1   Version dated: February 25, 2019

Interdependent Phenotypic and Biogeographic Evolution Driven by Biotic Interactions

5   IGNACIO QUINTERO[1,2] AND MICHAEL J. LANDIS[1]

6   [1] *Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, 06511, USA*

7   [2] *Institut de Biologie de l'École Normale Supérieure (IBENS), ENS, 75005 Paris, France*

8   **Corresponding author:** Ignacio Quintero, Institut de Biologie de l'École Normale

9   Supérieure (IBENS), ENS, 75005 Paris, France; E-mail: ignacioquinterom@gmail.com,

10   Phone: +33 0610902475

11   *Abstract.*— Biotic interactions are hypothesized to be one of the main processes shaping

12   trait and biogeographic evolution during lineage diversification. Theoretical and empirical

13   evidence suggests that species with similar ecological requirements either spatially exclude

14   each other, by preventing the colonization of competitors or by driving coexisting

15   populations to extinction, or show niche divergence when in sympatry. However, the extent

16   and generality of the effect of interspecific competition in trait and biogeographic evolution

17   has been limited by a dearth of appropriate process-generating models to directly test the

18   effect of biotic interactions. Here, we formulate a phylogenetic parametric model that

19 allows interdependence between trait and biogeographic evolution, thus enabling a direct

20 test of central hypotheses on how biotic interactions shape these evolutionary processes.

21 We adopt a Bayesian data augmentation approach to estimate the joint posterior

22 distribution of trait histories, range histories, and co-evolutionary process parameters

23 under this analytically intractable model. Through simulations, we show that our model is

24 capable of distinguishing alternative scenarios of biotic interactions. We apply our model

25 to the radiation of Darwin's finches—a classic example of adaptive divergence—and find

26 support for *in situ* trait divergence in beak size, convergence in traits such as beak shape

27 and tarsus length, and strong competitive exclusion throughout their evolutionary history.

28 Our modeling framework opens new possibilities for testing more complex hypotheses

29 about the processes underlying lineage diversification. More generally, it provides a robust

30 probabilistic methodology to model correlated evolution of continuous and discrete

31 characters.

32 (Keywords: biotic interactions, competition, trait evolution, historical biogeography,

33 Bayesian, data augmentation)

34   One of the major goals of biogeography is to explain the dramatic variation in

35 species richness across the planet. Ultimately, any difference in species richness between

36 two regions stems from contrasting frequencies of speciation, extinction or dispersal events

37 (Ricklefs 1987). While diversification processes alone drive the total number of species

38 through time, range evolution dynamics cannot be ignored when explaining spatial

39 gradients of biodiversity (Wiens and Donoghue 2004). Indeed, the increase in richness

40 within an area can only be the result of a new species eventually coming into (or remaining

41 in) sympatry (Weir and Price 2011; Pigot and Tobias 2013). This necessarily involves two

general processes: that of lineage splitting followed by that of establishing coexistence. Yet, we still lack a basic understanding on the generality and magnitude of the different processes that shape the geographical and phenotypic evolution of diversifying lineages (Mayr 1970; MacColl 2011; Tobias et al. 2014; Clarke et al. 2017).

Evidence suggests that the great majority of speciation processes, at least in terrestrial animals, involve an allopatric phase, with few conclusive examples demonstrating parapatric or sympatric speciation in nature (Mayr 1970; Coyne and Orr 2004; Rundell and Price 2009), but see (Stroud and Losos 2016). The prevailing view asserts that new species arise from geographically isolated populations that evolve sufficient morphological, ecological, physiological, behavioral and/or genetic differences to act as reproductive barriers. These incipient species usually fill very similar ecological niches since the initial driver of reproductive isolation was chance separation by geographical barriers (Kozak and Wiens 2006; Rundell and Price 2009; Cadena et al. 2011; Smith et al. 2014). Equivalent ecological requirements are supposed to make long-term coexistence untenable, following the competitive exclusion principle (Gause 1934; Hardin 1960; Macarthur and Levins 1967). Recent radiations often follow this principle, with closely related species occupying similar habitats but separated by physical barriers (recognized more than one century ago as the "general law of distribution"; Jordan 1905; Rundell and Price 2009). For species to attain sympatry, and thus elevate local richness, coexistence theory predicts that species must diverge sufficiently along one or more niche axes to avoid competition (Elton 1946; Hardin 1960; Macarthur and Levins 1967; Diamond 1978; Grether et al. 2009; Godoy et al. 2014).

