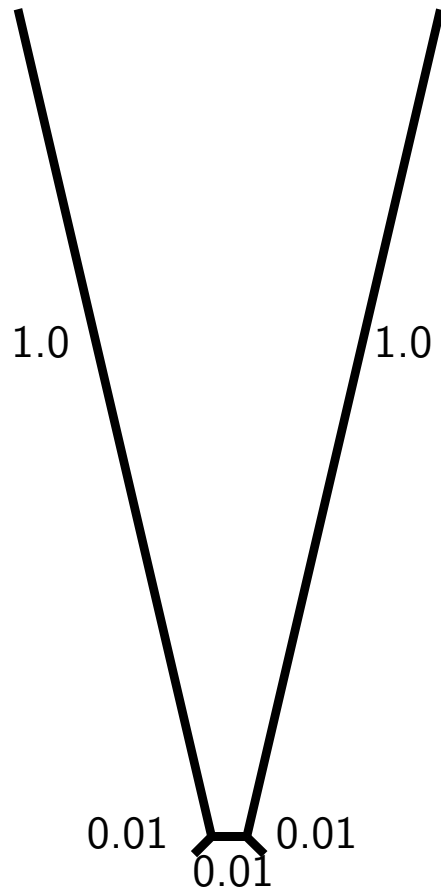


Qualitative description of parsimony

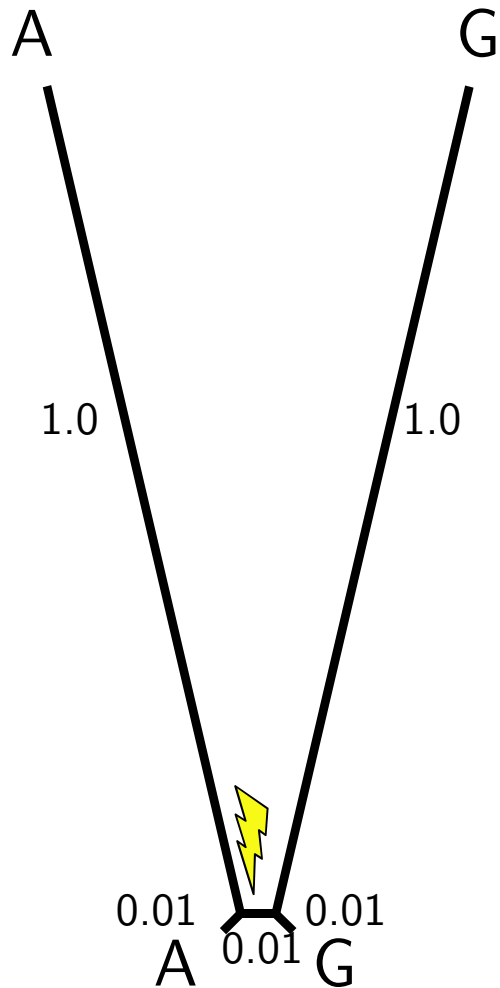
- Enables estimation of ancestral sequences.
- Even though parsimony always seeks to minimize the number of changes, it can perform well even when changes are not rare.
- Does not “prefer” to put changes on one branch over another
- Hard to characterize statistically
 - the set of conditions in which parsimony is guaranteed to work well is very restrictive (low probability of change and not too much branch length heterogeneity);
 - Parsimony often performs well in simulation studies (even when outside the zones in which it is guaranteed to work);
 - Estimates of the tree can be extremely biased.

Long branch attraction



Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401-410.

