Maximum Likelihood
Phylogenetic Biology - Week 3

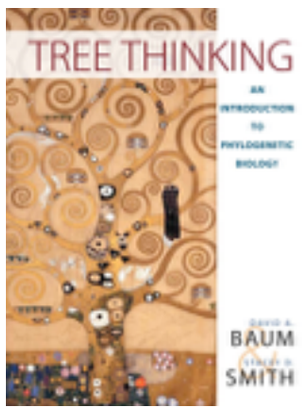
Biology 1425

Professor: Casey Dunn, dunnlab.org
Brown University

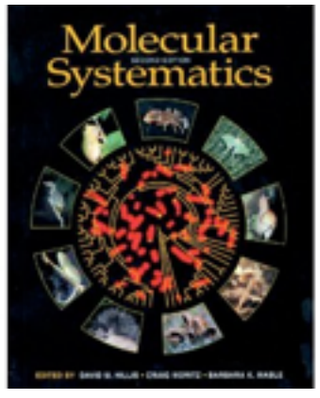
Front matter...

Sources

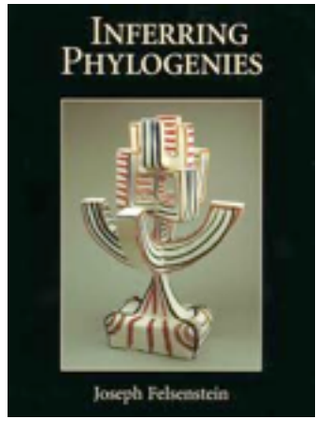
Some non-original content is drawn from:



Baum, D and S. Smith (2012) Tree Thinking: and Introduction to Phylogenetic Biology. Roberts and Company Publishers. ISBN 9781936221165



Swofford, D. L., Olsen, G. J., Waddell, P. J., & Hillis, D. M. (1996). Phylogenetic inference. In: Molecular Systematics, Second Edition. eds: D. M. Hillis, C Moritz, & B. K. Mable. Sinauer Associates. ISBN 9780878932825



Felsenstein, J. (2003) Inferring Phylogenies. Sinauer Associates. ISBN 978-0878931774

Other non-original content is referenced by url.

Probability density function

The probability of the hypothesis (H) given the data (D):

$$P(H|D)$$

Non-negative everywhere; area under the curve sums to one.

Likelihood

Likelihood is the probability of the data (D) given a hypothesis (H):

$$P(D|H)$$

In our case, the data is our aligned matrix (homologous characters and their observed states) and the hypothesis is a particular tree and model of character evolution.

Maximum Likelihood

$$P(D|H)$$

The hypothesis with the highest likelihood given the data and model.

Likelihood

For complex data and hypotheses, like those encountered in phylogenetics, the likelihood for any given hypothesis (even the most likely one) is very, very small.

The likelihood function is not a probability distribution function - the area need not sum to one.

Likelihood

To calculate likelihood, we need:

- Data (eg, character matrix)
- A model of evolution
- Hypothesis (eg, tree and model parameters)
- A mechanism to calculate the likelihood

given the above

Data

A matrix with:

- rows for taxa
- columns for characters
- cells are character states for each character for each taxon

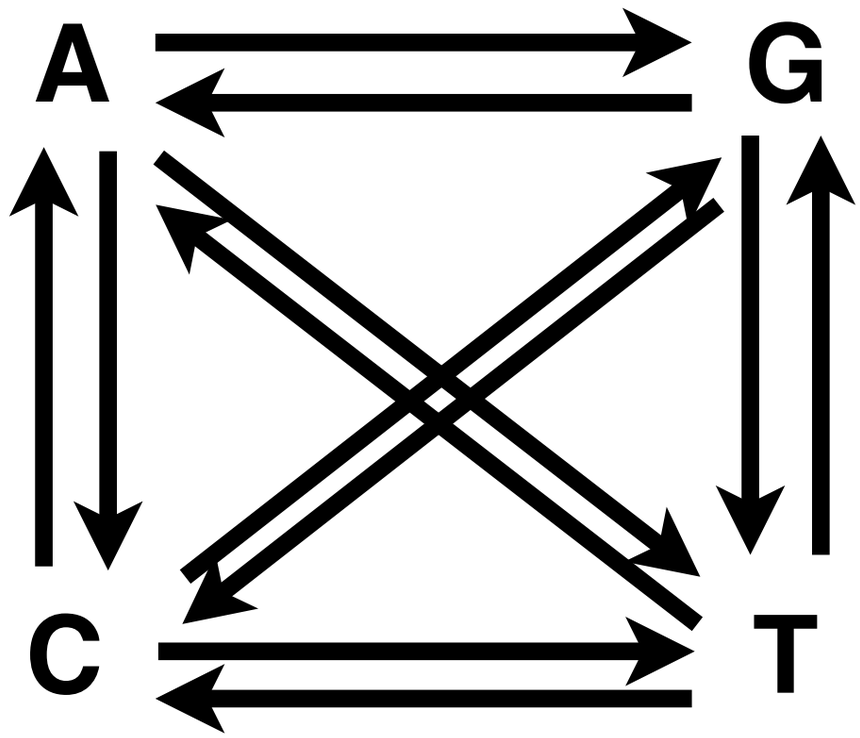
Model

Rate matrix

The instantaneous rate of given substitutions

Q - Rate matrix

Rate of change from one character state to another in infinitesimal time (ie, along a very short branch)



