Appendix 2

The six-parameter version of the symmetric model can be depicted by a matrix of instantaneous rates of substitution:

$$\mathbf{Q} = \begin{bmatrix} - & r_{AC} & r_{AG} & r_{AT} \\ r_{CA} & - & r_{CG} & r_{CT} \\ r_{GA} & r_{GC} & - & r_{GT} \\ r_{TA} & r_{TC} & r_{TG} & - \end{bmatrix} = \begin{bmatrix} - & \mu_1 & \mu_2 & \mu_3 \\ \mu_1 & - & \mu_4 & \mu_5 \\ \mu_2 & \mu_4 & - & \mu_6 \\ \mu_3 & \mu_5 & \mu_6 & - \end{bmatrix}$$

where r_{ij} is the instantaneous rate of change from base i to base j. The matrix can be rescaled before the probabilities of substitution for a branch are calculated so that branch lengths can be interpreted as the expected number of changes per site. This results in the non-identifiablity of the 6 parameters – multiplying each element of a matrix by any constant, k, will result in the same likelihood because the matrices will be identical after rescaling.

The 5RR form of the symmetric model is an identifiable parameterization produced by dividing every element in the matrix by μ_6. The 5RR parameterization can be depicted:

$$\mathbf{Q} = \begin{bmatrix} - & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & - & \rho_4 & \rho_5 \\ \rho_2 & \rho_4 & - & 1 \\ \rho_3 & \rho_5 & 1 & - \end{bmatrix}$$

where $\rho_i = \frac{\mu_i}{\mu_6}$ for $1 \leq i \leq 5$.

The ST1 parameterization is another identifiable form of the model produced by constraining the rates (before rescaling) to sum to 1.0:

$$\mathbf{Q} = \begin{bmatrix} - & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & - & \phi_4 & \phi_5 \\ \phi_2 & \phi_4 & - & \left(1 - \sum_{i=0}^{5} \phi_i\right) \\ \phi_3 & \phi_5 & \left(1 - \sum_{i=0}^{5} \phi_i\right) & - \end{bmatrix}$$

For the sake of brevity we will use ϕ_6 to represent the quantity $1 - \sum_{i=0}^{5} \phi_i$, despite the fact that ϕ_6 is not a free parameter of the model.

To transform a point in the 5RR parameterization to a point in the ST1 form of the model, one uses the equation

$$\phi_i = \frac{\rho_i}{1 + \rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5}$$

for $i \leq 5$ and

$$\phi_6 = \frac{1}{1 + \rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5}$$

Note that

$$\frac{\partial \phi_i}{\partial \rho_j} = \begin{cases} (1 - \phi_i)\phi_6 & \text{if } i = j \\ -\phi_i \phi_6 & \text{if } i \neq j \end{cases}$$

so the Jacobian matrix for the transformation is

$$\mathbf{J} = \begin{bmatrix} (1-\phi_1)\phi_6 & -\phi_1\phi_6 & -\phi_1\phi_6 & -\phi_1\phi_6 & -\phi_1\phi_6 \\ -\phi_2\phi_6 & (1-\phi_2)\phi_6 & -\phi_2\phi_6 & -\phi_2\phi_6 & -\phi_2\phi_6 \\ -\phi_3\phi_6 & -\phi_3\phi_6 & (1-\phi_3)\phi_6 & -\phi_3\phi_6 & -\phi_3\phi_6 \\ -\phi_4\phi_6 & -\phi_4\phi_6 & -\phi_4\phi_6 & (1-\phi_4)\phi_6 & -\phi_4\phi_6 \\ -\phi_5\phi_6 & -\phi_5\phi_6 & -\phi_5\phi_6 & -\phi_5\phi_6 & (1-\phi_5)\phi_6 \end{bmatrix}$$

and the determinant is

$$|\mathbf{J}| \;=\; (1-\phi_1-\phi_2-\phi_3-\phi_4-\phi_5)^{-6}$$

In the 5RR parameterization, the determinant of the Jacobian of the transformation is:

$$\begin{aligned} |\mathbf{J}| &= \left(1-\frac{\rho_1}{1+\sum\rho_i}-\frac{\rho_2}{1+\sum\rho_i}-\frac{\rho_3}{1+\sum\rho_i}-\frac{\rho_4}{1+\sum\rho_i}-\frac{\rho_5}{1+\sum\rho_i}\right)^{-6} \\ &= \left(\frac{1+(\sum\rho_i)-\rho_1-\rho_2-\rho_3-\rho_4-\rho_5}{1+\sum\rho_i}\right)^{-6} \\ &= (1+\rho_1+\rho_2+\rho_3+\rho_4+\rho_5)^6 \end{aligned}$$

The uniform Dirichlet prior distribution has the density k for every combination of points. We can solve for k:

$$\begin{aligned} 1 &= \int_0^1 \int_0^{1-\phi_1} \int_0^{1-\phi_1-\phi_2} \int_0^{1-\phi_1-\phi_2-\phi_3} \int_0^{1-\phi_1-\phi_2-\phi_3-\phi_4} k\,d\phi_5 d\phi_4 d\phi_3 d\phi_2 d\phi_1 \\ &= \frac{k}{120} \end{aligned}$$

