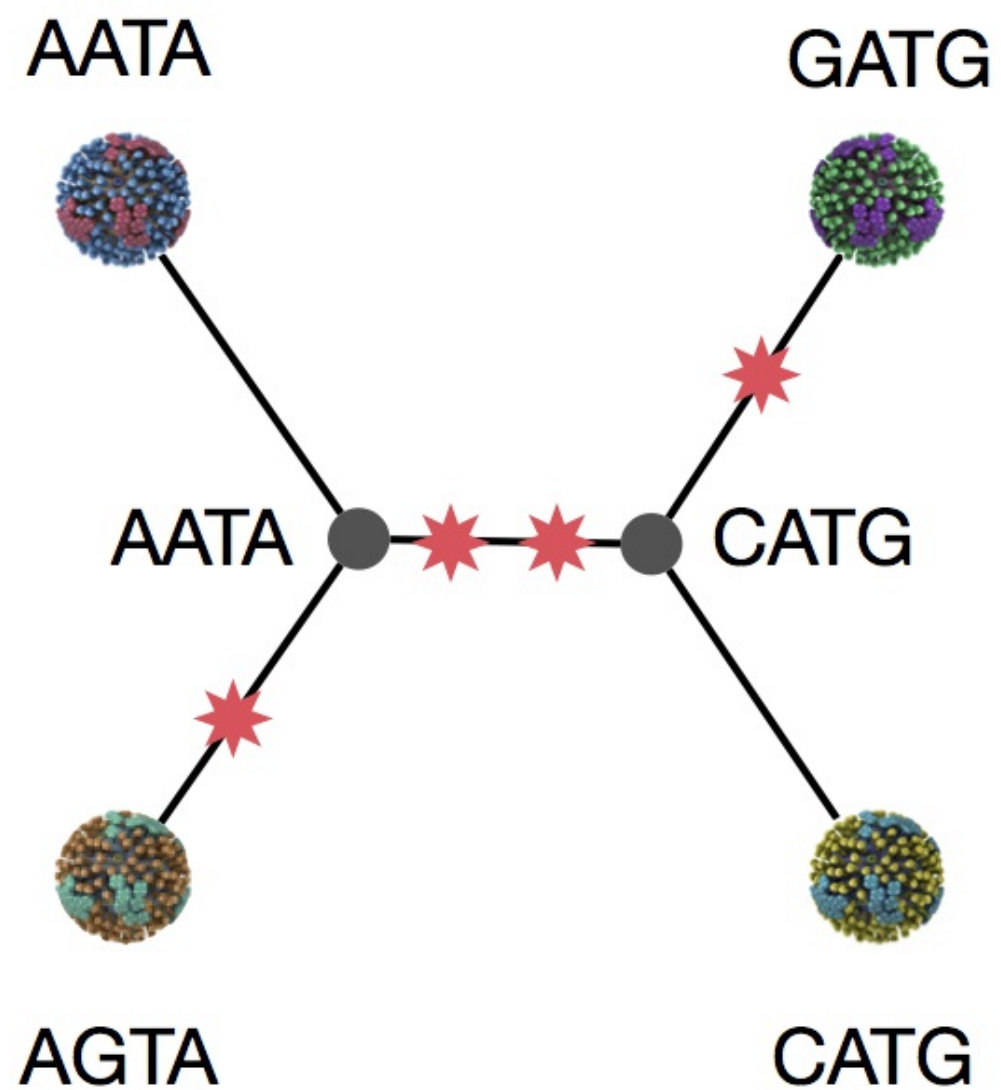
2+1+0+1 mutations

Tree #2



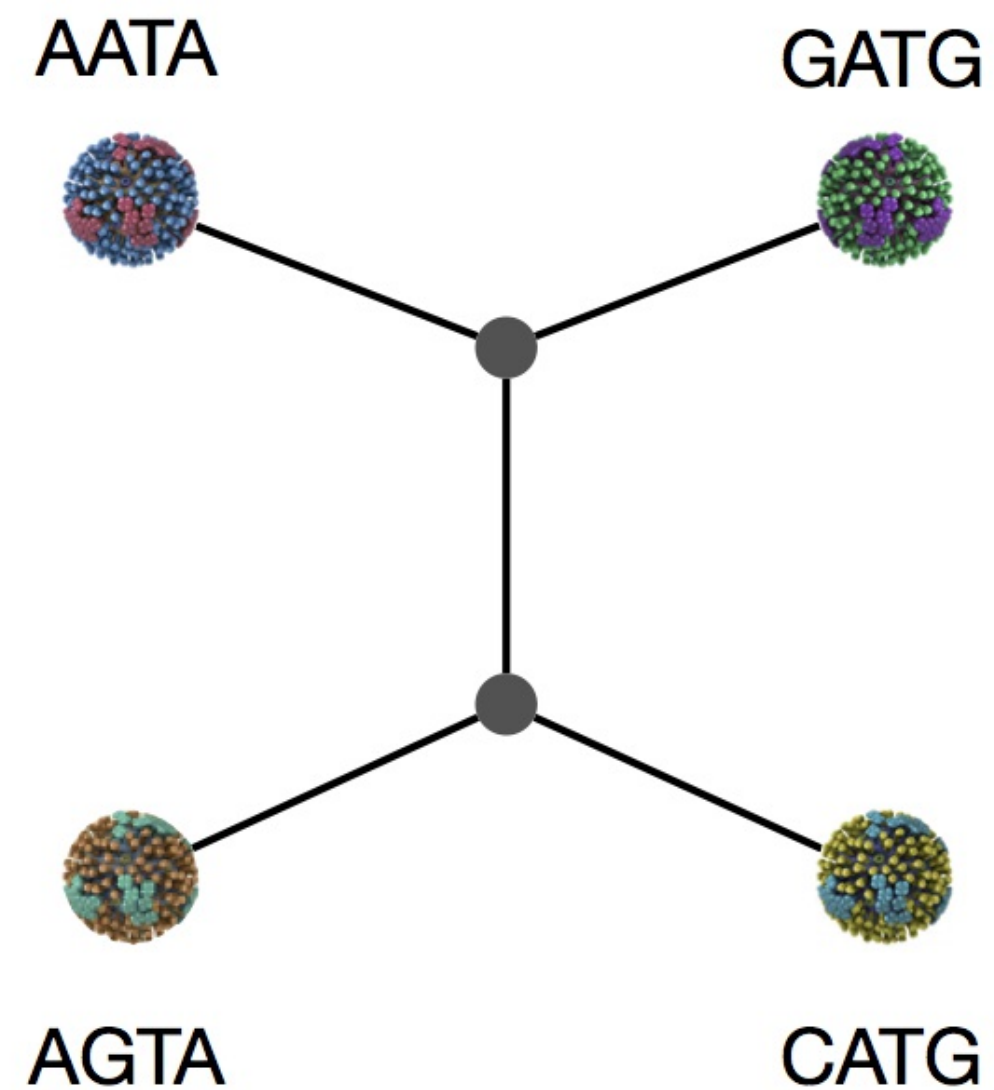
Exercise: which topology is more likely under parsimony?

Tree #1




4 mutations minimum

Tree #2



? mutations minimum



Example of how phylogeny structures site patterns

Buffalo. 2015

Inferring phylogenies with likelihood

Maximum likelihood (ML) inference

In ML, you have some set of data \mathbf{D} and a model for generating this data. This model has parameters θ . The probability of observing data is $\Pr(\mathbf{D} \mid \theta)$. The best parameter point estimate $\hat{\theta}$ is simply the value that maximizes $\Pr(\mathbf{D} \mid \theta)$.

Maximum likelihood (ML) inference

For example, if we have data \mathbf{D} from a Bernoulli observation model representing k successes in n trials, then the probability of observing k and n given coin flip probability parameter p is simply