Consequently, biotic interactions seem to be paramount in shaping trait and biogeographic distributions of evolving lineages. The effects of biotic interactions during evolutionary radiations can be broadly categorized in three ways: by limiting (or enhancing) geographical expansion (Rundell and Price 2009; Ricklefs 2010; Weir and Price 2011; Pigot and Tobias 2013; Tobias et al. 2014; Pigot et al. 2018), by promoting (or reducing) local extinction (Slatkin 1974; Simberloff and Boecklen 1991; Valone and Brown

69   1995), and by inducing niche divergence (or convergence) in coexisting species (Lack 1954;

70   Rohwer 1973; Schluter 2000; Davies et al. 2007; Pfennig and Pfennig 2012). While there

71   are experimental tests and suitable models for shallow divergences under population

72   genetic or ecological models (e.g., Lotka 1924; Neuhauser and Pacala 1999; Schluter 2000;

73   Scheffer and van Nes 2006), the long-term evolutionary consequences of biotic interactions

74   measured at ecological time-scales remain difficult to characterize. Except for a few

75   illuminating—but serendipitous—fossil sequences (Elredge 1974; Schindel and Gould 1977),

76   our understanding has been mostly restricted to tests of phylogenetic community structure

77   metrics, such as measures of trait under/over-dispersion juxtaposed to null models (Webb

78   et al. 2002; Cavender-Bares et al. 2009), and correlative analyses, such as sister-species

79   comparisons between allopatric species and those that have achieved secondary sympatry

80   (Schluter et al. 1985; Davies et al. 2007; Pigot and Tobias 2013; Anacker and Strauss 2014;

81   Freeman 2015; Cadotte et al. 2017; McEntee et al. 2018). Though insightful, such

82   pattern-based studies rely on non-generative models that do not disentangle how the

83   processes are driven by biotic interactions over evolutionary timescales. The different

84   stages of biotic interactions unfold in a complex interplay between phenotype and

85   geographical distribution, often ephemeral through the evolutionary history of species

86   (Brown and Wilson 1956), and most probably lost when evidence is restricted to

87   contemporaneous observations (Schindel and Gould 1977). To understand this interplay,

88   generative phylogenetic models are needed that allow for the reciprocity of trait-range

89   distributions during radiations that unfold over millions of years.

90   Event-based phylogenetic models have pivotally advanced our understanding of trait

91   and range dynamics of lineages through time (e.g, Butler and King 2004; Ree et al. 2008;

92   Lemey et al. 2010; Goldberg et al. 2011; Uyeda and Harmon 2014; Gill et al. 2017).

93   Standard phylogenetic models, however, generally disregard one or several features that are

94   essential to an idealized model of trait-range evolution. Two key features are (1) that

95   lineages should evolve interdependently with one another and (2) that trait dynamics and

₉₆ range dynamics should be capable of influencing one another. Addressing the first

₉₇ challenge, Nuismer and Harmon (2015) derived a stochastic differential equation (SDE) to

₉₈ test for the effect of biotic interactions under a phylogenetic tree and present day species

₉₉ data (see also Clarke et al. 2017). Because species must be in sympatry to interact, Drury

₁₀₀ et al. (2016) and Clarke et al. (2017) extended the framework to limit species interactions

₁₀₁ to those times when lineages were estimated to be in sympatry. Drury et al.'s and Clarke