$$Q = \begin{array}{c} \\ \text{From:} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} \end{array} \begin{array}{c} \begin{matrix} \text{To:} \\ \text{A} \end{matrix} \\ \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) \\ \mu g \pi_A \\ \mu h \pi_A \\ \mu i \pi_A \end{pmatrix} \end{array} \begin{array}{c} \\ \text{C} \\ \begin{pmatrix} \mu a \pi_C \\ -\mu(g\pi_A + d\pi_G + e\pi_T) \\ \mu j \pi_C \\ \mu k \pi_C \end{pmatrix} \end{array} \begin{array}{c} \\ \text{G} \\ \begin{pmatrix} \mu b \pi_G \\ \mu d \pi_G \\ -\mu(h\pi_A + j\pi_C + f\pi_T) \\ \mu l \pi_G \end{pmatrix} \end{array} \begin{array}{c} \\ \text{T} \\ \begin{pmatrix} \mu c \pi_T \\ \mu e \pi_T \\ \mu f \pi_T \\ -\mu(i\pi_A + k\pi_C + l\pi_G) \end{pmatrix} \end{array}$$

(SOWH 1996)

$$Q = \begin{array}{c} \\ \text{From:} \\ \begin{matrix} \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{matrix} \end{array} \begin{array}{c} \begin{matrix} \text{A} & \text{C} & \text{G} & \text{T} \end{matrix} \\ \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu g\pi_A & -\mu(g\pi_A + d\pi_G + e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu h\pi_A & \mu j\pi_C & -\mu(h\pi_A + j\pi_C + f\pi_T) & \mu f\pi_T \\ \mu i\pi_A & \mu k\pi_C & \mu l\pi_G & -\mu(i\pi_A + k\pi_C + l\pi_G) \end{pmatrix} \end{array}$$

μ	Mean instantaneous rate
$a, b, \ldots l$	Relative rates
$\pi_A, \pi_C, \pi_G, \pi_T$	Equilibrium frequencies of states

$$Q = \begin{array}{c} \\ \text{From:} \\ \begin{matrix} \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{matrix} \end{array} \begin{array}{c} \begin{matrix} \text{A} & \text{C} & \text{G} & \text{T} \end{matrix} \\ \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu g\pi_A & -\mu(g\pi_A + d\pi_G + e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu h\pi_A & \mu j\pi_C & -\mu(h\pi_A + j\pi_C + f\pi_T) & \mu f\pi_T \\ \mu i\pi_A & \mu k\pi_C & \mu l\pi_G & -\mu(i\pi_A + k\pi_C + l\pi_G) \end{pmatrix} \end{array}$$

μ Mean instantaneous rate

$a, b, \ldots l$ Relative rates

$\pi_A, \pi_C, \pi_G, \pi_T$ Equilibrium frequencies of states

Rows sum to 0. This is because the A's that exist before the change are reduced by the number of A's that become C, G, and T.

$$Q = \begin{array}{c} \\ \text{From:} \\ \begin{matrix} \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{matrix} \end{array} \begin{array}{c} \begin{matrix} \text{A} & \text{C} & \text{G} & \text{T} \end{matrix} \\ \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu g\pi_A & -\mu(g\pi_A + d\pi_G + e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu h\pi_A & \mu j\pi_C & -\mu(h\pi_A + j\pi_C + f\pi_T) & \mu f\pi_T \\ \mu i\pi_A & \mu k\pi_C & \mu l\pi_G & -\mu(i\pi_A + k\pi_C + l\pi_G) \end{pmatrix} \end{array}$$

μ	Mean instantaneous rate
$a, b, \ldots l$	Relative rates
$\pi_A, \pi_C, \pi_G, \pi_T$	Equilibrium frequencies of states

Assumption: that the state frequencies approach equilibrium.

“The rate of change to each base is proportional to the equilibrium frequency” - (SOWH 1996)

Given enough time, the sequence will evolve to the equilibrium frequencies even if it doesn’t start at the,.

$$Q = \begin{array}{c} \\ \text{From:} \\ \end{array} \begin{array}{c} \begin{array}{cccc} \text{A} & \text{C} & \text{G} & \text{T} \end{array} \\ \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu g\pi_A & -\mu(g\pi_A + d\pi_G + e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu h\pi_A & \mu j\pi_C & -\mu(h\pi_A + j\pi_C + f\pi_T) & \mu f\pi_T \\ \mu i\pi_A & \mu k\pi_C & \mu l\pi_G & -\mu(i\pi_A + k\pi_C + l\pi_G) \end{pmatrix} \end{array} \begin{array}{c} \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{array}$$

$$= R\Pi$$

Where:

$$\mathbf{R} = \begin{pmatrix} — & \mu a & \mu b & \mu c \\ \mu g & — & \mu d & \mu e \\ \mu h & \mu j & — & \mu f \\ \mu i & \mu k & \mu l & — \end{pmatrix} \qquad \mathbf{\Pi} = \begin{pmatrix} \pi_A & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_G & 0 \\ 0 & 0 & 0 & \pi_T \end{pmatrix}$$