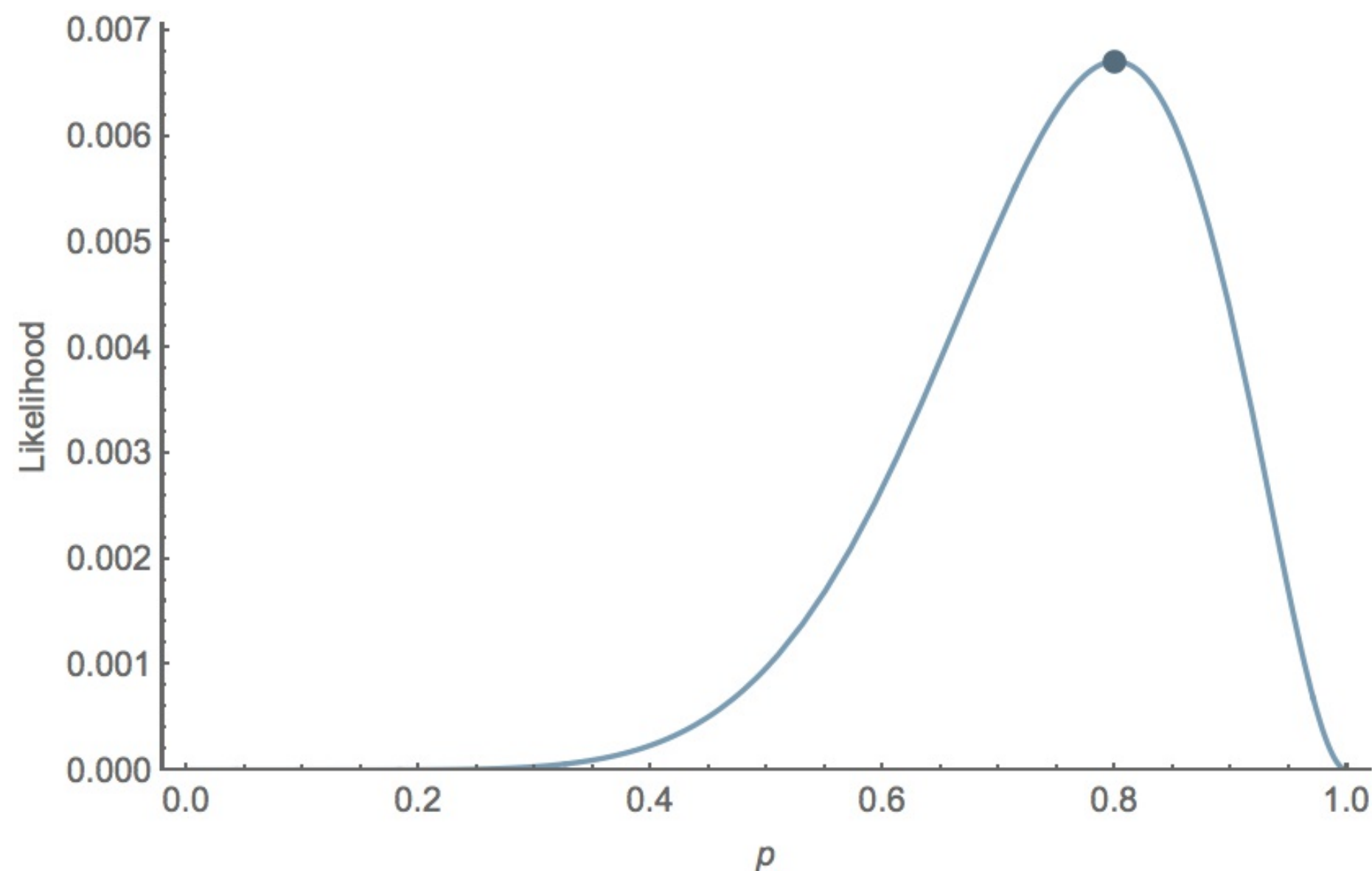
$$\Pr(k, n | p) = p^k (1 - p)^{n-k}.$$

Maximum likelihood (ML) inference

For the Bernoulli model $\Pr(k, n \mid p) = p^k (1 - p)^{n-k}$, we have

$\hat{p} = \frac{k}{n}$. For example, with $k = 8$ and $n = 10$, $\hat{p} = 0.8$ and the