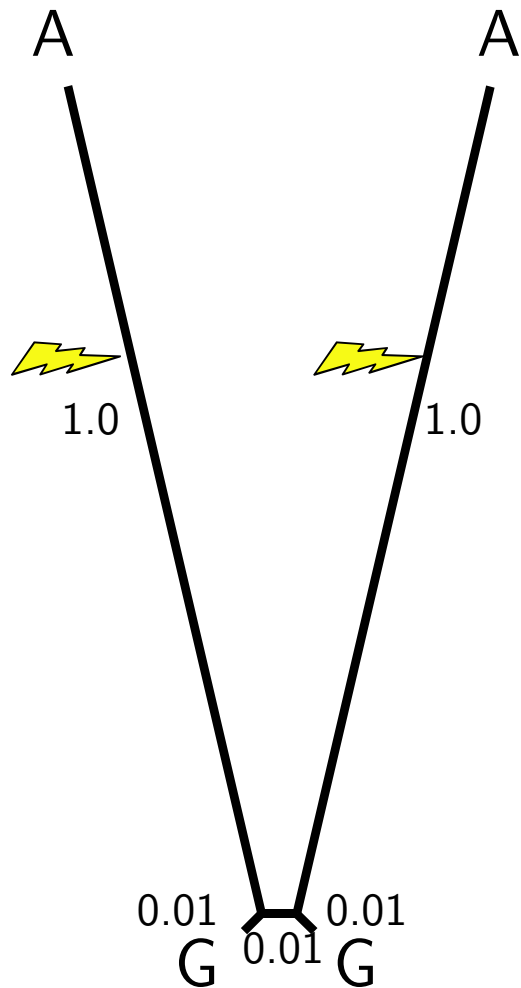
Long branch attraction



Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401-410.

The probability of a parsimony informative site due to inheritance is very low, (roughly 0.0003).

Long branch attraction



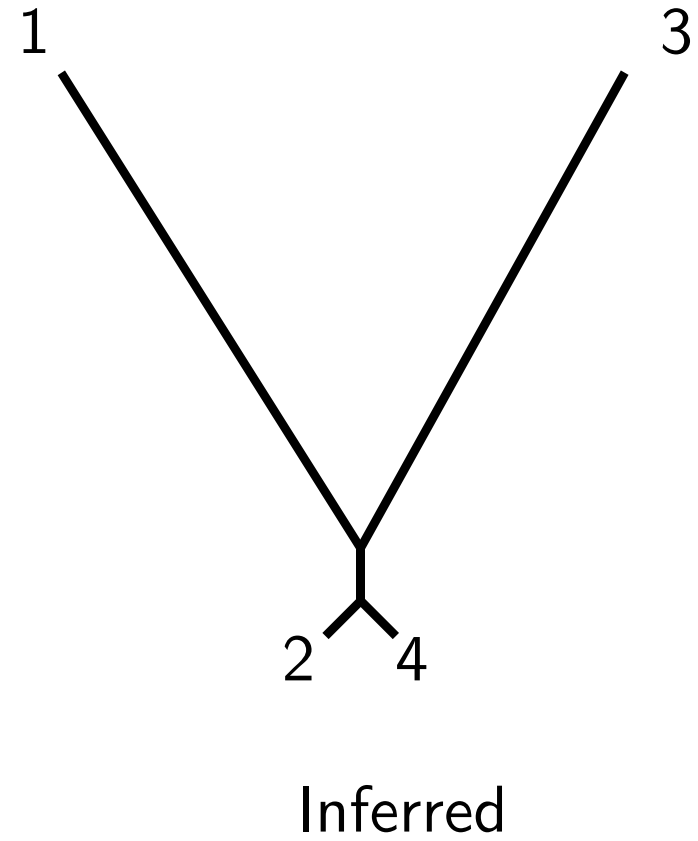
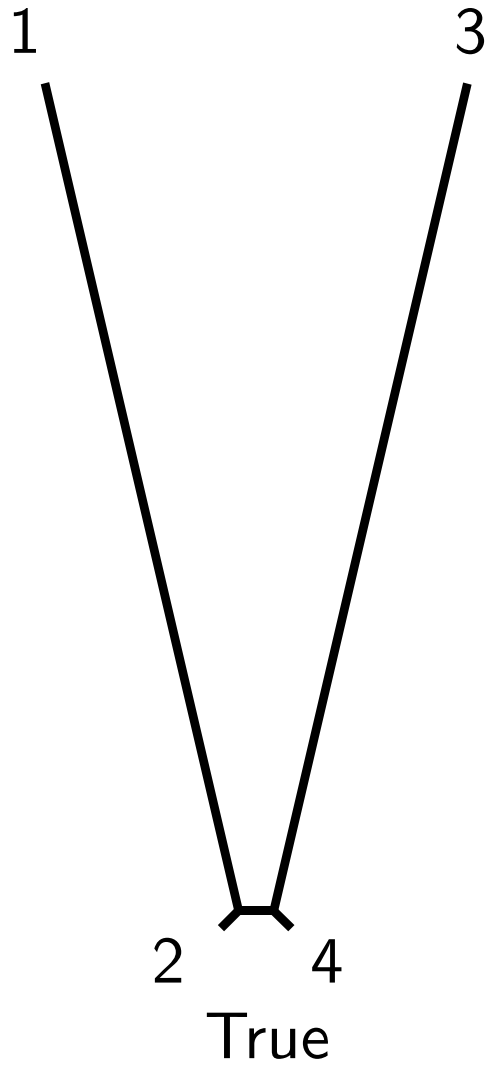
Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401-410.