And — is selected for rows of **Q** to sum to 0

$$\mathbf{R} = \begin{pmatrix} — & \mu a & \mu b & \mu c \\ \mu g & — & \mu d & \mu e \\ \mu h & \mu j & — & \mu f \\ \mu i & \mu k & \mu l & — \end{pmatrix} \qquad \mathbf{\Pi} = \begin{pmatrix} \pi_A & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_G & 0 \\ 0 & 0 & 0 & \pi_T \end{pmatrix}$$

New assumption: time reversibility, e.g. the relative rate of change from A to C is the same as C to A.

$$g = a, h = b, i = c, j = d, k = e, l = f$$

Gives:

$$\mathbf{R} = \begin{pmatrix} — & \mu a & \mu b & \mu c \\ \mu a & — & \mu d & \mu e \\ \mu b & \mu d & — & \mu f \\ \mu c & \mu e & \mu f & — \end{pmatrix}$$

This model is simpler - it has only 6 relative rate parameters.

This model is simpler - it has only 6 relative rate parameters.

$$\mathbf{R} = \begin{pmatrix} — & \mu a & \mu b & \mu c \\ \mu a & — & \mu d & \mu e \\ \mu b & \mu d & — & \mu f \\ \mu c & \mu e & \mu f & — \end{pmatrix}$$

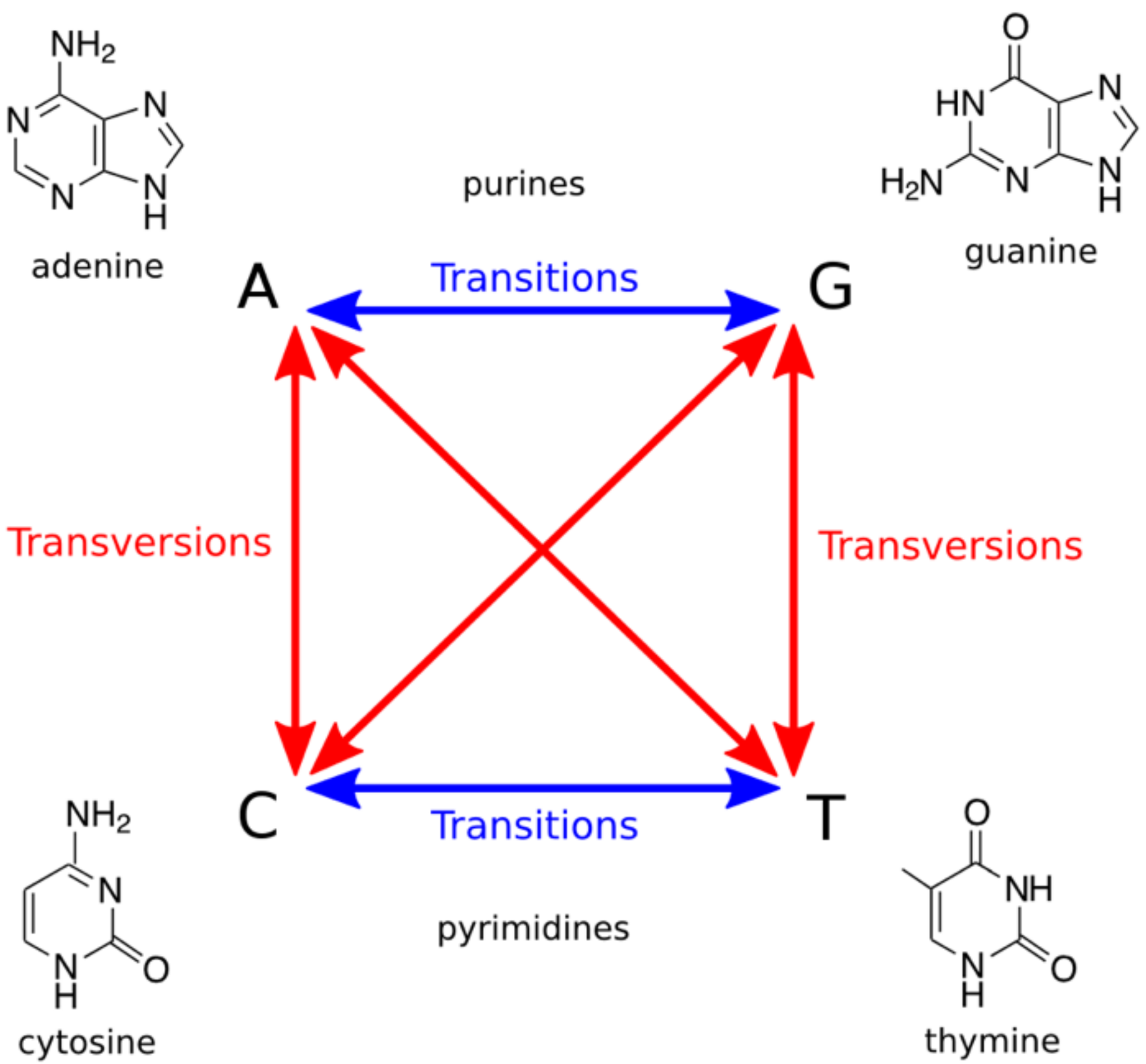
$$\mathbf{\Pi} = \begin{pmatrix} \pi_A & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_G & 0 \\ 0 & 0 & 0 & \pi_T \end{pmatrix}$$

