21.3.1 Jukes-Cantor model

The simplest model is the *Jukes-Cantor* model (1969) which has equal rates of mutation between all bases so that $q_{ij} = 1$ for $i \neq j$,

$$Q = \beta \begin{bmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{bmatrix}.$$

In this case, $\beta = 1/3$ so

$$Q = \begin{bmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/3 & -1 & 1/3 & 1/3 \\ 1/3 & 1/3 & -1 & 1/3 \\ 1/3 & 1/3 & 1/3 & -1 \end{bmatrix}.$$

The equilibrium of this process is $\pi = (1/4, 1/4, 1/4, 1/4)$.

The transition matrix $P(t) = \exp(Qt)$ for the Jukes-Cantor model has off-diagonal entries

$$P_{ij}(t) = \frac{1}{4} - \frac{1}{4}\exp(-t\mu)$$

and diagonal entries

$$P_{ii}(t) = \frac{1}{4} + \frac{3}{4}\exp(-t\mu).$$

21.3.2 Kimura model

The Kimura model (1980) distinguishes between *transitions* ($A \longleftrightarrow G$ and $C \longleftrightarrow T$ state changes) and *transversions* (state changes from a purine to pyrimidine or *vice versa*). The model assumes base frequencies are equal for all characters. This transition/transversion bias is governed by the κ parameter and the Q matrix is:

$$Q = \beta \begin{bmatrix} -2-\kappa & 1 & \kappa & 1 \\ 1 & -2-\kappa & 1 & \kappa \\ \kappa & 1 & -2-\kappa & 1 \\ 1 & \kappa & 1 & -2-\kappa \end{bmatrix},$$

The normalized Q is obtained by setting $\beta = \frac{1}{2+\kappa}$. This model has one free parameter, κ. The transition probabilities are:

$$p_{ij}(d) = \begin{cases} \frac{1}{4} + \frac{1}{4}\exp(-\frac{4}{\kappa+2}d) + \frac{1}{2}\exp(-\frac{2\kappa+2}{\kappa+2}d) & \text{if } i = j \\ \frac{1}{4} + \frac{1}{4}\exp(-\frac{4}{\kappa+2}d) - \frac{1}{2}\exp(-\frac{2\kappa+2}{\kappa+2}d) & \text{if transition} \\ \frac{1}{4} - \frac{1}{4}\exp(-\frac{4}{\kappa+2}d) & \text{if transversion} \end{cases}.$$

21.3.3 F81 and HKY models

In 1981, Joe Felsenstein proposed a model that extends the Jukes-Cantor model to allow for unequal equilibrium base frequencies, that is π for which $\pi_a \neq \pi_b$. This is known as the F81 model. The F81 model has 3 parameters, one less than the number of equilibrium base frequencies since there is the restriction that $\sum_i \pi_i = 1$.

In 1985, the F81 model was extended to incorporate the Kimura model, so allows different rates for transitions and transversions as well as unequal base frequencies. The resulting model is known as the HKY model and has rate matrix of the form:

$$Q = \beta \begin{bmatrix} \cdot & \pi_C & \kappa\pi_G & \pi_T \\ \pi_A & \cdot & \pi_G & \kappa\pi_T \\ \kappa\pi_A & \pi_C & \cdot & \pi_T \\ \pi_A & \kappa\pi_C & \pi_G & \cdot \end{bmatrix}$$

where the diagonal elements are defined in the usual way so that the row sums are zero. The transition matrix P can be calculated analytically for this model but it is omitted here.

21.3.4 GTR model

In 1986, the most general reversible model was developed which can have an arbitrary stationary distribution, and given the restriction of reversibility, 6 parameters for adjusting the rates of mutation between bases. The rate matrix is

$$Q = \beta \begin{bmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{bmatrix}.$$

The diagonal elements are calculated in the normal way.

Where the normalization is $\beta = 1/[2(a\pi_A\pi_C + b\pi_A\pi_G + c\pi_A\pi_T + d\pi_C\pi_G + e\pi_C\pi_T + \pi_G\pi_T)]$

This model has 9 parameters to be specified: the parameters of the equilibrium distribution, $\pi = (\pi_A, \ldots, \pi_T)$, (since $\sum_i \pi_i = 1$, this only counts for 3 parameters) and the parameters $a, b, c, d, e, f > 0$. Note the form of the Q matrix here is chosen so that π is indeed the equilibrium distribution, that is, as $t \to \infty$, every row of $P(t) \to \pi$. Recall that $P(t) = \exp(tQ)$, where exp() is the matrix exponential.

The same modelling tools can be used when the bases are the 20 amino acids, the difference being that the Q matrix is now 20×20.

21.4 Estimating the maximum likelihood tree

According to the substitution model we are using, the best tree is the one which maximises the likelihood $\mathcal{L}(T) = \Pr(D|T)$ under that model. This is called the maximum

likelihood tree. Since there is no way to analytically find the maximum likelihood tree under general model of mutation, we can use similar techniques to those used for maximum parsimony to find something close to the maximum likelihood tree.

That is, we can start at some tree and use a stochastic search to propose new trees which are accepted if they have a higher likelihood. Note that we have the added complication when dealing with likelihoods that branch lengths now influence the likelihood of a tree, so for each tree topology, the branch lengths need to be optimised.

The hill-climbing algorithm we introduced in the context of parsimony trees is restated here for likelihood trees:

- choose an initial tree and calculate its likelihood.
- Iterate:
  - modify the tree and calculate its likelihood.
  - if the modified tree has a higher likelihood then the unmodified tree, keep it. Else, return to the previous tree
  - stop when no or minimal increase in likelihood occurs

Modifications to the tree can either change the tree topology (shape) or the length of the branches. The same topology changing operations as we used in the equivalent parsimony algorithm, such as SPR, can be extended to work with trees with explicit branch lengths as we have here. Modifications that change only the branch lengths are also used in this context.