Metabarcoding model development in `NIMBLE`

A. J. Rominger

10 April 2017

The model

We know the total number of reads N_{reads}, the number of species S, the total number of individuals N, and the correlation structure from the phylogeny Σ_{phy}. We also know the vector of number of reads per species x_{read}. We do not know the rate of copy number evolution σ_{copy} nor the rate of primer affinity evolution σ_{primer}. We also most importantly don't know the actual vector of abundances x. We want to model the number of reads as multinomial conditioned on N_{reads} and S, and with a probability vector p given by

$$p \sim \text{dirch}(x\nu\lambda)$$

where ν is a variable proportional to copy number and λ is a variable proportional to primer affinity. We will assume that $\log\nu$ evolved according to a Brownian motion process with rate σ_{copy} and mean $\mu_{copy} = 0$. We will assume that $\text{logit}_{min=\alpha}(\lambda)$ evolved according to an independent Brownian motion process with rate σ_{primer} and mean $\mu_{primer} = 0$. $\text{logit}_{min=\alpha}$ is a modified *logit* function which assumes λ is constrained by a lower bound $0 \leq \alpha \leq 1$. Its corresponding modified *expit* function is

$$\text{expit}(x) = \alpha + \frac{1 - \alpha}{1 + \exp(-x)}$$

The variance-covariance matrix for a Brownian-motion process on a known phylogeny is given as

$$\Sigma_{i,j} = \sigma d_{MRCA(i,j)}$$

where $d_{MRCA(i,j)}$ is the distance from the root to the most recent common ancestor of tips i and j. $d_{MRCA(i,i)}$ is simply the depth of the root.

Thus our full model is

$$x_{reads} \sim \text{multinom}(N_{reads}, p_1, \ldots, p_S)$$

where

$$p \sim \text{dirch}(x\nu\lambda)$$

and ν and λ follow the following distributions

$$\log(\nu) \sim \text{mvnorm}(0, \sigma_{copy} d_{MRCA(i,j)})$$

$$\text{logit}_{min=\alpha}(\nu) \sim \text{mvnorm}(0, \sigma_{primer} d_{MRCA(i,j)})$$

with priors

$$x/N \sim \text{dirch}(N/S, \ldots, N/S)$$

$$\frac{1}{\sigma_{copy}} \sim \Gamma(0.001, 0.001)$$

$$\frac{1}{\sigma_{primer}} \sim \Gamma(0.001, 0.001)$$

$$\alpha \sim \text{Unif}(0, 1)$$