

1 SUPPLEMENTAL INFORMATION

2 **Modeling phylogenetic biome shifts on a planet with a past**

3 Michael J. Landis^{1,2,*}, Erika J. Edwards^{2,3}, and Michael J. Donoghue^{2,3}

4
5 ¹*Department of Biology, Washington University in St. Louis, One Brookings Drive,*
6 *St. Louis, Missouri 63130, USA*

7 ²*Department of Ecology & Evolutionary Biology, Yale University, PO Box 208106,*
8 *New Haven, Connecticut 06520, USA*

9 ³*Division of Botany, Yale Peabody Museum of Natural History, P.O. Box 208118,*
10 *New Haven, Connecticut 06520, USA*

11 * Email correspondence: michael.landis@wustl.edu

12 **Supplement 1: Computing the value of the rate matrix, $Q(m)$**

13 We provide an example for computing $Q(m)$ for a simple toy system with two
 14 biomes (A and B) and two regions (X and Y). Rate matrices for other time intervals are
 15 computed similarly for all values of m , so we focus only on the first time slice ($m = 1$). In
 16 our example, we assume the availability/connectivity of biomes and regions at time $m = 1$
 17 is represented by the adjacency matrix for geography,

18
$$A_G(1) = \begin{matrix} X & Y \\ Y & X \end{matrix} \begin{pmatrix} 1 & y \\ y & 1 \end{pmatrix},$$

19 for biome A,

20
$$A_A(1) = \begin{matrix} X & Y \\ Y & Y \end{matrix} \begin{pmatrix} 1 & y \\ y & y \end{pmatrix},$$

21 and for biome B,

22
$$A_B(1) = \begin{matrix} X & Y \\ Y & X \end{matrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

23 where availability (diagonal) or connectivity (off-diagonal) is coded as strong (1), weak
 24 ($0 < y < 1$; estimated), or marginal (0). These adjacency matrices are then used to
 25 (potentially) rescale the dispersal rates among regions and the biome shift rates among
 26 biomes for three rate matrices: for the geographically uninformative rate matrix,

27
$$Q_U = \begin{matrix} AX \\ AY \\ BX \\ BY \end{matrix} \begin{pmatrix} - & \delta_{XY} & \beta_{AB} & 0 \\ \delta_{YX} & - & 0 & \beta_{AB} \\ \beta_{BA} & 0 & - & \delta_{XY} \\ 0 & \beta_{BA} & \delta_{YX} & - \end{pmatrix},$$

28 the geography-dependent rate matrix,

29
$$Q_G(1) = \begin{matrix} AX \\ AY \\ BX \\ BY \end{matrix} \begin{pmatrix} - & [A_G(1)]_{XY}\delta_{XY} & \beta_{AB} & 0 \\ [A_G(1)]_{YX}\delta_{YX} & - & 0 & \beta_{AB} \\ \beta_{BA} & 0 & - & [A_G(1)]_{XY}\delta_{XY} \\ 0 & \beta_{BA} & [A_G(1)]_{YX}\delta_{YX} & - \end{pmatrix}$$

$$30 \quad = \begin{matrix} AX \\ AY \\ BX \\ BY \end{matrix} \begin{pmatrix} - & y\delta_{XY} & \beta_{AB} & 0 \\ y\delta_{YX} & - & 0 & \beta_{AB} \\ \beta_{BA} & 0 & - & y\delta_{XY} \\ 0 & \beta_{BA} & y\delta_{YX} & - \end{pmatrix},$$

31 and the biome-dependent rate matrix,

$$32 \quad Q_B(1) = \begin{matrix} AX \\ AY \\ BX \\ BY \end{matrix} \begin{pmatrix} - & [A_A(1)]_{XY}\delta_{XY} & [A_B(1)]_{XX}\beta_{AB} & 0 \\ [A_A(1)]_{YX}\delta_{YX} & - & 0 & [A_B(1)]_{YY}\beta_{AB} \\ [A_A(1)]_{XX}\beta_{BA} & 0 & - & [A_B(1)]_{YX}\delta_{XY} \\ 0 & [A_A(1)]_{YY}\beta_{BA} & [A_B(1)]_{YX}\delta_{YX} & - \end{pmatrix}$$

$$33 \quad = \begin{matrix} AX \\ AY \\ BX \\ BY \end{matrix} \begin{pmatrix} - & y\delta_{XY} & 0\beta_{AB} & 0 \\ y\delta_{YX} & - & 0 & 1\beta_{AB} \\ 1\beta_{BA} & 0 & - & 0\delta_{XY} \\ 0 & y\beta_{BA} & 0\delta_{YX} & - \end{pmatrix}.$$