The probability of a parsimony informative site due to inheritance is very low, (roughly 0.0003).

The probability of a misleading parsimony informative site due to parallelism is much higher (roughly 0.008).

Long branch attraction

Parsimony is almost guaranteed to get this tree wrong.

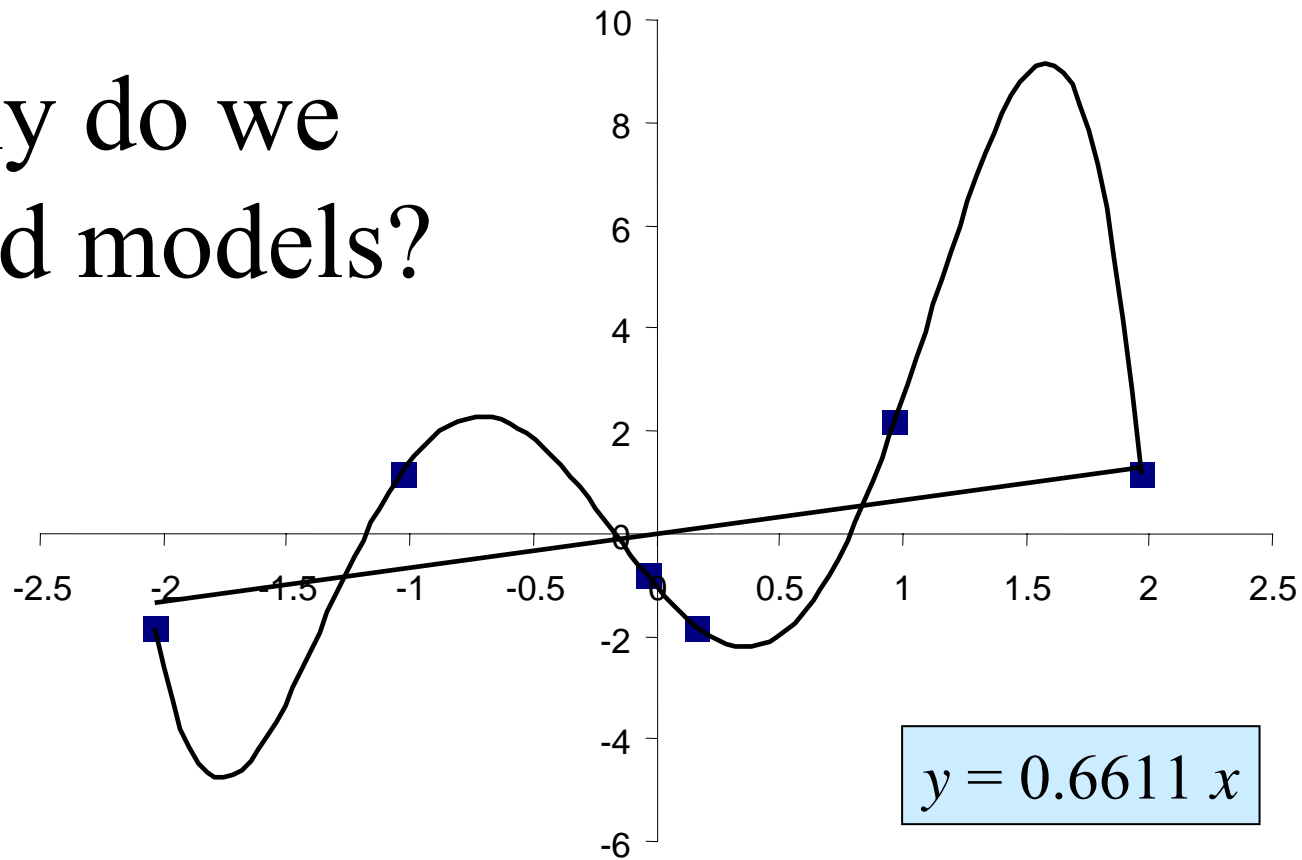


Inconsistency

- Statistical Consistency (roughly speaking) is converging to the true answer as the amount of data goes to ∞ .
- Parsimony based tree inference is *not* consistent for some tree shapes. In fact it can be “positively misleading”:
 - “Felsenstein zone” tree
 - Many clocklike trees with short internal branch lengths and long terminal branches (Penny *et al.*, 1989, Huelsenbeck and Lander, 2003).
- Methods for assessing confidence (e.g. bootstrapping) will indicate that you should be very confident in the wrong answer.

$$y = -1.5972 x^5 + 23.167 x^4 - 126.18 x^3 + 319.17 x^2 - 369.22 x + 155.67$$

Why do we
need models?