₁₀₂ et al.'s methods relies on pre-estimating a distribution of ancestral ranges, and then

₁₀₃ conditioning on those histories to estimate ancestral trait dynamics. One consequence of

₁₀₄ this is that the range dynamics unidirectionally influence trait evolution. The second

₁₀₅ challenge relates to how multiple traits within a single lineage co-evolve. For discrete traits,

₁₀₆ Sukumaran and Knowles (2018) proposed a joint dependence between geographical and

₁₀₇ binary traits in a discrete setting by treating the two traits as a single compound trait,

₁₀₈ then modeling the evolution of that trait with an appropriately structured rate matrix.

₁₀₉ Lartillot and Poujol (2011) introduced a phylogenetic method that jointly models the

₁₁₀ co-evolution of continuous traits, discrete traits, and (hidden) lineage-specific evolutionary

₁₁₁ rates or parameters. And while Lartillot and Poujol's software implementation of the

₁₁₂ method, `coevol`, is specialized to study how molecular substitution processes are

₁₁₃ unidirectionally shaped by life history traits, the underlying design of `coevol`'s inference

₁₁₄ machinery is suited to more general problems in which continuous traits influence the

₁₁₅ instantaneous transition rates for models of discrete trait evolution. This is to say that

₁₁₆ fitting phylogenetic models with either interactions between lineages or with interactions

₁₁₇ between characters are both challenging problems, each in its own right.

₁₁₈   In our work, we build upon these pioneering studies to develop a new parametric

₁₁₉ model to test for the effect of biotic interactions on the interplay between trait evolution

₁₂₀ and biogeographic history. First, to better reflect theoretical expectations, we reformulate

₁₂₁ the SDE describing trait evolution such that the pressure from coexisting species is

₁₂₂ stronger when lineage traits are most similar, and wanes as traits diverge. Second, instead

123 of supplying a pre-estimated distribution of biogeographic histories, we simultaneously

124 infer biogeographic and trait histories to model interdependence among trait evolution,

125 sympatry, dispersal, and biotic interactions. Third, we allow trait evolution to directly

126 affect the colonization and local extinction rates of lineages throughout their biogeographic

127 history. Specifically, the colonization and local extinction rates for a lineage at a given time

128 depend on the trait values of lineages present across the different biogeographic areas.

129 Notably, our generative model allows direct examination of the distinct contributions of

130 *pre-* and *post*-sympatric niche divergence while attaining secondary contact. For instance,

131 a lineage attempting to colonize a given area might be limited by the similarity among its

132 trait value and those from the species in that area (i.e., competitive exclusion), suggesting

133 a role of *pre*-sympatric niche divergence for successful colonization. Conversely, a lineage

134 could readily colonize any area, independent of the trait distribution found there, but be

135 forced to change because of strong *in situ* interspecific competition, indicating

136 *post*-sympatric niche divergence. We note, however, that we do not model the intricacies of

137 geographic speciation at the nodes and assume that allopatric speciation does not occur;

138 we leave the modeling of this important speciational process to forthcoming work.

139   Our method fits the model using data augmentation within a Bayesian framework

140 to perform parameter inference, enabling accurate propagation of uncertainty in the

141 posterior distributions by integrating over all trait and biogeographic scenarios found likely

142 by the model. This algorithm has the added advantage of returning joint posterior

143 reconstructions of trait and biogeographic histories, which can be used in post hoc analyses

144 and visualizations. To assess the behavior of our model and to validate our method, we

145 first measure how well it fits a variety of datasets that were simulated under a breadth of

146 evolutionary scenarios. Subsequently, we fit the model to the adaptive radiation of

147 Galápagos finches, an evolutionary system that has been instrumental in exploring

148 phenomena including character displacement, competitive exclusion, and local extirpation

149 due to competition pressure (Lack 1947; Schluter et al. 1985; Grant and Grant 2006).