34 The value of the rate matrix for the regional biome shift process for a given time

35 slice, $Q(m)$, is then determined as by weighted mixture of these rate matrices. If we

36 suppose that $w = (w_U, w_G, w_B) = (0.1, 0.2, 0.7)$, then we have

$$37 \quad Q(1) = 0.1Q_U \times 0.2Q_G(1) \times 0.7Q_B(1).$$

38 To assign exact rate values to $Q(1)$, we set $\delta_{XY} = \delta_{YX} = 0.6$, $\beta_{AB} = \beta_{BA} = 0.4$, and $y = 0.3$,

39 producing four dispersal event rate values

$$40 \quad [Q(1)]_{(AX),(AY)} = 0.6 \times ((0.1)(1.0) + (0.2)(0.3) + (0.7)(0.3)) = 0.222$$

$$41 \quad [Q(1)]_{(AY),(AX)} = 0.6 \times ((0.1)(1.0) + (0.2)(0.3) + (0.7)(0.3)) = 0.222$$

$$42 \quad [Q(1)]_{(BX),(BY)} = 0.6 \times ((0.1)(1.0) + (0.2)(0.3) + (0.7)(0.0)) = 0.096$$

$$43 \quad [Q(1)]_{(BY),(BX)} = 0.6 \times ((0.1)(1.0) + (0.2)(0.3) + (0.7)(0.0)) = 0.096$$

44 and four biome shift rates values

$$45 \quad [Q(1)]_{(AX),(BX)} = 0.4 \times ((0.1)(1.0) + (0.2)(1.0) + (0.7)(0.0)) = 0.120$$

$$46 \quad [Q(1)]_{(BX),(AX)} = 0.4 \times ((0.1)(1.0) + (0.2)(1.0) + (0.7)(1.0)) = 0.400$$

$$47 \quad [Q(1)]_{(AY),(BY)} = 0.4 \times ((0.1)(1.0) + (0.2)(1.0) + (0.7)(1.0)) = 0.400$$

48 $[Q(1)]_{(BY),(AY)} = 0.4 \times ((0.1)(1.0) + (0.2)(1.0) + (0.7)(0.3)) = 0.204$

49 with all remaining non-diagonal elements being zero, and each diagonal element equaling
50 the negative sum of the rates departing that row's biome-region state. When fitting the
51 time-stratified regional biome shift model to phylogenetic datasets, we populate the rate
52 matrices for each time interval in a similar manner, $Q = (Q(1), Q(2), \dots, Q(M))$, and further
53 multiply each rate matrix by the global scaling factor, μ , to control the overall rate of the
54 process. Note that dispersal rates between regions X and Y are symmetrical, while shifts
55 between biomes asymmetrical, due to differences in the availability in biomes A and B
56 within regions X and Y.

57

58 ***Supplement 2: Sensitivity analysis of biome structure models***

59 Biased models may predispose inferences towards certain results, despite the
60 presence of data that would otherwise support alternative results. In our case, it is possible
61 that inferences drawn from our posterior distribution of stochastic mappings (e.g. how
62 'biome match' proportions differ before and after the Oligocene; Fig. 6) are not driven by
63 information in the data through the likelihood function, but rather that the signal is
64 determined largely by constraints imposed by the *Paleobiome*, *Modern Biome*, and *Null*
65 *Biome* structure models. We wished to address this concern. To do so, we viewed the biome
66 structure models as empirically structured priors, which allowed us to ask whether those
67 priors have outsized influence on posterior estimates. Put another way, does our choice of
68 empirically structured prior induce biased results?