$$R\Pi = \boldsymbol{Q} = \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu a\pi_A & -\mu(a\pi_A + d\pi_G + e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu b\pi_A & \mu d\pi_C & -\mu(b\pi_A + d\pi_C + f\pi_T) & \mu f\pi_T \\ \mu c\pi_A & \mu e\pi_C & \mu f\pi_G & -\mu(c\pi_A + e\pi_C + f\pi_G) \end{pmatrix}$$

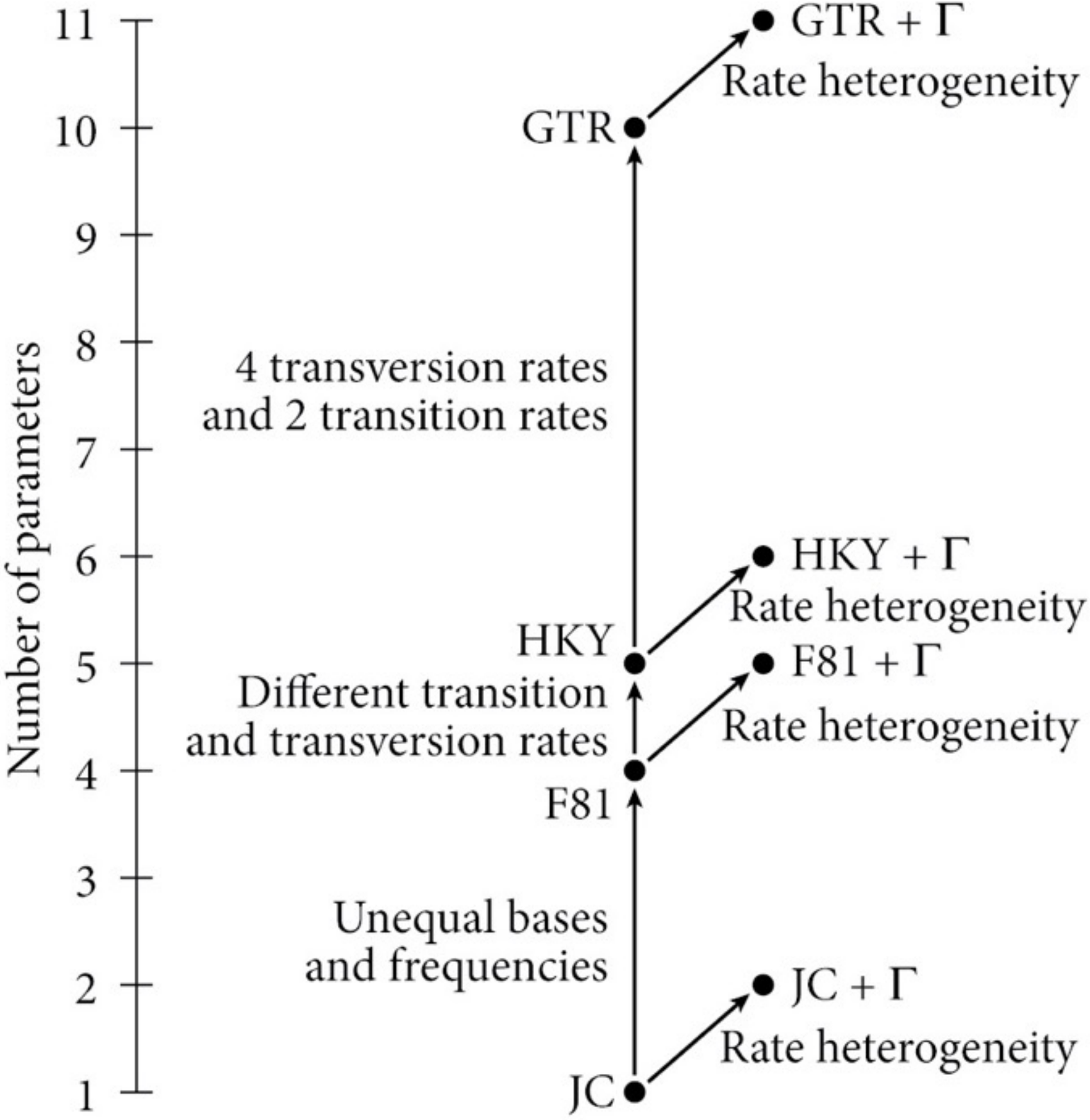
This is the General Time Reversible (**GTR**) model.