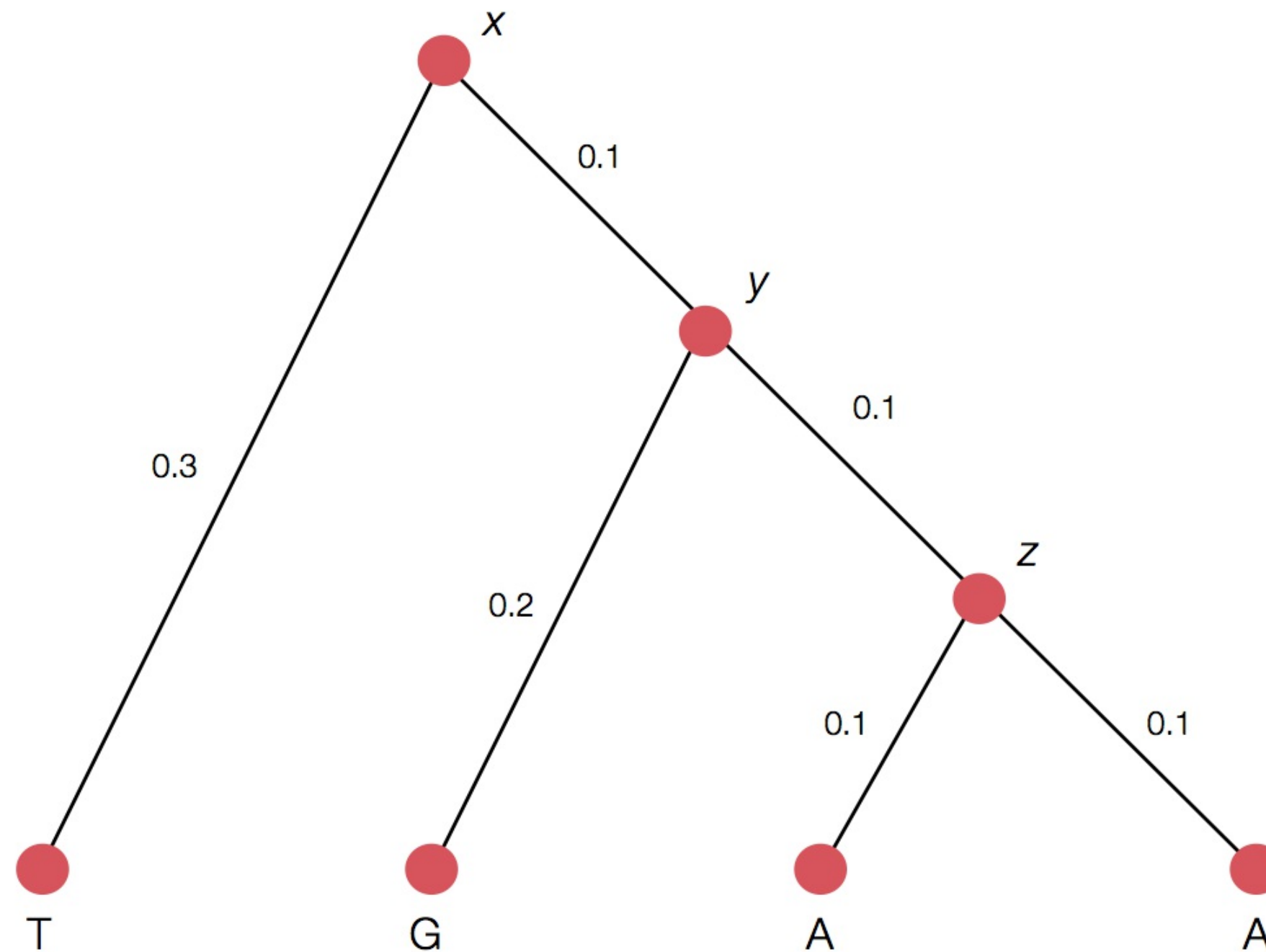
likelihood curve follows



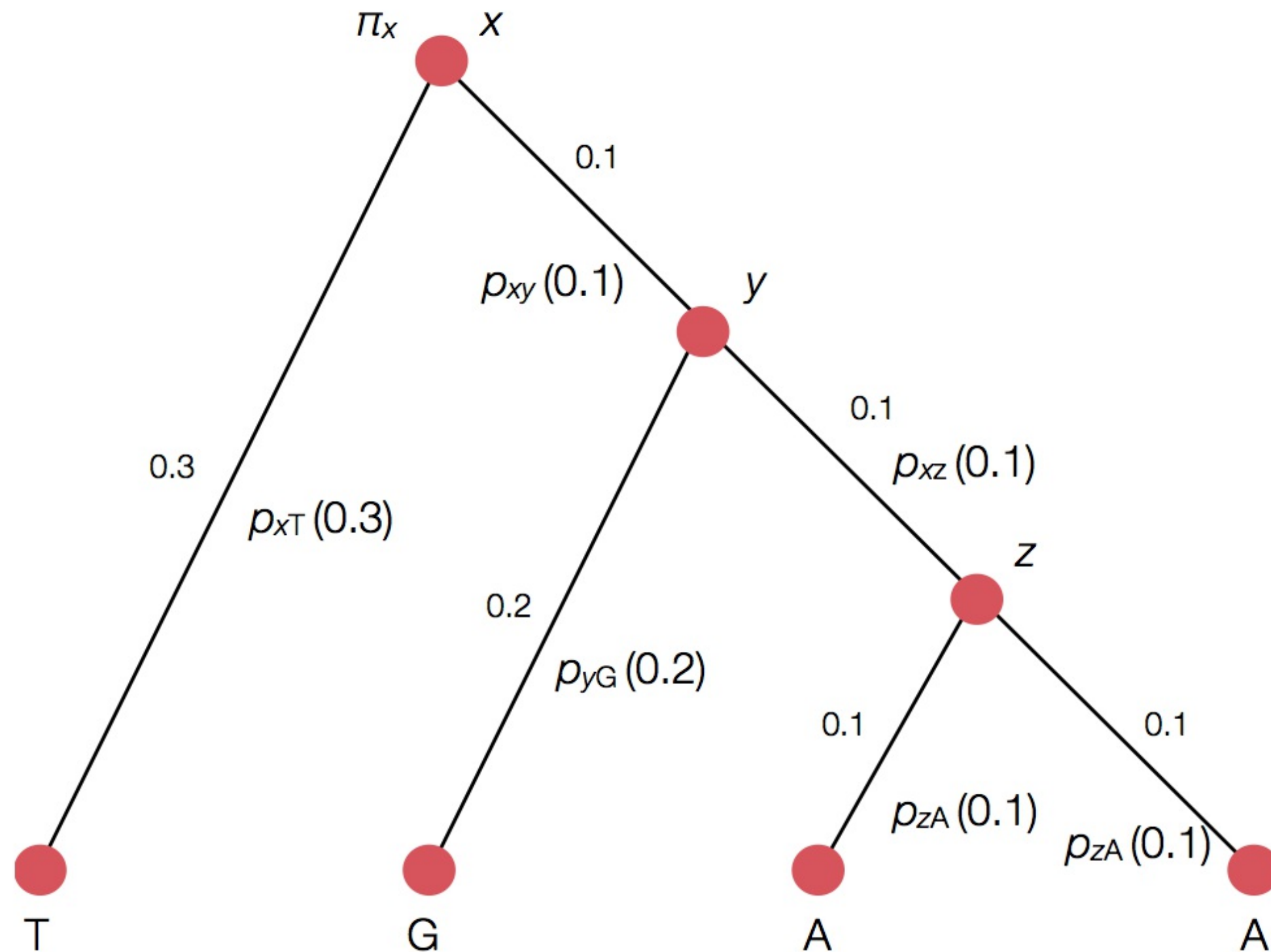
Data likelihood

In phylogenetics, \mathbf{D} are the observed tip sequences and θ is the phylogenetic tree including topology and branch lengths.