69 To assess whether this was an issue, we performed a sensitivity analysis by fitting
70 the *Paleobiome*, *Modern Biome*, and *Null Biome* models to the *Viburnum* dataset in the same

71 manner as described in the main text, with three notable changes. First, we forced the
72 model likelihood function to return the value of 1 for all prior settings. In a Bayesian
73 context, this will cause the posterior distribution to be identical to the prior. Even though
74 the prior is technically data-independent, we nonetheless generated stochastic mappings
75 that conditioned on the biome-region states at the tips of the phylogeny when sampling
76 model parameters from the prior (not the posterior). These stochastic mappings therefore
77 represent a sample of evolutionary histories that are compatible with the comparative
78 dataset under an assumed biome structure model (*Paleobiome*, *Modern Biome*, and *Null*
79 *Biome*) for parameter values as supported by our prior. Standard stochastic mapping
80 algorithms are typically applied using model parameters that explain the data well (e.g.
81 using maximum likelihood estimators or posterior samples for model parameters). Much of
82 prior parameter space, however, explains a particular dataset extremely poorly. We found
83 that our stochastic mapping algorithm was numerically unstable for some of our most
84 extreme rates (e.g. when μ equaled 10^{-4} or 10^1 events/Myr). To improve algorithmic
85 stability, we made our second modification to the original analysis settings, by using a more
86 conservative prior on the global event rate for biome-region state transitions, substituting
87 the original prior of $\mu \sim \text{Loguniform}(10^{-4}, 10^1)$ with $\mu \sim \text{Loguniform}(10^{-3}, 10^0)$. Posterior
88 estimates under the original prior were generally in the range of 10^{-2} to 10^{-1} and exclude
89 values as extreme as 10^{-4} or 10^1 , and both priors share an expectation of 10^{-1} , so we believe
90 this substitution is wholly appropriate given the unusual nature of the problem we are
91 attempting to address. Finally, our sensitivity analyses differed in that we ran MCMC for
92 fewer iterations to collect the same number of samples, since it is generally easier to
93 sample from the prior than it is to do so from the posterior. In addition to generating the

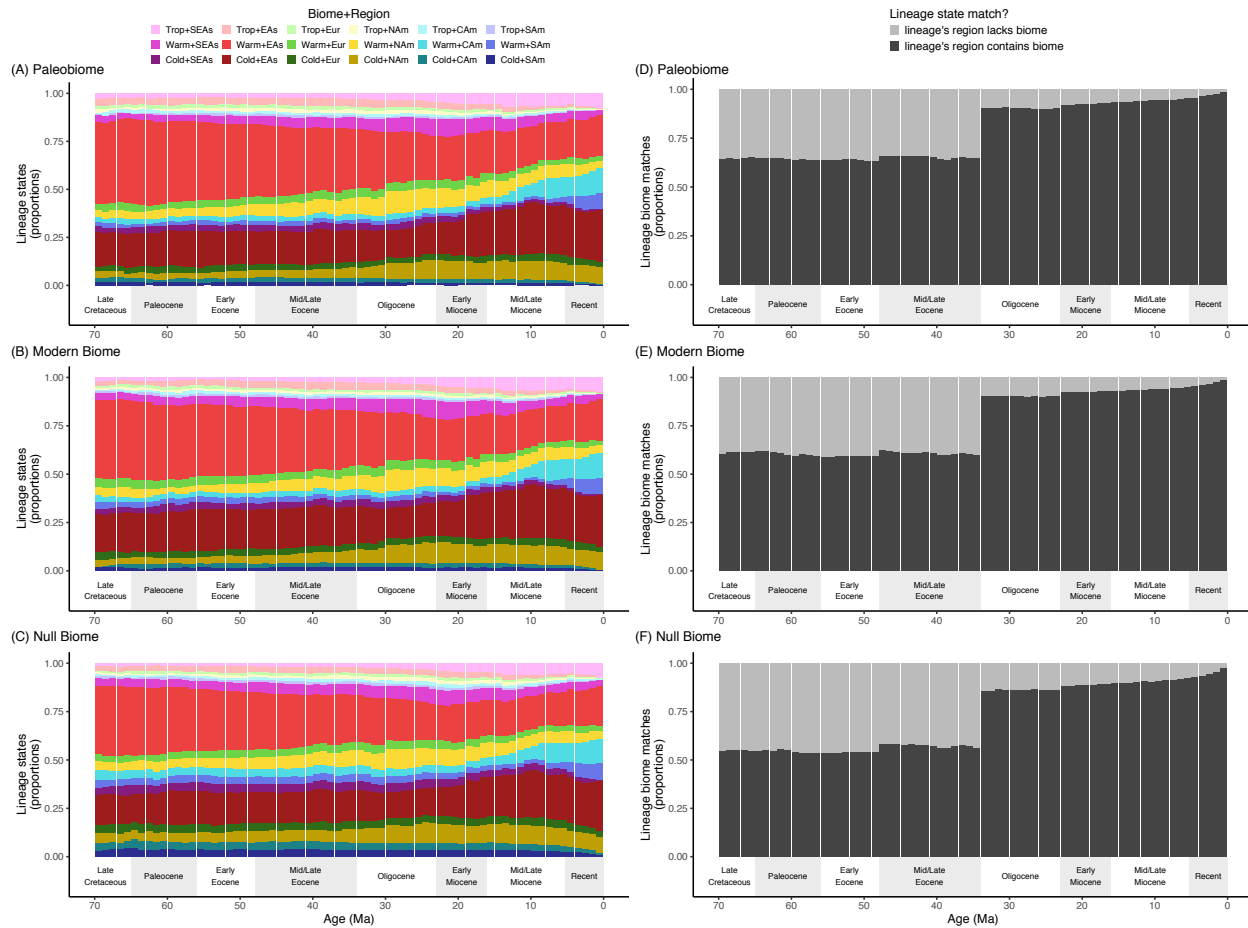
94 three induced prior distributions of stochastic mappings, we also generated the induced
95 prior distribution of root state frequencies.

