 DOI: 10.1086/667574

Appendix B from R. S. Etienne and B. Haegeman, "A Conceptual and Statistical Framework for Adaptive Radiations with a Key Role for Diversity Dependence"

(Am. Nat., vol. 180, no. 4, p. E75)

Computing the Likelihood When There Are Missing Species

The algorithm in box 1 gives the likelihood for the diversity-dependent model without decoupling even when there are species missing from the phylogeny. The likelihood is simply element $q + m$, q being the number of extant species in the phylogeny and m being the number of extant species missing from the phylogeny, of the vector of probabilities resulting from the integration of the ordinary differential equation divided by a combinatorial factor:

$$L_m(\text{tree}) = \frac{Q_{q+m}(t_\text{p})}{\binom{q+m}{m}}. \tag{B1}$$

This combinatorial factor is explained by Etienne et al. (2012). Here we explain how missing species can be dealt with in the model with decoupling of diversity dependence. The total likelihood for clades M and S combined is given by

$$L_m[\text{M \& S}] = \sum_{\{m_\text{M}, m_\text{S} | m_\text{M}+m_\text{S}=m\}} P(m_\text{M}, m_\text{S}|m) L_{m_\text{M}}[\text{M}] L_{m_\text{S}}[\text{S}], \tag{B2}$$

where $L_{m_\text{M}}[\text{M}]$ is the likelihood of the main clade M, with m_M missing species (see box 1; eq. [B2]),

$$L_{m_\text{M}}[\text{M}] = \frac{Q^\text{M}_{q_\text{M}+m_\text{M}}(t_\text{p})}{\binom{q_\text{M}+m_\text{M}}{m_\text{M}}}, \tag{B3}$$

and, similarly, $L_{m_\text{S}}[\text{S}]$ is the likelihood of the subclade S with the remaining $m_\text{S} = m - m_\text{M}$ missing species,

$$L_{m_\text{S}}[\text{S}] = \frac{Q^\text{S}_{q_\text{S}+m_\text{S}}(t_\text{p})}{\binom{q_\text{S}+m_\text{S}}{m_\text{S}}}, \tag{B4}$$

and $P(m_\text{M}, m_\text{S}|m)$ is the probability of m_M out of m missing species being assigned to clade M and the remainder, $m_\text{S} = m - m_\text{M}$, being assigned to clade S. Because all topologies have equal probability (Etienne et al. 2012), the probability of selecting m_M out of $q_\text{M} + m_\text{M}$ species and m_S out of $q_\text{S} + m_\text{S}$ species, where q_M and q_S are the number of species in clades M and S, respectively, is simply given by the hypergeometric distribution

$$P(m_\text{M}, m_\text{S}|m) = \frac{\binom{q_\text{M}+m_\text{M}}{m_\text{M}}\binom{q_\text{S}+m_\text{S}}{m_\text{S}}}{\binom{q_\text{M}+q_\text{S}+m}{m}}. \tag{B5}$$

Substituting equations (B3), (B4), and (B5) into (B2) we find

$$\begin{aligned} L_m[\text{M \& S}] &= \sum_{\{m_\text{M}, m_\text{S} | m_\text{M}+m_\text{S}=m\}} \frac{\binom{q_\text{M}+m_\text{M}}{m_\text{M}}\binom{q_\text{S}+m_\text{S}}{m_\text{S}}}{\binom{q_\text{M}+q_\text{S}+m}{m}} \frac{Q^\text{M}_{q_\text{M}+m_\text{M}}(t_\text{p})}{\binom{q_\text{M}+m_\text{M}}{m_\text{M}}} \frac{Q^\text{S}_{q_\text{S}+m_\text{S}}(t_\text{p})}{\binom{q_\text{S}+m_\text{S}}{m_\text{S}}}, \\ &= \frac{\sum_{j=0}^{m} Q^\text{M}_{q_\text{M}+j}(t_\text{p}) Q^\text{S}_{q_\text{S}+m-j}(t_\text{p})}{\binom{q_\text{M}+q_\text{S}+m}{m}}. \end{aligned} \tag{B6}$$

which is the final result of box 2.