Can further simplify GTR by reducing the number of parameters:

- Can set some relative rates equal to each other
- Can set some frequencies equal to each other



http://upload.wikimedia.org/wikipedia/commons/3/35/Transitions-transversions-v3.png



Baum and Smith 2012, Figure 8.10

F81 model

Q - Rate matrix

From: \ To:	A (freq = π_A)	C (freq = π_C)	G (freq = π_G)	T (freq = π_T)
A (freq = π_A)	$-m(\pi_C + \pi_G + \pi_T)$	$\pi_C m$	$\pi_G m$	$\pi_T m$
C (freq = π_C)	$\pi_A m$	$-m(\pi_A + \pi_G + \pi_T)$	$\pi_G m$	$\pi_T m$
G (freq = π_G)	$\pi_A m$	$\pi_C m$	$-m(\pi_A + \pi_C + \pi_T)$	$\pi_T m$
T (freq = π_T)	$\pi_A m$	$\pi_C m$	$\pi_G m$	$-m(\pi_A + \pi_C + \pi_G)$

Baum and Smith 2012, Figures 8.7, 8.8

Can increase complexity by adding to the number of parameters:

- Allow rates to differ between sites (eg, gamma), see Lewis slides “Rate Heterogeneity”
- Allow equilibrium frequencies to vary between sites (eg, CAT)

Hypothesis

For now, we will treat the hypothesis as fixed. The challenge at hand his to calculate the likelihood of a given tree and set of model parameters given the data.

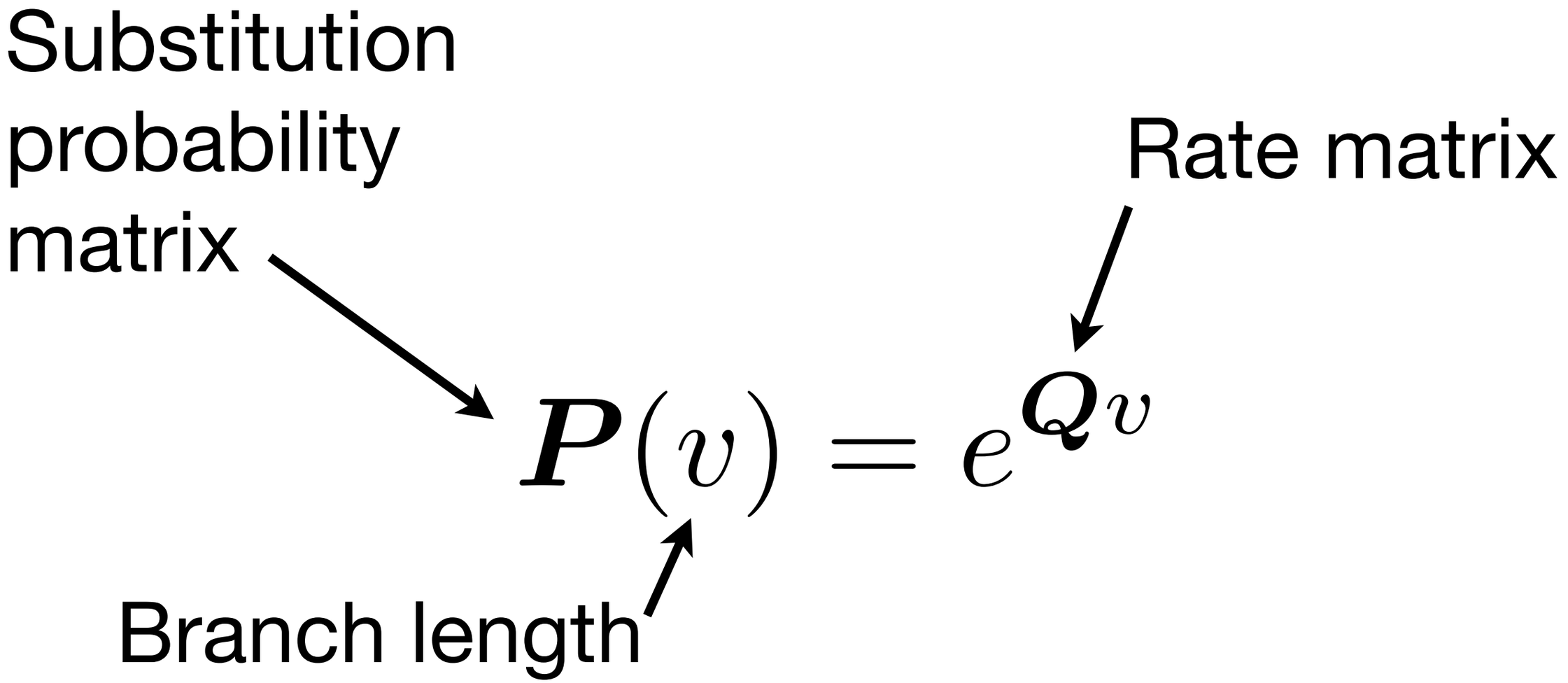
Calculating likelihood

Substitution probability matrix

The probability of a given substitution occurring in a given interval (branch length) t. Because of reversals, there are an infinite number of histories that could have given rise to the particular substitution. Can be derived from the rate matrix.