Models

- Models help us intelligently **interpolate between our observations** for purposes of **making predictions**
- **Adding parameters** to a model generally increases its fit to the data
- **Underparameterized** models lead to poor fit to observed data points
- **Overparameterized** models lead to poor prediction of future observations
- Criteria for choosing models include likelihood ratio tests, AIC, BIC, Bayes Factors, etc.
 - all provide a way to choose a model that is neither underparameterized nor overparameterized

The Poisson distribution

Probability distribution on the number of events when:

1. events are assumed to be independent,
2. the *rate* of events some constant, μ , and
3. the process continues for some duration of time, t .

The expectation of the number of events is $\nu = \mu t$.

Note that ν can be any non-negative number, but the Poisson is a discrete distribution – it gives the probabilities of the number of events (and this number will always be a non-negative integer).

The Poisson distribution

$$\Pr(k \text{ events} | \text{Expected \# is } \nu) = \frac{\nu^k e^{-\nu}}{k!}$$

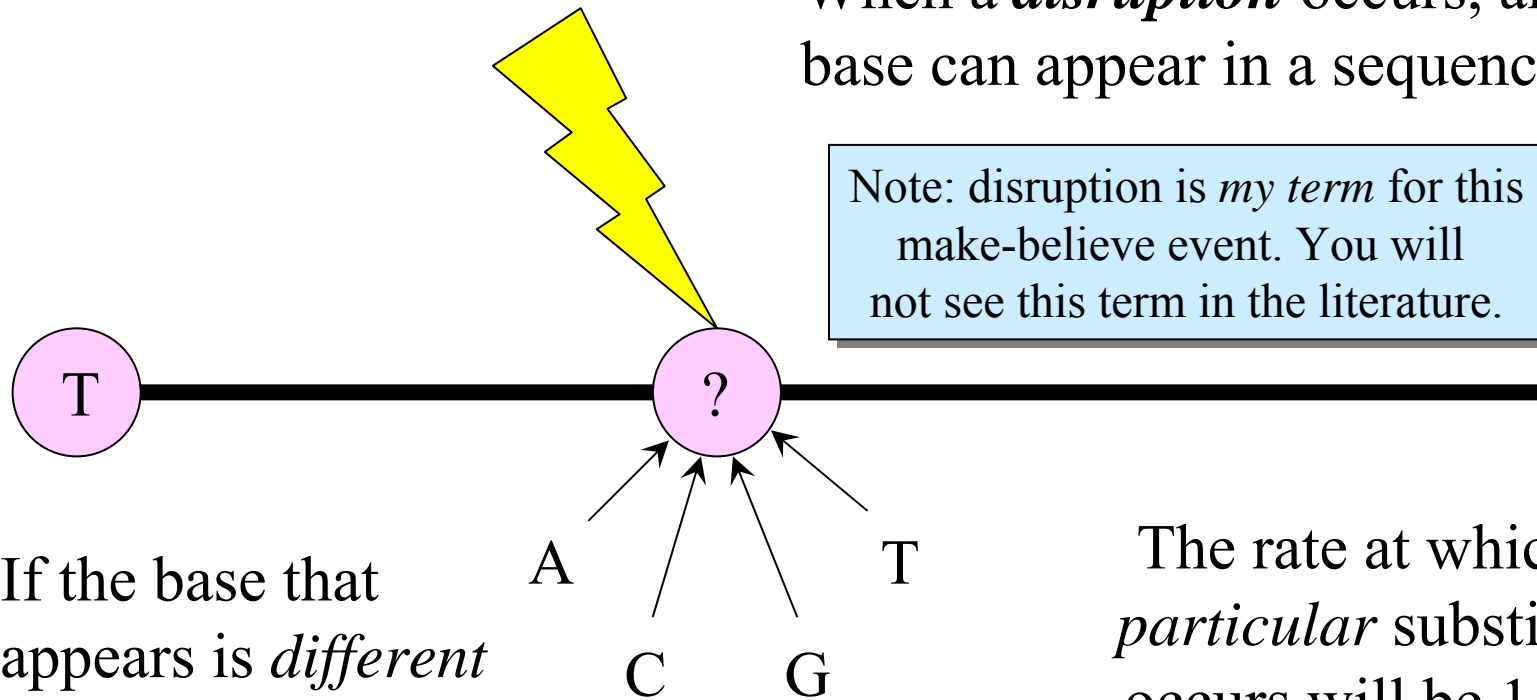
$$\Pr(0 \text{ events}) = \frac{\nu^0 e^{-\nu}}{0!} = e^{-\nu} = e^{-\mu t}$$

$$\Pr(\geq 1 \text{ events}) = 1 - e^{-\nu} = 1 - e^{-\mu t}$$

"Disruptions" vs. substitutions

When a *disruption* occurs, any base can appear in a sequence.