96 Our sensitivity analyses show that regardless of which biome structure model we
97 assumed, all reconstructions produced nearly identical lineage state proportions through
98 time (Fig S1A–C) and nearly identical lineage-biome match proportions through time (Fig
99 S1D–F). In general, posterior estimates for lineage-state proportions supported only an
100 East Asian origin of *Viburnum* (Fig. 6A–C), whereas our prior-based estimates awarded
101 diffuse support to a broad range of alternative regions of origin (Fig. 6D–F). Conflicting
102 posterior estimates of a tropical, warm temperate, or cold temperate origin of *Viburnum*
103 that were contingent on which biome structure model was analyzed (Fig. 6A–C) is
104 completely erased under the prior-based inference (Fig. S1A–C). Lineage-biome match
105 proportions under the prior lingered around 65% until the end of the Oligocene, after
106 which they rose to ~85%, which are most similar to the posterior estimates under the
107 *Modern* and *Null* biomes (Fig. 6E,F). By contrast, posterior estimates under the *Paleobiome*
108 model (Fig. 6D) inferred high proportions of lineage-biome matches (>95%) over all time
109 intervals.

110 Prior distributions of root state frequencies were generally insensitive to which
111 biome structure model was assumed (Fig. S2). The *Paleobiome*, *Modern Biome*, and *Null*
112 *Biome* structures all have median state frequencies that are close $1/18 \approx 0.056$, i.e. the
113 value one expects if all 18 states had equal prior probability. Highest posterior densities are
114 also similar across biome structure models and biome-region states. In contrast, posterior
115 root state frequency estimates (Fig. 7) departed significantly from the value 0.056, and in
116 ways that reflect the given biome structure model. For example, the small posterior

117 probabilities for the Cold+SEAs and Cold+SAm root states do not include the value 0.056 in
118 their 95% highest posterior densities (HPDs). In contrast, the *Null Biome* prior and
119 posterior density have fairly similar medians, although their HPDs differ.

120 We did not find compelling support that our posterior estimates for *Viburnum*,
121 especially for those results under the *Paleobiome* model, are due to inherent and
122 overwhelming biases in the underlying biome structure models. Rather, posterior
123 inferences under alternative biome structure models differed in large part because of how
124 each model fit its parameters to the datasets through the model (i.e. through the likelihood
125 function). Because these results will not hold for any comparative dataset or for any biome
126 structure model imaginable, future researchers wishing to use our biome shift model to
127 test hypotheses in other empirical systems are advised to perform their own sensitivity
128 analyses.