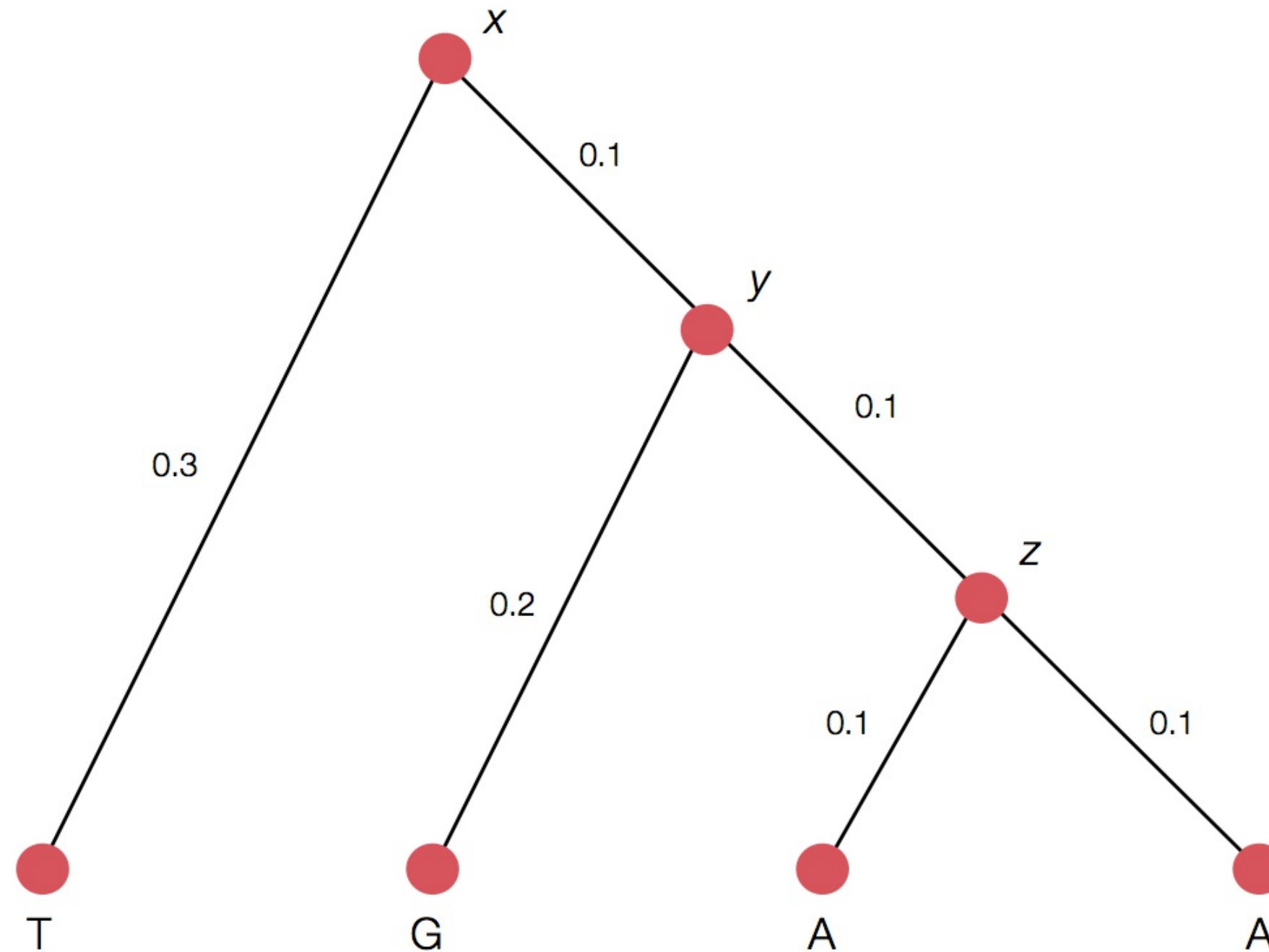
Calculating the likelihood of a single topology