$\boldsymbol{P}$ - Substitution probability matrix

Substitution probability matrix



This is called matrix exponentiation

F81 model

Q - Rate matrix

		To:			
		A (freq = π_A)	C (freq = π_C)	G (freq = π_G)	T (freq = π_T)
From:	A (freq = π_A)	$-m(\pi_C + \pi_G + \pi_T)$	$\pi_C m$	$\pi_G m$	$\pi_T m$
	C (freq = π_C)	$\pi_A m$	$-m(\pi_A + \pi_G + \pi_T)$	$\pi_G m$	$\pi_T m$
	G (freq = π_G)	$\pi_A m$	$\pi_C m$	$-m(\pi_A + \pi_C + \pi_T)$	$\pi_T m$
	T (freq = π_T)	$\pi_A m$	$\pi_C m$	$\pi_G m$	$-m(\pi_A + \pi_C + \pi_G)$

P - Substitution probability matrix

		To:			
		A	C	G	T
From:	A	$\pi_A + (1 - \pi_A)e^{-mt}$	$\pi_C(1 - e^{-mt})$	$\pi_G(1 - e^{-mt})$	$\pi_T(1 - e^{-mt})$
	C	$\pi_A(1 - e^{-mt})$	$\pi_C + (1 - \pi_C)e^{-mt}$	$\pi_G(1 - e^{-mt})$	$\pi_T(1 - e^{-mt})$
	G	$\pi_A(1 - e^{-mt})$	$\pi_C(1 - e^{-mt})$	$\pi_G + (1 - \pi_G)e^{-mt}$	$\pi_T(1 - e^{-mt})$
	T	$\pi_A(1 - e^{-mt})$	$\pi_C(1 - e^{-mt})$	$\pi_G(1 - e^{-mt})$	$\pi_T + (1 - \pi_T)e^{-mt}$

Baum and Smith 2012, Figures 8.7, 8.8

F81 model

As the branch length goes to 0, **P** becomes a diagonal matrix

	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

As the branch length goes to infinity, the rows become the equilibrium base frequencies

$\boldsymbol{P}$

		To: A	To: C	To: G	To: T
From:	A	$\pi_A + (1-\pi_A)e^{-mt}$	$\pi_C(1-e^{-mt})$	$\pi_G(1-e^{-mt})$	$\pi_T(1-e^{-mt})$
	C	$\pi_A(1-e^{-mt})$	$\pi_C + (1-\pi_C)e^{-mt}$	$\pi_G(1-e^{-mt})$	$\pi_T(1-e^{-mt})$
	G	$\pi_A(1-e^{-mt})$	$\pi_C(1-e^{-mt})$	$\pi_G + (1-\pi_G)e^{-mt}$	$\pi_T(1-e^{-mt})$
	T	$\pi_A(1-e^{-mt})$	$\pi_C(1-e^{-mt})$	$\pi_G(1-e^{-mt})$	$\pi_T + (1-\pi_T)e^{-mt}$

Baum and Smith 2012, Figure 8.8

Calculating likelihood

The data:

	1							j						N
(1)	C	...	G	G	A	C	A	C	G	T	T	T	A	... C
(2)	C	...	A	G	A	C	A	C	C	T	C	T	A	... C
(3)	C	...	G	G	A	T	A	A	G	T	T	A	A	... C
(4)	C	...	G	G	A	T	A	G	C	C	T	A	G	... C

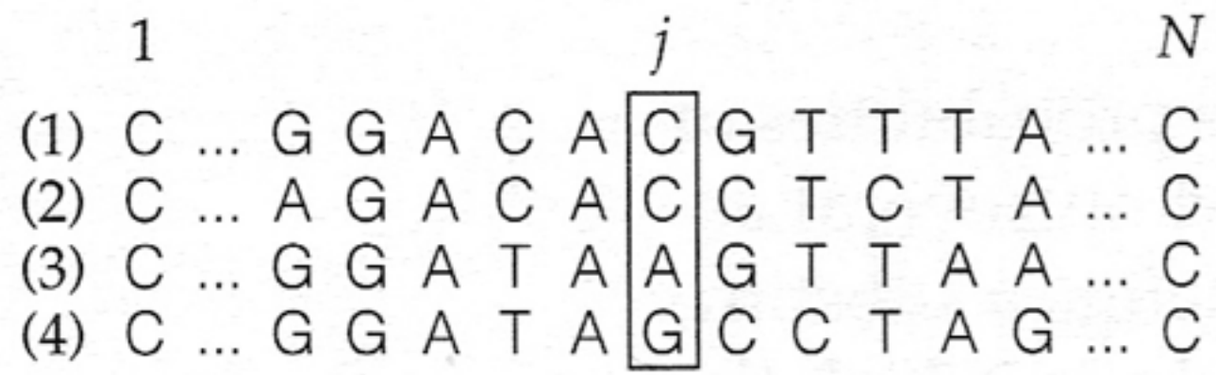
Need to:

- Calculate the probability of observing the data in each column given the tree and model (sum over all alternative histories congruent with tree)
- Calculate the probability of observing the entire matrix (multiply probabilities across columns)