So, under a uniform Dirichlet prior, $k = 120$ and the probability density is 120 at every point in the ST1 parameteriztion. Thus, if the model is expressed in terms of the 5RR parameterization, but one wishes to use a prior that is equivalent to the uniform Dirichlet prior in the ST1 form of the model, then the prior density for any point in the 5RR space is given by:

$$\pi(\boldsymbol{\rho}) = 120\,(1 + \rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5)^{-6} \tag{6}$$

Calculation of probabilities in hypercubes

One can calculate the prior probability assigned to different regions of the 5RR parameter space (or any other parameterization) under different priors by computing definite integrals of the probability density function. For the $0 \leq \rho_i \leq 2$ hypercube the prior probability under a prior equivalent to the uniform Dirichlet is

$$\begin{aligned}
\Pr(0 \leq \rho_i \leq 2) &= 120 \int_0^2 \int_0^2 \int_0^2 \int_0^2 \int_0^2 (1 + \rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5)^{-6} d\rho_1 d\rho_2 d\rho_3 d\rho_4 d\rho_5 \\
&= -(9 + \rho_1)^{-1} + 4(7 + \rho_1)^{-1} - 6(5 + \rho_1)^{-1} + 4(3 + \rho_1)^{-1} - (1 + \rho_1)^{-1} \Big|_{\rho_1=0}^{\rho_1=2} \\
&\approx 0.3694
\end{aligned}$$

and for the $98 \leq \rho_i \leq 100$ hypercube the probability is

$$\begin{aligned}
\Pr(98 \leq \rho_i \leq 100) &= 120 \int_{98}^{100} \int_{98}^{100} \int_{98}^{100} \int_{98}^{100} \int_{98}^{100} (1 + \rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5)^{-6} d\rho_1 d\rho_2 d\rho_3 d\rho_4 d\rho_5 \\
&\approx 2.58 * 10^{-13}
\end{aligned}$$

If *a priori* $\rho_i \sim U(0, m)$, in other words the prior used with the 5RR parameterization is a uniform distribution from 0 to some arbitrary cutoff point, m, then the prior probabity for any hypercube $\ell \leq \rho_i \leq u$ may be calculated as follows:

$$\begin{aligned}
\Pr(\ell \leq \rho_i \leq u) &= \frac{1}{m^5} \int_\ell^u \int_\ell^u \int_\ell^u \int_\ell^u \int_\ell^u d\rho_1 d\rho_2 d\rho_3 d\rho_4 d\rho_5 \\
&= \left(\frac{u - \ell}{m} \right)^5
\end{aligned}$$

Because the probability only depends upon $u - \ell$, the regions with $0 \leq \rho_i \leq 2$ and $98 \leq \rho_i \leq 100$ have identical prior probabilities under this prior.

Appendix 3. Equivalence of Hyper-prior on 5RR and Uniform Dirichlet on ST1

One can use exponential priors on the rate parameters in the 5RR model with an exponential scaling factor λ. Taking a hierarchical approach, one could place an exponential prior with mean of ν on λ. Marginalizing over λ reveals how changing ν alters the joint prior on the rate parameters in the 5RR model:

$$\begin{aligned}
\rho_i &\sim \text{Exp}(\lambda) \\
\lambda &\sim \text{Exp}(\nu) \\
f(\boldsymbol{\rho}|\lambda) &= \prod_{i=1}^{5} \lambda e^{-\lambda \rho_i} \\
f(\lambda|\nu) &= \nu e^{-\nu\lambda} \\
f(\boldsymbol{\rho}|\nu) &= \nu \int_0^\infty \left(e^{-\nu\lambda} \prod_{i=1}^{5} \lambda e^{-\lambda\rho_i} \right) d\lambda
\end{aligned}$$

If we use ρ to indicate $\sum_{i=1}^{5} \rho_i$, then:

$$\begin{aligned}
f(\boldsymbol{\rho}|\nu) &= \nu \int_0^\infty \lambda^5 e^{-(\nu+\rho)\lambda} d\lambda \\
&= \nu \left(-\frac{120}{(\nu+\rho)^6} + \frac{120\lambda}{(\nu+\rho)^5} - \frac{60\lambda^2}{(\nu+\rho)^4} + \frac{20\lambda^3}{(\nu+\rho)^3} - \frac{5\lambda^4}{(\nu+\rho)^2} + \frac{\lambda^5}{(\nu+\rho)} \right) e^{-(\nu+\rho)\lambda} \Bigg|_{\lambda=0}^{\lambda=\infty} \\
&= \frac{120\nu}{(\nu+\rho)^6}
\end{aligned}$$

Thus, if $\nu = 1$, the prior density is identical to the density given in equation (6) as the 5RR prior that is identical to a uniform Dirichlet on the ST1 parameterization.