### JC69 Distance Formula

The JC69 distance formula is based on the Jukes-Cantor model, which assumes equal evolutionary rates for all four nucleotide substitutions.

### JC69 rate matrix

<table>
<thead>
<tr>
<th>“From” state</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“To” state</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
</tr>
</tbody>
</table>

### K80 (K2P) rate matrix

The K80 model is more accurate than JC69 because it accounts for different substitution probabilities.

### “Transition/transversion ratio” vs. “transition/transversion rate ratio”

Possible transitions: Possible transversions:

$$\frac{E[\text{No. transitions}]}{E[\text{No. transversions}]} = \frac{E[\text{rate of transitions}]}{E[\text{rate of transversions}]}$$

### Equivalence to JC69

The transition/transversion rate ratio is equivalent to the JC69 model when the transition rate and the transversion rate are equal.
F81 rate matrix


\[
\begin{array}{cccc}
A & C & G & T \\
A & & & \\
C & & & \\
G & & & \\
T & & & \\
\end{array}
\]

no. parameters: 22

equivalence to JC69

HKY85 rate matrix


\[
\begin{array}{cccc}
A & C & G & T \\
A & & & \\
C & & & \\
G & & & \\
T & & & \\
\end{array}
\]

no. parameters: 23

equivalence to JC69, F81

F84 vs. HKY85

F84 model:

\[\mu\] rate of process generating all types of substitutions
\[k\mu\] rate of process generating only transitions

Becomes F81 model if \(k = 0\)

HKY85 model:

\[\beta\] rate of process generating only transversions
\[k\beta\] rate of process generating only transitions

Becomes F81 model if \(k = 1\)

GTR rate matrix


\[
\begin{array}{cccc}
A & C & G & T \\
A & & & \\
C & & & \\
G & & & \\
T & & & \\
\end{array}
\]

no. parameters: 24

equivalence to JC69, F81, K80, HKY 25
AND means MULTIPLY

Using 2 dice, what is the probability of

○ AND ● ?

AND rule in phylogenetics

OR means ADD

Using one die, what is the probability of

○ OR ● ?

Combining AND and OR

What is the probability that the sum of two dice is 7?
Using both AND and OR in phylogenetics

Dependence Example

A = my normal route

Pr(route = A | normal) = 0.95
Pr(route = B | normal) = 0.05

B = I go out of my way

Dependence vs Independence

B = I go out of my way

Pr(route = A | accident) = 0.05
Pr(route = B | accident) = 0.95
Likelihood of a single sequence

First 32 nucleotides of the ψη-globin gene of gorilla:

\[ \text{GAAGTCCTTGAGAAATAAACTGCACACACTGG} \]

\[ L = \]

\[ \log L = \]

Likelihood ratio test

Find *maximum* logL under F81 (unconstrained) model:

Find *maximum* logL under JC69 (constrained) model:

-43.1: 375, 219, 219, 187
-44.4: 25, 25, 25, 25

Likelihood ratio test

Calculate the likelihood ratio test statistic:

Calculate the degrees of freedom:

Likelihood of the simplest tree

First 32 nucleotides of the ψη-globin gene of gorilla and orangutan:

gorilla  \text{GAAGTCCTTGAGAAATAAACTGCACACACTGG}
orangutan  \text{GAACTCCTTGAGAAATAAACTGCACACACTGG}

first site: G \leftrightarrow G
second site: A \leftrightarrow G
Maximum likelihood estimation

First 32 nucleotides of the \( \psi \eta \)-globin gene of gorilla and orangutan:

<table>
<thead>
<tr>
<th>Gorilla</th>
<th>GAAGTCCTGGAGAAATAAACAGCACACTGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orangutan</td>
<td>GAAGTCCTGGAGAAATAAACAGCACACTGG</td>
</tr>
</tbody>
</table>

\[
L = \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} + \frac{3}{4} e^{-4\nu/3} \right) \right]^{30} \left[ \left( \frac{1}{4} \right) \left( \frac{1}{4} - \frac{1}{4} e^{-4\nu/3} \right) \right]^{12}
\]