129

130 **Figure S1.** Ancestral proportions of lineage state frequencies through time for *Viburnum* as

131 estimated under the prior (Supplement 2). The left column (A–C) shows the lineages

132 biome-region states, where regions differ by color and biomes differ by shading (see

133 legend). Proportions of reconstructed lineages in each biome-region state are shown for

134 estimates under the *Paleobiome* (A), *Modern Biome* (B), and *Null Biome* (C) settings. The

135 right column (D–F) shows the proportion of lineages with biome states that match (dark)

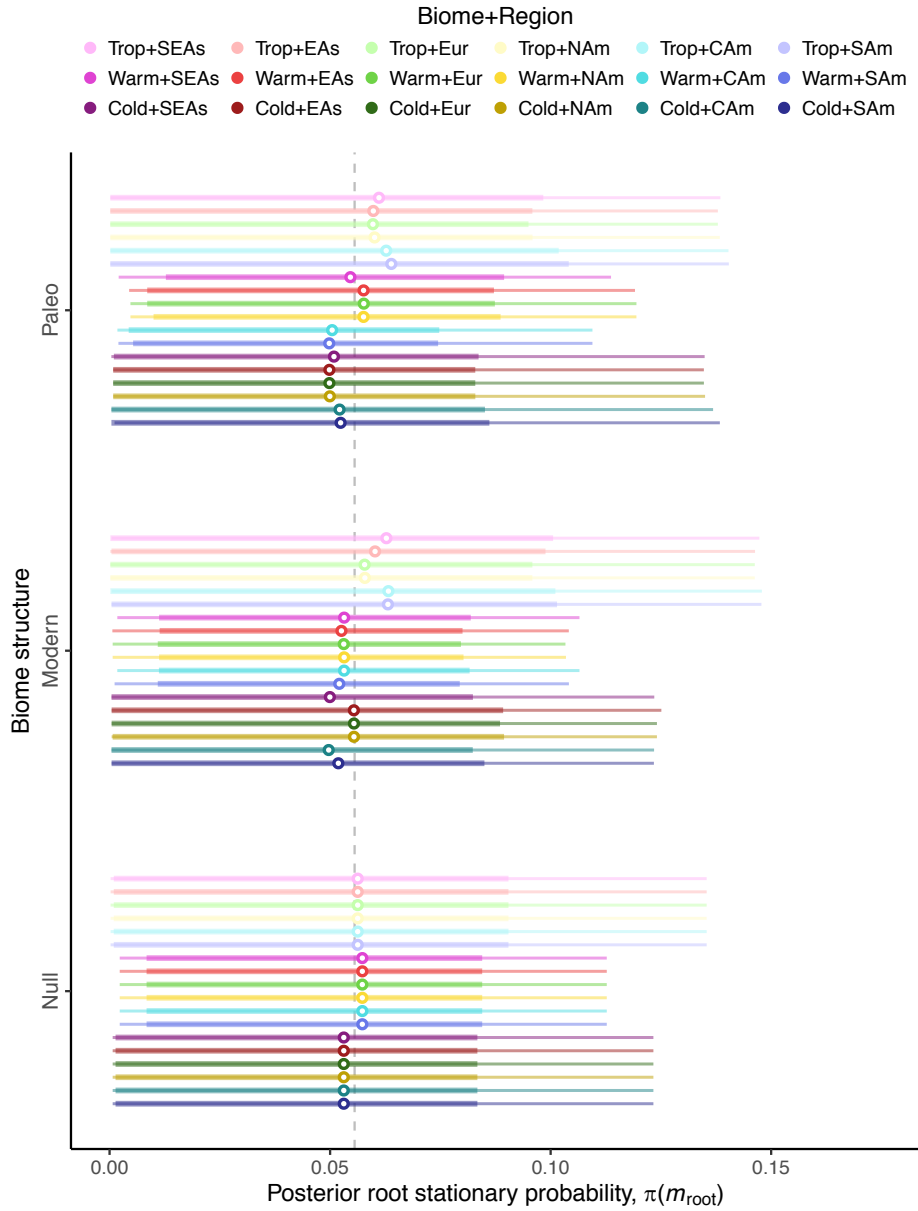
136 or mismatch (light) the non-marginal biomes that are locally accessible given any lineage’s

137 location, as defined under the *Paleobiome* structure (see main text for details). Proportions

138 of reconstructed lineages with biome match and mismatch scores are shown for estimates

139 under the *Paleobiome* (D), *Modern Biome* (E), and *Null Biome* (F) settings.

140



141

142 **Figure S2.** Stationary probabilities for the *Viburnum* root state during the Late Cretaceous

143 under the prior. Prior stationary probabilities for $\pi(m_{\text{root}})$ are given for each biome

144 structure model (grouped rows) and for each biome-region state (colors) as posterior

145 means (points) and credible intervals (HPD80, thick lines; HPD95, thin lines).

146