Transition probability for each branch



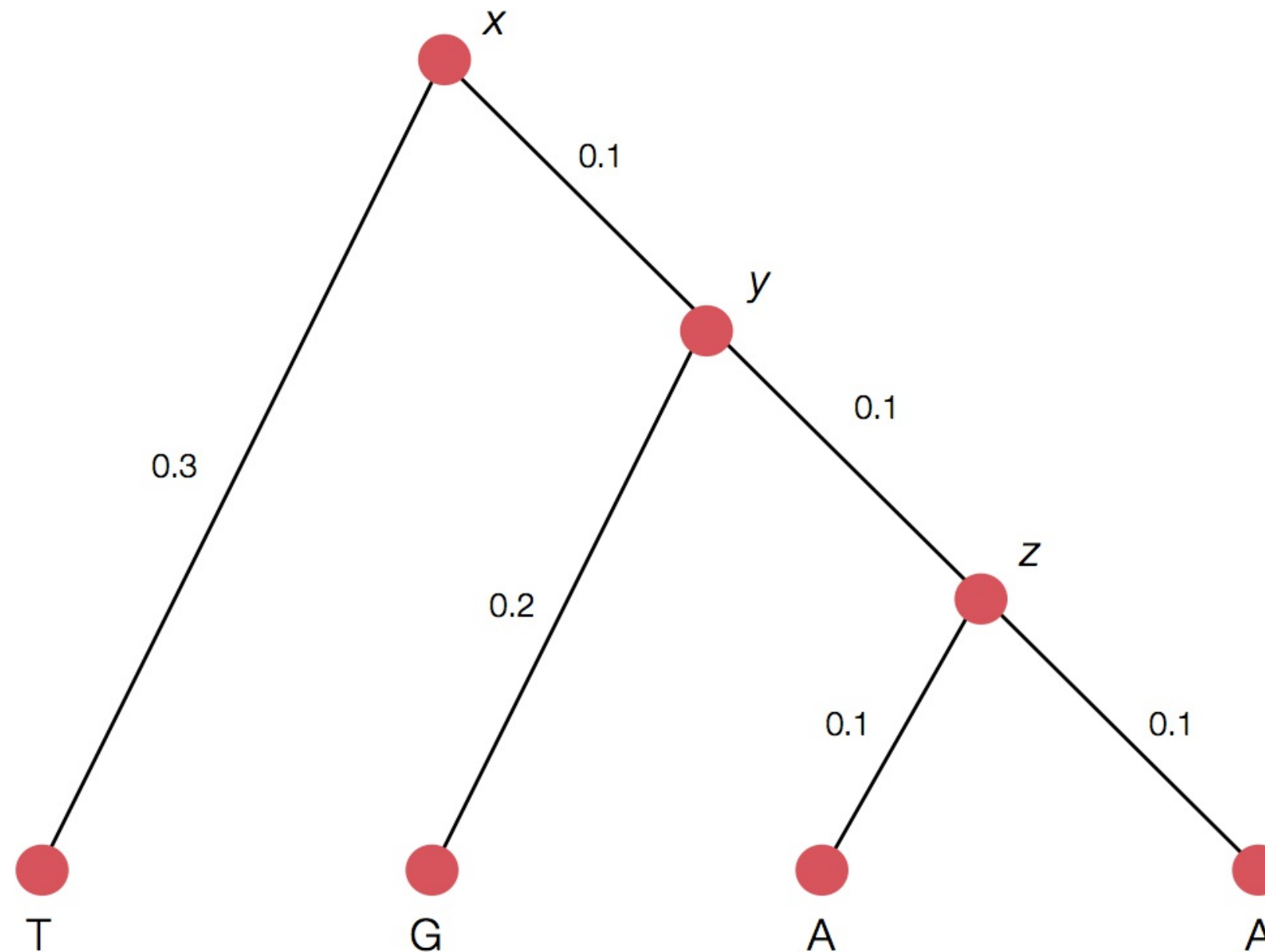
Rather than enumerating all possible
internal states



$$L(\mathbf{x}) = \sum_x \sum_y \sum_z \pi_x p_{xT}(0.3) p_{yG}(0.2) p_{xy}(0.1) p_{xz}(0.1) p_{zA}(0.1) p_{zA}(0.1)$$

$4 \times 4 \times 4 \times 7 = 448$ transition matrices

Can use Felsenstein's pruning algorithm to speed up calculation



$$L(\mathbf{x}) = \left(\sum_x \pi_x p_{xT}(0.3) \left(\sum_y p_{yG}(0.2) p_{xy}(0.1) \left(\sum_z p_{xz}(0.1) p_{zA}(0.1) p_{zA}(0.1) \right) \right) \right)$$

$(4 \times 2) + (4 \times 2) + (4 \times 3) = 28$ transition matrices



ML inference

Inference becomes a search for the tree that maximizes the likelihood of observing tip sequences. Lots of computation goes into this.

Bayesian phylogenetic inference



Bayesian inference

Generally, it's difficult to make probability statements using frequentist statistics. You cannot directly say that model 1 is twice as likely as model 2. People misuse p values in this sort of fashion all the time.

Bayes' rule

Bayes' rule forms the basis of Bayesian inference, it states:

$$\Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B)}$$

Bayes' rule exercise

For example, let's say we have an Ebola test that is 99% **sensitive** and 99% **specific** (meaning if someone has Ebola it will report true 99% of the time and if someone doesn't have Ebola it will report false 99% of the time). Let's further say that 0.1% of the population has Ebola. If we select a random individual and observe a positive test result, what is the probability that they actually have Ebola?

Bayes' rule exercise

If we select a random individual and observe a positive test result, what is the probability that they actually have Ebola? I.e.

$$\begin{aligned}\Pr(\mathbf{P} \mid \mathbf{E}) &= 0.99, \Pr(\mathbf{N} \mid \mathbf{E}) = 0.01, \\ \Pr(\mathbf{P} \mid \mathbf{NE}) &= 0.01, \Pr(\mathbf{N} \mid \mathbf{NE}) = 0.99 \\ \text{and } \Pr(\mathbf{E}) &= 0.001.\end{aligned}$$

In this case, what is $\Pr(\mathbf{E} \mid \mathbf{P})$?

