1 Materials and Methods

1.1 Taxon occurrence information

1.2 Model specification

A taxon's ecotype is modeled as a realization from a categorical distribution with K possible outcomes, where each outcome has its own probability.

$$\text{softmax}(y_k) = \frac{\exp(y_k)}{\sum_{k=1}^{K} exp(y_k)} \tag{1}$$

N samples. K categories. C cohorts. D is number of individual-level covariates. E is number of cohort-level covariates. π is a length K vector such that $\sum_{k=1}^{K} \pi_k = 1$. X is a $D \times N$ matrix of individual-level covariates. U is a $E \times C$ matrix of cohort-level covariates. β is a $K \times D$ matrix of regression coefficients for the individual-level predictors. γ is a $K \times E$ number of regression coefficients for the group-level predictors.

$$\begin{aligned}
y_i &\sim \text{Categorical}(K, \pi_{ik}) \\
\pi_{ik} &= \frac{\eta_{ik}}{\sum_{k=1}^{K} \eta_{ik}} \\
\eta_{ik} &= \alpha_{kc[i]} + \beta_k X_i \\
\alpha_{kc} &\sim \mathcal{N}(\alpha'_k + \gamma U_c, \sigma_k) \\
\alpha'_k &\sim \mathcal{N}(0, 5) \\
\beta &\sim \mathcal{N}(0, 1) \\
\gamma &\sim \mathcal{N}(0, 1) \\
\sigma_k &\sim \text{C}^{+}(1)
\end{aligned} \tag{2}$$

Where $\pi_{iK} = 0$ for $i = 1, 2, \dots, N$. This last statement is necessary because the softmax function is invariable to the addition of a constant (Eq. 1), thus this insures identifiability.

1.3 Posterior inference