1 T92 model of DNA evolution

1. If we set $\kappa = 1$ and $g = 0.5$, indicative of universal transition-transversion rates and homogeneous nucleotide frequencies, then the Q-matrix resolves to:

$$Q = \begin{pmatrix} - & 1/4 & 1/4 & 1/4 \\ 1/4 & - & 1/4 & 1/4 \\ 1/4 & 1/4 & - & 1/4 \\ 1/4 & 1/4 & 1/4 & - \end{pmatrix}$$

With the product of Q and μ, the overall rate of substitution, we can obtain the Jukes-Cantor model.

2. If $q_{ij} > 0$ for all $i \neq j$, then it is defined to have an existing stationary distribution (π). For the T92 model Q, $\kappa > 0 < g < 1$, which indicates that it does in-fact have a stationary distribution.

3. Given the information we have, it seems most intuitive to use the GC-content to model the distribution. To be stationary it must satisfy $\pi^T Q = 0^T$, so the distribution of $\pi_{(A,G,C,T)} = ((1-g)/2,\ g/2,\ g/2,\ (1-g)/2)$

$$(0,0,0,0) = \left(\tfrac{1-g}{2},\ \tfrac{g}{2},\ \tfrac{g}{2},\ \tfrac{1-g}{2}\right) \begin{pmatrix} -(\kappa g+1)/2 & \kappa g/2 & g/2 & (1-g)/2 \\ \kappa(1-g)/2 & -(\kappa - \kappa g + 1)/2 & g/2 & (1-g)/2 \\ (1-g)/2 & g/2 & -(\kappa - \kappa g + 1)/2 & \kappa(1-g)/2 \\ (1-g)/2 & g/2 & \kappa g/2 & -(\kappa g+1)/2 \end{pmatrix}$$

Since it is symmetrical, the equations match for A-T and G-C.

$$\pi_{A,T} = \frac{-(\kappa g+1)(1-g)}{4} + \frac{g\kappa(1-g)}{4} + \frac{g(1-g)}{4} + \frac{(1-g)^2}{4} = 0$$

$$\pi_{G,C} = \frac{\kappa g(1-g)}{4} - \frac{g(\kappa - \kappa g + 1)}{4} + \frac{g^2}{4} + \frac{g(1-g)}{4} = 0$$

Thereby $\pi = ((1-g)/2, g/2, g/2, (1-g)/2)$ is the stationary distribution of this T92 model.

4. If $\pi_i q_{ij} = \pi_j q_{ji}$ for all $i \neq j$ then it is shown to be time-reversible. This can be shown by equating all transition-transversion pairs:

$$\text{AC - CA} = \frac{(1-g)}{2}\frac{g}{2} = \frac{g}{2}\frac{(1-g)}{2}$$

$$\text{AG - GA} = \frac{(1-g)}{2}\frac{\kappa g}{2} = \frac{g}{2}\frac{\kappa(1-g)}{2}$$

$$\text{AT - TA} = \frac{(1-g)}{2}\frac{(1-g)}{2} = \frac{(1-g)}{2}\frac{(1-g)}{2}$$

$$\text{CG - GC} = \frac{g}{2}\frac{(-\kappa + \kappa g - 1)}{2} = \frac{(-\kappa + \kappa g - 1)}{2}\frac{g}{2}$$

$$\text{CT - CT} = \frac{g}{2}\frac{(\kappa(1-g))}{2} = \frac{(\kappa g)}{2}\frac{(1-g)}{2}$$

$$\text{GT - GT} = \frac{g}{2}\frac{(1-g)}{2} = \frac{(1-g)}{2}\frac{g}{2}$$

5. The overall rate of substitution, μ, is:

$\mu = \sum_{i\in(A,G,C,T)} \sum_{j\neq i} \pi_i q_{ij} = \sum_{i\in(A,G,C,T)} \pi_i(-q_{ii}) = \frac{1-g}{2}\frac{(\kappa g+1)}{2} + \frac{g}{2}\frac{(\kappa-\kappa g+1)}{2} + \frac{g}{2}\frac{(\kappa-\kappa g+1)}{2} + \frac{1-g}{2}\frac{(\kappa g+1)}{2}$

$$\mu = -\kappa g^2 + \kappa g + \frac{1}{2}$$

6. The Q-matrix was made using $q_{ij} = \nu_i p_{ij}, i \neq j$ with $\nu_i = -q_{ii}$. If we use the Q-matrix without respect to time, we obtain discrete transitions, and the P-matrix can be formulated as $p_{ij} = q_{ij}/-q_{ii}$, exploiting the properties of the diagonal in the Q. Thus, we get

$$P = \begin{pmatrix} 0 & \frac{\kappa g}{(\kappa g+1)} & \frac{g}{(\kappa g+1)} & \frac{(1-g)}{(\kappa g+1)} \\ \frac{\kappa(1-g)}{\kappa-\kappa g+1} & 0 & \frac{g}{\kappa-\kappa g+1} & \frac{(1-g)}{\kappa-\kappa g+1} \\ \frac{(1-g)}{\kappa-\kappa g+1} & \frac{g}{\kappa-\kappa g+1} & 0 & \frac{\kappa(1-g)}{\kappa-\kappa g+1} \\ \frac{(1-g)}{(\kappa g+1)} & \frac{g}{(\kappa g+1)} & \frac{\kappa g}{(\kappa g+1)} & 0 \end{pmatrix}$$