Bayesian inference

Bayesian inference applies Bayes' rule in a likelihood context, so

that

$$\Pr(\theta | D) = \frac{\Pr(D | \theta) \Pr(\theta)}{\Pr(D)},$$

where D is data and θ are parameters. $\Pr(D)$ is constant with respect to θ , so that $\Pr(\theta | D) \propto \Pr(D | \theta) \Pr(\theta)$. This

relationship is often referred to as

posterior \propto likelihood \times prior.

Bayesian inference for Bernoulli model

Following our previous Bernoulli example, we've observed k successes in n trials, and so the likelihood

$\Pr(k, n \mid p) = p^k (1 - p)^{n-k}$. We'll assume a flat prior

$\Pr(p) = 1$. In this case, the marginal likelihood follows

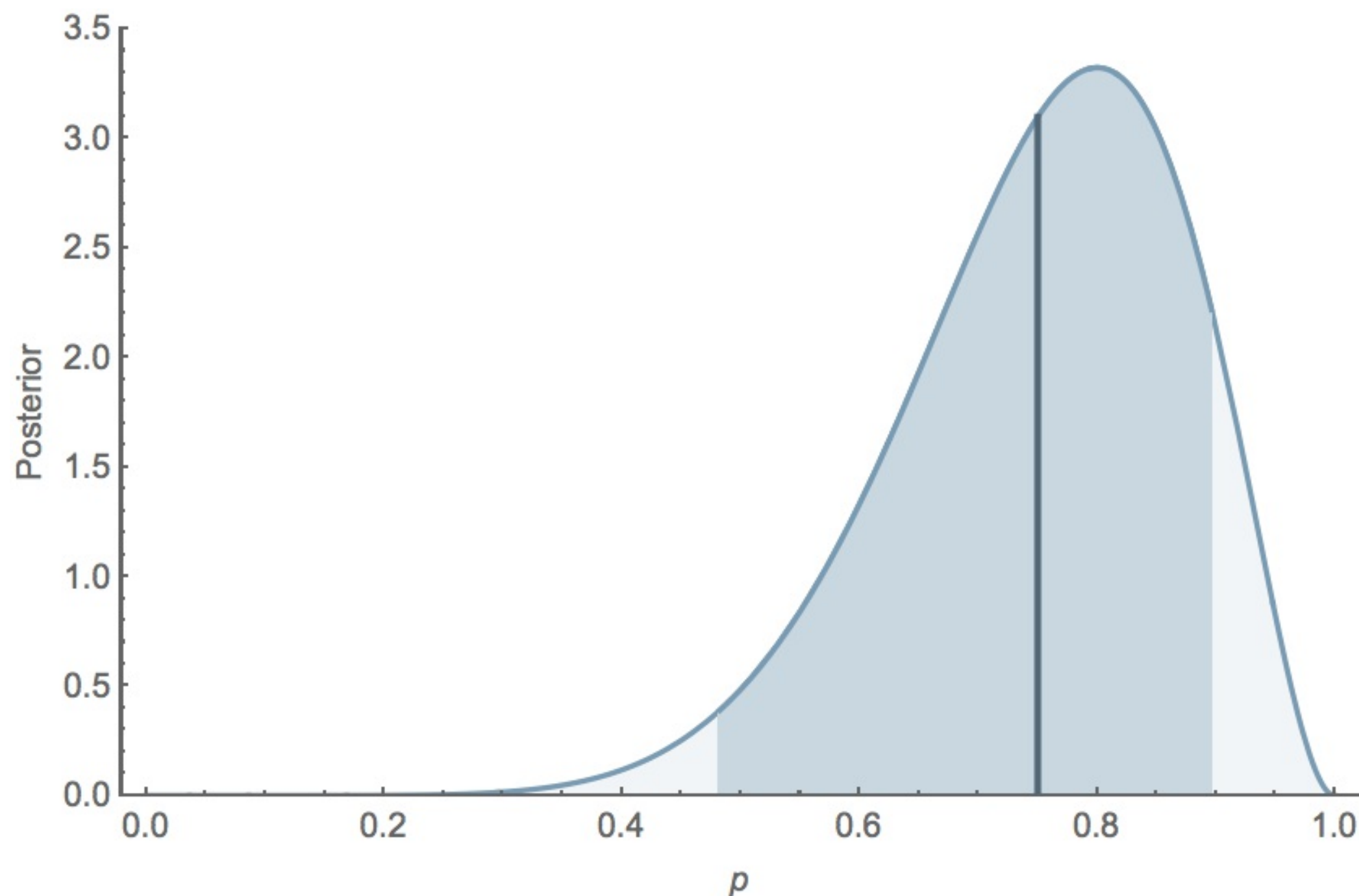
$$\Pr(k, n) = \int_0^1 \Pr(k, n \mid p) \Pr(p) dp = \frac{k! (n - k)!}{(n + 1)!}.$$