Although our present work focuses on the reciprocal evolution of continuous-valued ecological traits and discrete-valued ranges within and between lineages, our inference framework is extensible to more general models of co-evolution than studied here.

To our knowledge, this is the first study that models biogeographic history and continuous trait evolution as interdependent with one another. This allows assaying previously untestable hypotheses explaining the biogeographic history of clades at the intersection of evolutionary biology and ecology.

Model

Current approaches for interdependent trait evolution between lineages

Nuismer and Harmon (2015) introduced a continuous trait model where traits of lineages depend on traits of other contemporaneous lineages, allowing biotic interactions among lineages to drive trait divergence and convergence. We follow their derivation of the model, but note that we have modified the notation for the following equations to match analogous parameters in our model. Under the assumption that all lineages are able to interact with each other at any given time (i.e., all are sympatric), weak natural selection and fixed additive genetic variance and population sizes, the change in population mean phenotype for species i is given by the following Stochastic Differential Equation (SDE; Eq. S38 in Nuismer and Harmon 2015)

$$x_i(t + dt) = x_i(t) + \psi(\theta - x_i(t))dt + \omega_x(\mu(t) - x_i(t))dt + \sigma dW_t, \tag{1}$$

where ψ represents the strength of selection, θ the selective optimum, ω_x the strength and directionality of competitive interactions, μ the expected value of mean phenotypes among all species, σ the diffusion rate, and W_t the Wiener process (i.e., standard Brownian motion of Gaussian increments with mean 0 and variance 1). This model couples genetic drift and

172 stabilizing selection (i.e., single-peak Ornstein-Uhlenbeck) with competitive co-evolutionary

173 dynamics; when $\psi = 0$, the model collapses to a random drift with competitive

174 interactions; if, additionally, $\omega_x = 0$, the model becomes a Brownian motion. Lastly, when

175 $\omega_x \leq 0$, species traits are repelled from a shared average; when $\omega_x > 0$, species traits

176 converge to this average.

177 The above model assumes that all species in the phylogenetic tree have been

178 sympatric along their evolutionary history, which is often not the case. Drury et al. (2016)

179 expanded on this competition model to incorporate a sympatry matrix among lineages

180 through time. The sympatry matrix effectively limits any interspecific effects upon trait

181 evolution to only those lineages in sympatry at a given time. To do so, let $A(t)$ represent a

182 time-varying sympatry matrix where entry $\mathbf{A}_{i,j}(t) = 1$ if species i and j are sympatric at

183 time t and 0 otherwise. Then, the change in trait value is given by the following SDE

$$x_i(t + dt) = x_i(t) + \omega_x \left(\left(\frac{\sum_j \mathbf{A}_{i,j}(t) x_j(t)}{\sum_j \mathbf{A}_{i,j}(t)} \right) - x_i(t) \right) dt + \sigma dW_t. \tag{2}$$

184 Note, because of non-identifiability, the model assumes no directional selection ($\psi = 0$;

185 Drury et al. 2016). The likelihood of the parameters of interest, ω_x, σ, and the ancestral

186 state estimate of the MRCA, is a Multivariate Normal density with mean equal to the

187 MRCA state and the scalar product of σ with the resulting variance-covariance matrix

188 (Manceau et al. 2017). Drury et al. (2016) derived the SDEs governing the expected

189 variance-covariance through time, and use numerical integration to solve from the root to

190 the tips.

191 Clarke et al. (2017) proposed a different SDE where species phenotypes are assumed

192 to have normal distributions that phenotypically displace one another in trait space based

193 on their degree of overlap.

$$x_i(t + dt) = x_i(t) + \omega_x \sum_j \mathbf{A}_{i,j}(t) f_{\text{overlap}}(x_i(t), x_j(t)) dt + \sigma dW_t. \tag{3}$$

194 This equation has the advantage of summing over the relative repelling forces from each

195 sympatric species to determine trait evolution instead than just being driven by a

196 community average (Clarke et al. 2017).