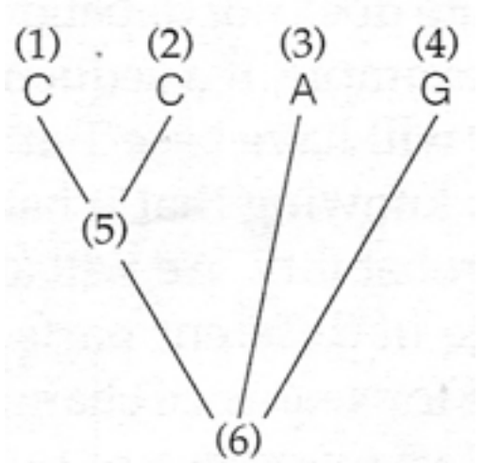
See Lewis slides “Combining probabilities”

Calculating likelihood

The data:



1 of 3 possible trees:



Likelihood for site j:

$$L_{(j)} = \text{Prob}\left(\begin{array}{cccc} C & C & A & G \\ & A & & \\ & & A & \end{array}\right) + \text{Prob}\left(\begin{array}{cccc} C & C & A & G \\ & C & & \\ & & A & \end{array}\right)$$

$$+ \quad \ldots \quad + \text{Prob}\left(\begin{array}{cccc} C & C & A & G \\ & G & & \\ & & C & \end{array}\right)$$

$$+ \quad \ldots \quad + \text{Prob}\left(\begin{array}{cccc} C & C & A & G \\ & T & & \\ & & T & \end{array}\right)$$

Swofford et al 1996, Figure 10

Calculating likelihood

The data:

	1														N
								j							
(1)	C	...	G	G	A	C	A	C	G	T	T	T	A	...	C
(2)	C	...	A	G	A	C	A	C	C	T	C	T	A	...	C
(3)	C	...	G	G	A	T	A	A	G	T	T	A	A	...	C
(4)	C	...	G	G	A	T	A	G	C	C	T	A	G	...	C

Likelihood of all sites:

$$L = L_{(1)} \bullet L_{(2)} \bullet \ldots \bullet L_{(N)} = \prod_{j=1}^{N} L_{(j)}$$

Log likelihood of all sites:

$$\ln L = \ln L_{(1)} + \ln L_{(2)} + \ldots + \ln L_{(N)} = \sum_{j=1}^{N} \ln L_{(j)}$$

Swofford et al 1996, Figure 10

Maximum likelihood

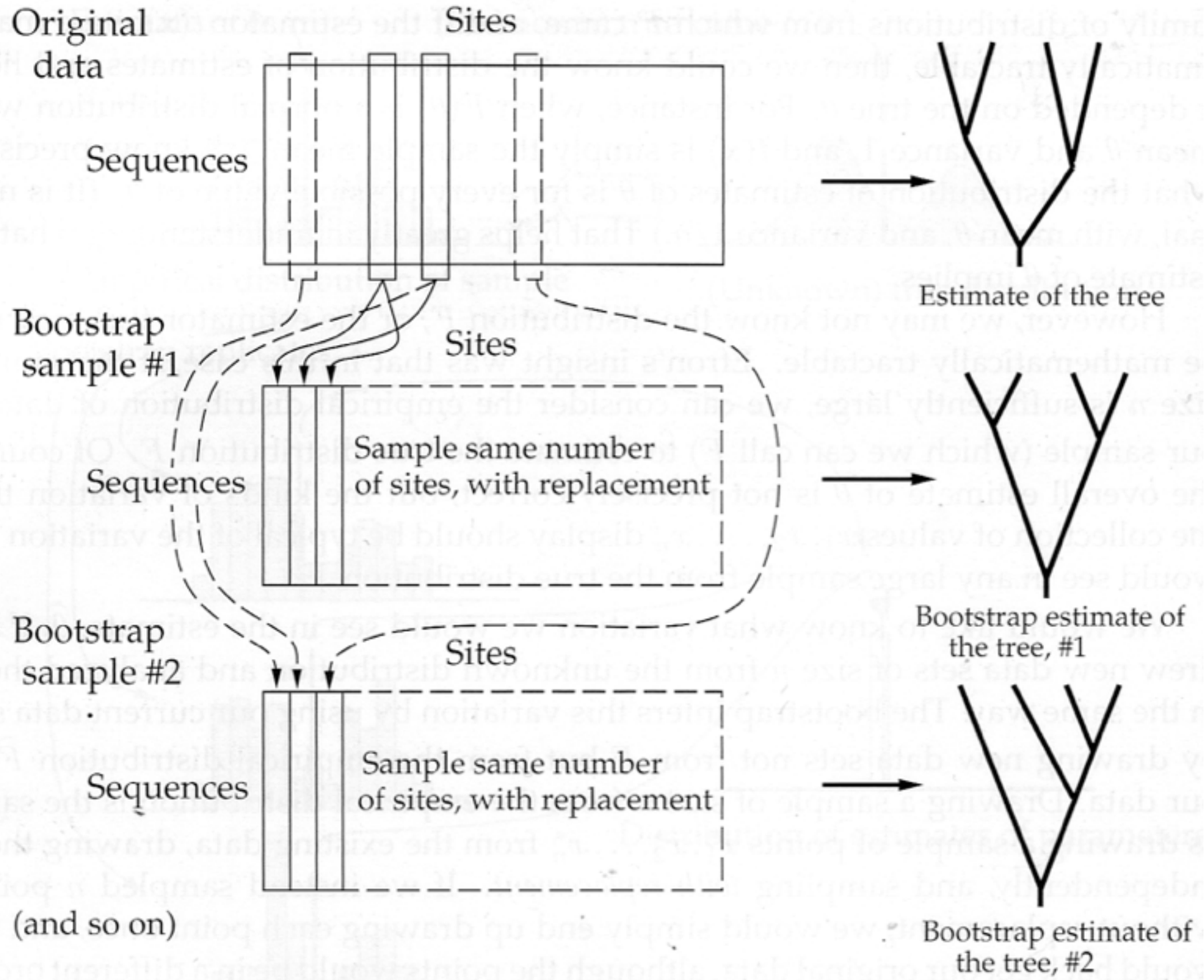
Propose many trees.