Note: disruption is *my term* for this make-believe event. You will not see this term in the literature.



If the base that appears is *different* from the base that was already there, then a **substitution** event has occurred.

The rate at which any *particular* substitution occurs will be 1/4 the disruption rate (assuming equal base frequencies)

Probability of T→G over time t

If μ is the rate of disruptions, and a branch is t units of time long then: Let's use θ for the rate of any particular "disruption."

$$\mu_{TA} = \mu_{TC} = \mu_{TG} = \mu_{TT} = \theta$$

$$\mu = 4\theta$$

Furthermore, given that there is a disruption the chance of any particular change is $\frac{1}{4}$

Probability of $T \rightarrow G$ over time t

$$\Pr(0 \text{ disruptions}|t) = e^{-\mu t}$$

$$\Pr(\text{at least 1 disruption}|t) = 1 - e^{-\mu t}$$

$$\Pr(\text{last disruption leads to } G) = 0.25$$

$$\begin{aligned}\Pr(T \rightarrow G|t) &= 0.25 (1 - e^{-\mu t}) \\ &= 0.25 (1 - e^{-4\theta t})\end{aligned}$$

JC69 model

- Bases are assumed to be equally frequent (all 0.25)
- Assumes rate of substitution (α) is the same for all possible substitutions
- Usually described as a 1-parameter model (the parameter being α)
- Remember, however, that each edge in a tree can have its own α , so there are really as many parameters in the model as there are edges in the tree!

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pages 21-132 in H. N. Munro (ed.), *Mammalian Protein Metabolism*. Academic Press, New York.

JC transition probabilities

$$\Pr(T \rightarrow A|t) = 0.25 (1 - e^{-4\theta t})$$

$$\Pr(T \rightarrow C|t) = 0.25 (1 - e^{-4\theta t})$$

$$\Pr(T \rightarrow G|t) = 0.25 (1 - e^{-4\theta t})$$

$$\Pr(T \rightarrow T|t) = 0.25 (1 - e^{-4\theta t})$$

but this only adds up to:

$$(1 - e^{-4\theta t})$$

instead of 1!

We left out the probability of no disruptions: $e^{-4\theta t}$

So:

$$\Pr(T \rightarrow A|t) = 0.25 (1 - e^{-4\theta t})$$

$$\Pr(T \rightarrow C|t) = 0.25 (1 - e^{-4\theta t})$$

$$\Pr(T \rightarrow G|t) = 0.25 (1 - e^{-4\theta t})$$

$$\begin{aligned}\Pr(T \rightarrow T|t) &= e^{-4\theta t} + 0.25 (1 - e^{-4\theta t}) \\ &= 0.25 + 0.75e^{-4\theta t}\end{aligned}$$

JC transition probabilities

$$\Pr(i \rightarrow j|t) = 0.25 (1 - e^{-4\theta t})$$

$$\Pr(i \rightarrow i|t) = 0.25 + 0.75e^{-4\theta t}$$

When $t = 0$, then $e^{-4\theta t} = 1$, and:

$$\Pr(i \rightarrow j|t) = 0$$

$$\Pr(i \rightarrow i|t) = 1$$

JC transition probabilities

$$\Pr(i \rightarrow j|t) = 0.25 (1 - e^{-4\theta t})$$

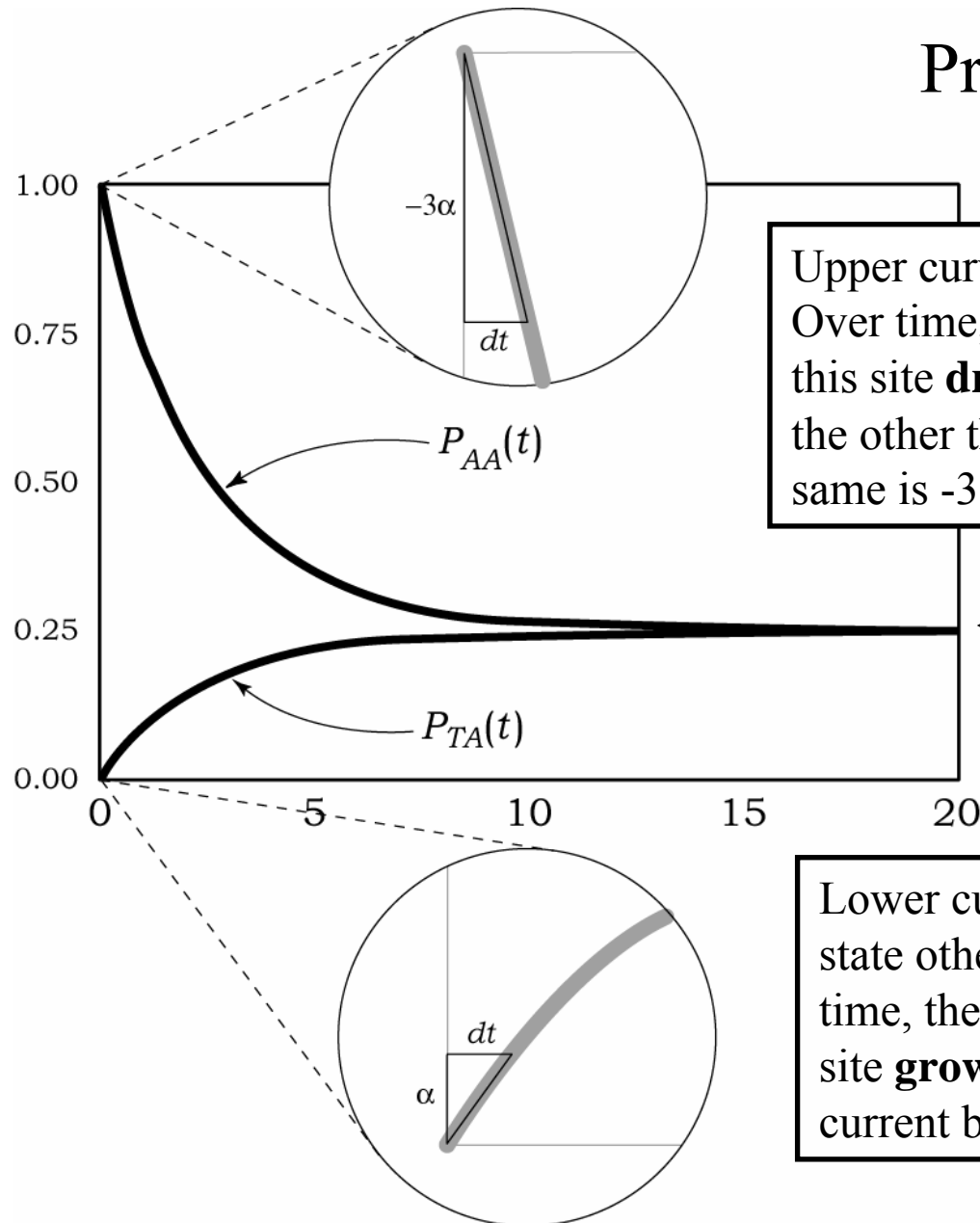
$$\Pr(i \rightarrow i|t) = 0.25 + 0.75e^{-4\theta t}$$

When $t = \infty$, then $e^{-4\theta t} = 0$, and:

$$\Pr(i \rightarrow j|t) = 0.25$$

$$\Pr(i \rightarrow i|t) = 0.25$$

Probability of “A present” as a function of time

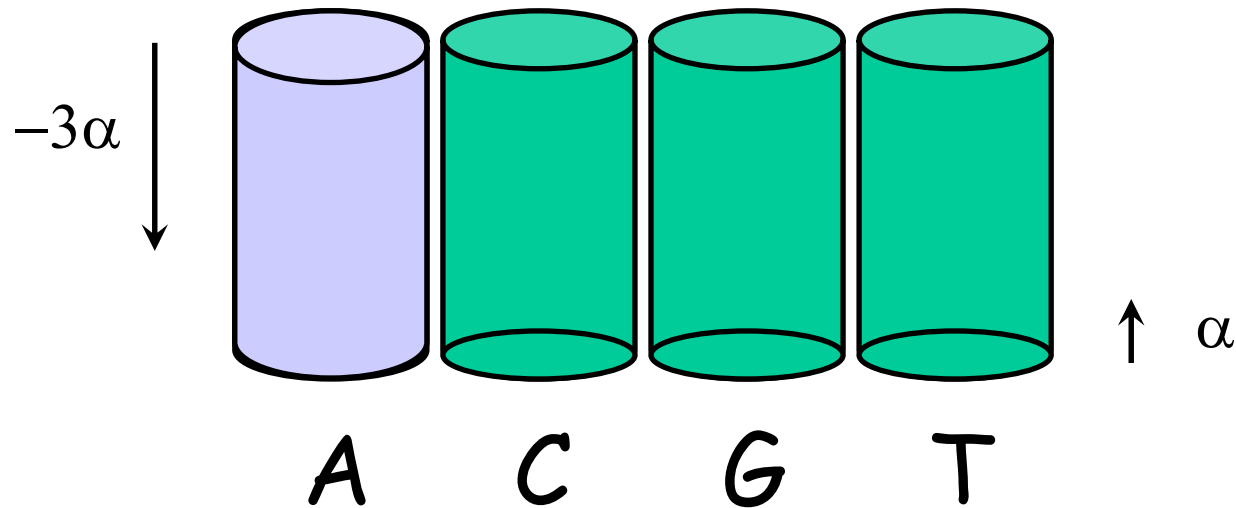


Upper curve assumes we started with A at time 0. Over time, the probability of still seeing an A at this site **drops** because rate of changing to one of the other three bases is 3α (so rate of staying the same is -3α).

The equilibrium relative frequency of A is 0.25

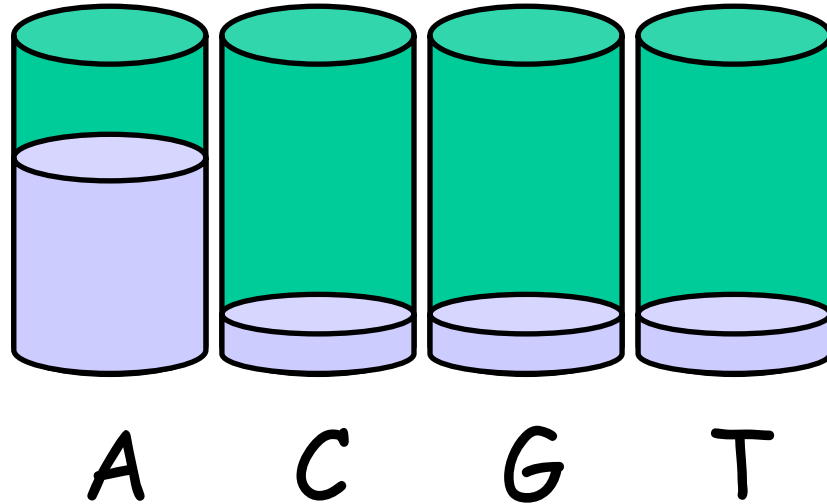
Lower curve assumes we started with some state other than A (T is used here). Over time, the probability of seeing an A at this site **grows** because the rate at which the current base will change into an A is α .

Water analogy (time 0)



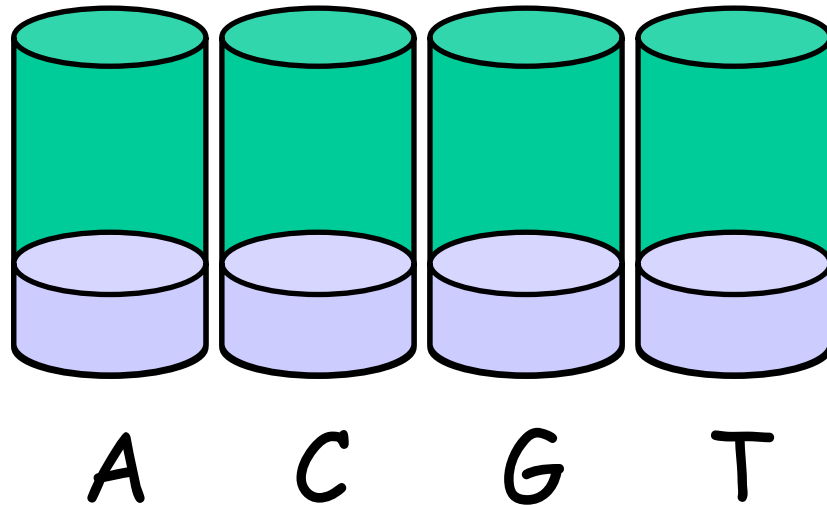
- Start with container A completely full and others empty
- Imagine that all containers are connected by tubes that allow same rate of flow between any two
- Initially, A will be losing water at 3 times the rate that C (or G or T) gains water

Water analogy (after some time)



A's level is not dropping as fast now because it is now also *receiving* water from C, G and T