And the full posterior follows

$$\Pr(p \mid k, n) = \frac{(n + 1)! p^k (1 - p)^{n-k}}{k! (n - k)!}.$$

Probability statements

If $k = 8$ and $n = 10$, the mean posterior $\mathbf{E}[p] = 0.75$, while the 95% credible interval extends from **0.482** to **0.896**, and the posterior distribution follows



Bayesian phylogenetic inference

Here, we are interested in the posterior distribution $\Pr(\tau, \mu \mid D)$, where D represents sequence data, τ represents the tree topology and μ represents mutational parameters (like transition vs tranversion rate). In this case,

$$\Pr(\tau, \mu \mid D) \propto \Pr(D \mid \tau, \mu) \Pr(\tau) \Pr(\mu).$$

Bayesian coalescent inference

In the case of the coalescent model, we are interested in coalescent rate parameter like λ . Here, we use λ to give the likelihood of observing a particular tree topology $\Pr(\tau \mid \lambda)$. This probability is the likelihood of observing the coalescent intervals seen in the tree.

Bayesian coalescent inference

Thus, the full model becomes

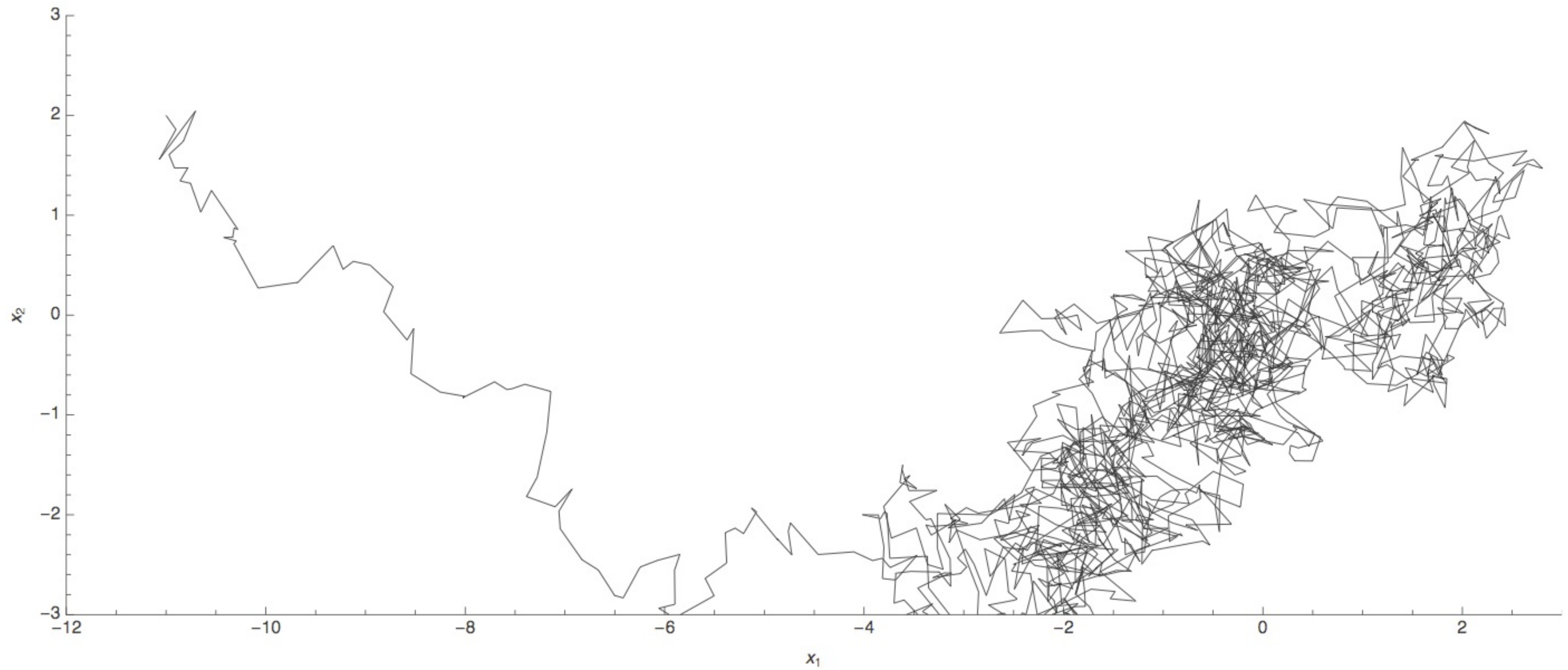
$$\Pr(\tau, \mu, \lambda \mid D) \propto \Pr(D \mid \tau, \mu) \Pr(\tau \mid \lambda) \Pr(\lambda) \Pr(\mu).$$

Bayesian approaches work well to build these sorts of nested models.



BEAST: Bayesian Evolutionary Analysis by Sampling Trees

Markov chain Monte Carlo (MCMC)



Markov chain Monte Carlo (MCMC)