For each tree optimize branch lengths and model parameters.

Pick the tree with the highest likelihood.

Felsenstein 2003, Figure 20.2

Bootstraps



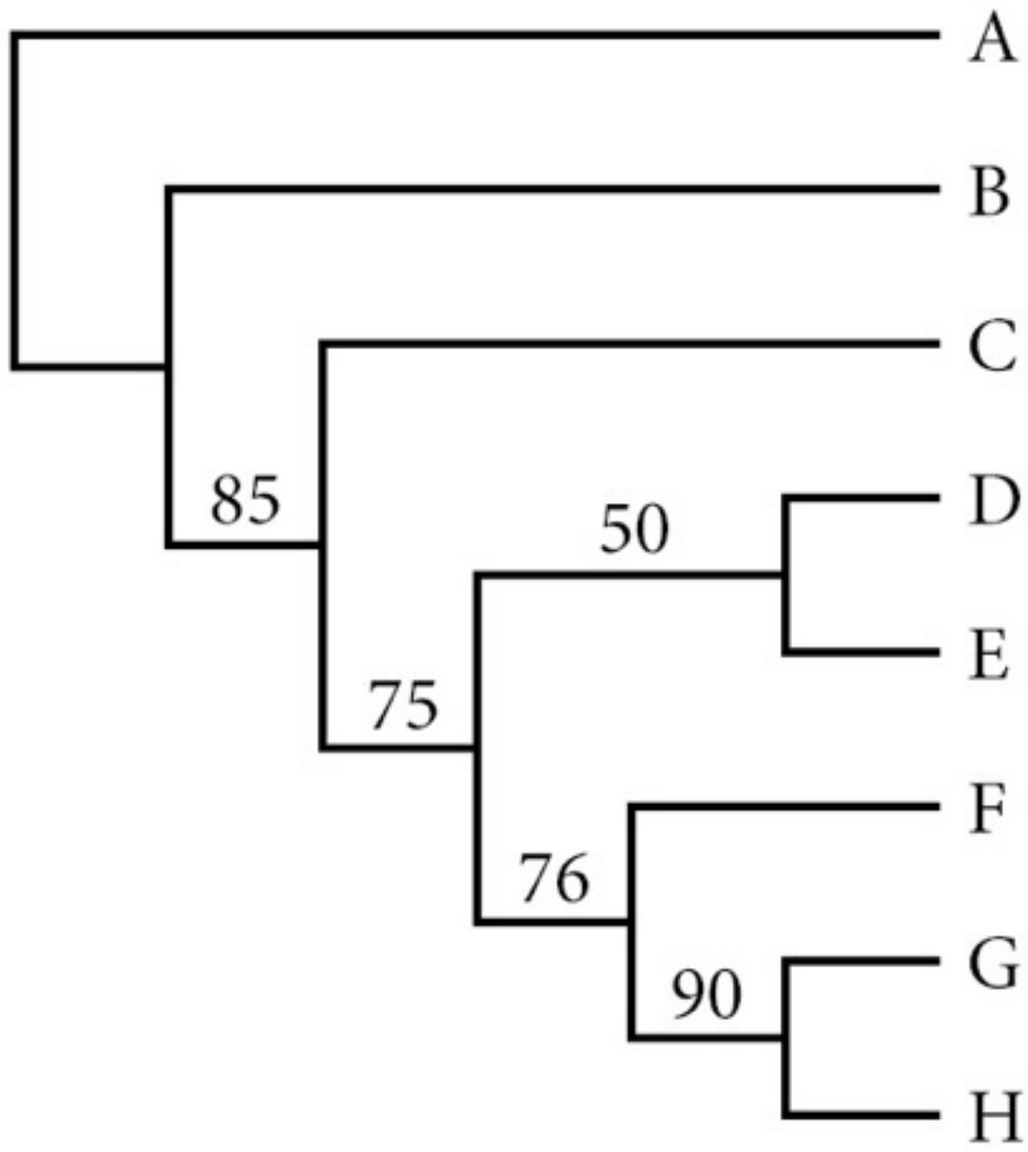
Felsenstein 2003, Figure 20.2

Calculate frequency of bipartitions across bootstrap trees.

Summarize these:

- Map them to the ML tree
- Calculate a consensus tree and map the frequencies to the consensus
- etc…

Consensus tree



The number indicates the percentage of bootstrap trees that include the corresponding edge (ie, branch).

Baum and Smith 2012, Figure 9.8