Water analogy (after a very long time)



Eventually, all containers are one fourth full and there is zero *net* volume change – **stationarity** (equilibrium) has been achieved

(Thanks to Kent Holsinger for this analogy)

JC instantaneous rate matrix - the Q matrix for JC

The 1 parameter is α (sometimes parameterized in terms of μ). This is the rate of replacements (“disruptions” that change the state):

		To State			
		A	C	G	T
From State	A	-3α	α	α	α
	C	α	-3α	α	α
	G	α	α	-3α	α
	T	α	α	α	-3α

Change probabilities

We can calculate a transition probability matrix as a function of time by:

$$\mathbf{P}(t) = e^{\mathbf{Q}t}$$

The important thing to note is the rates (\mathbf{Q} matrix) is multiplied by the time.

We can't separate rates and times since we always see the effect of their product.

Is a medium level of character divergence:

1. medium rate of change and medium amount of time,
2. high rate, but short time period,
3. low rate, but a long time period?

JC instantaneous rate matrix again

What if you do not know the length of time for a branch in the tree? We estimate branch lengths in terms of character divergence – the product of rate and time. What is important is that we know the relative rates of different types of substitutions, so JC can be expressed:

		To State			
		A	C	G	T
From State	A	-3	1	1	1
	C	1	-3	1	1
	G	1	1	-3	1
	T	1	1	1	-3

JC instantaneous rate matrix yet again

We estimate branch lengths in terms of expected number of changes *per site*. To do this we standardize the total rate of divergence in the Q matrix and estimate $\nu = \mu t = 3\alpha t$ for each branch.

		To State			
		A	C	G	T
From State	A	-1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
	C	$\frac{1}{3}$	-1	$\frac{1}{3}$	$\frac{1}{3}$
	G	$\frac{1}{3}$	$\frac{1}{3}$	-1	$\frac{1}{3}$
	T	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	-1

Kimura (1980) model or “the K80 model”

Transitions and transversions occur at different rates:

		To State			
		A	C	G	T
From State	A	$-2\beta - \alpha$	β	α	β
	C	β	$-2\beta - \alpha$	β	α
	G	α	β	$-2\beta - \alpha$	β
	T	β	α	β	$-2\beta - \alpha$

Kimura (1980) model or “the K80 model”.
Reparameterized.

Once again, we care only about the relative rates, so we can choose one rate to be frame of reference. This turns the 2 parameter model into a 1 parameter form:

		To State			
		A	C	G	T
From State	A	$-(2 + \kappa)\beta$	β	$\kappa\beta$	β
	C	β	$-(2 + \kappa)\beta$	β	$\kappa\beta$
	G	$\kappa\beta$	β	$-(2 + \kappa)\beta$	β
	T	β	$\kappa\beta$	β	$-(2 + \kappa)\beta$

Kimura (1980) model or “the K80 model”.
Reparameterized again.

		To State			
		A	C	G	T
From State	A	$-2 - \kappa$	1	κ	1
	C	1	$-2 - \kappa$	1	κ
	G	κ	1	$-2 - \kappa$	1
	T	1	κ	1	$-2 - \kappa$

Kappa is the transition/transversion rate ratio:

$$\kappa = \frac{\alpha}{\beta}$$

(if $\kappa = 1$ then we are back to JC).

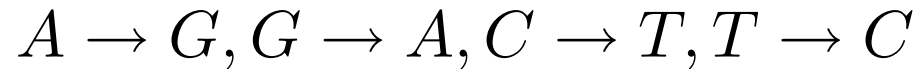
What is the instantaneous probability of an particular transversion?

$$\begin{aligned}\Pr(A \rightarrow C) &= \Pr(A) \Pr(\text{change to } C) \\ &= \frac{1}{4} (\beta dt)\end{aligned}$$

What is the instantaneous probability of an particular transition?

$$\begin{aligned}\Pr(A \rightarrow G) &= \Pr(A) \Pr(\text{change to } G) \\ &= \frac{1}{4} (\kappa \beta dt)\end{aligned}$$

There are four types of transitions:



and eight types of transversions:



$$\text{Ti/Tv ratio} = \frac{\text{Pr}(\text{any transition})}{\text{Pr}(\text{any transversion})} = \frac{4 \left(\frac{1}{4} (\kappa \beta dt) \right)}{8 \left(\frac{1}{4} (\beta dt) \right)} = \frac{\kappa}{2}$$

For K2P instantaneous transition/transversion ratio is one-half the instantaneous transition/transversion **rate ratio**