197   One concern with these (and similar) approaches is that biogeographic history is

198 inferred separately from trait evolutionary dynamics, and then conditioned upon when

199 estimating a competition effect on trait evolution. Biologically, the distribution of species

200 traits across areas is likely to directly affect dispersal patterns of lineages along their

201 biogeographic history. For example, extirpation rates might increase among competing

202 lineages while in sympatry, and dispersal rates might decrease for lineages attempting to

203 colonize areas occupied by competitors. More subtly, sequential inference schemes that

204 uniformly average over posterior samples often do not properly weigh the probability of

205 each "upstream" sample when aggregating results under the "downstream" model. This

206 forces the support for each upstream sample to be taken as equal under the downstream

207 model even when that is not true, resulting in the incorrect propagation of uncertainty in

208 species ranges – i.e. a range that is unlikely to be sampled under the trait model would be

209 awarded too much support. Jointly modeling trait and range evolution would circumvent

210 both of these issues, as we describe below.

211   *Mutually dependent trait and range evolution model*

212 *Hypotheses framework.*— There are three parameters that regulate the effect of biotic

213 interactions in our model. The magnitude and directionality of these parameters explicitly

214 examine three expected processes in which interspecific biotic interactions shape

215 biogeographic and trait evolution (Figure 1).

216   i. Sympatric competition driving character change is described by ω_x (i.e.,

217   *post*-sympatry effect of biotic interactions on trait evolution). If $\omega_x < 0$ or $\omega_x > 0$,

218   biotic interactions are driving character divergence and convergence, respectively. If

219       $\omega_x = 0$, no effect of biotic interactions is found when in sympatry, and the particular

220   trait follows a random walk.

221   ii. The effect of biotic interactions on successful colonization is regulated by ω_1 (i.e.,

222   *pre*-sympatric effect of biotic interactions). If $\omega_1 < 0$, lineages have lower rates of

223   successful colonization for areas inhabited by similar species, indicative of

224   competitive exclusion. If $\omega_1 > 0$, lineages have higher rates of successful colonization

225   for areas inhabited by similar species, presumably because of environmental filtering.

226   Evidently, if $\omega_1 = 0$, there is no effect of biotic interactions on rates of colonization.

227   iii. Finally, ω_0 describes the effect of biotic interactions on rates of local extinction. If

228       $\omega_0 > 0$, more divergent lineages within an area are less likely to go locally extinct,

229   suggesting that competition pressure drives population extirpation. If $\omega_0 < 0$,

230   phenotypically similar lineages within an area are less likely to go extinct, indicative

231   of environmental filtering. Again, if $\omega_0 = 0$, there is no effect of biotic interactions on

232   local extinction rates.

233   Table 1 summarizes the effect of model parameters upon the evolution of sympatric

234   lineages for reference.

235   Adopting a Bayesian perspective allows one to directly detect the effect of

236   sympatric interactions on trait and range evolution. When the 95% highest posterior

237   density (HPD) does not contain the value $\omega_x = 0$, we reject the hypothesis that traits

238   evolve independently among lineages. Similarly, we interpret HPDs that do not contain

239    $\omega_1 = 0$ or $\omega_0 = 0$ as evidence against colonization and extirpation rates being independent

240   of interspecific effects.

241   *Model details.*— We define a joint probabilistic model where rates of area gain and loss for

242   a species may depend on the trait values of all species present in the determined area, and

243   trait values may depend on the trait values of sympatric species (Figure 1). Given a fixed,

244   fully bifurcating and time-calibrated phylogenetic tree with n extant species, which we

245   assume as the true tree, and observed data at the tips, we model the biogeographic and

246   trait evolution across time. The crown age of the tree occurs at time 0, with time

Parameter	Value	Effect of sympatry
ω_x	0	No effect
	< 0	Traits diverge
	> 0	Traits converge
ω_1	0	No effect
	< 0	Lower colonization rates
	> 0	Higher colonization rates
ω_0	0	No effect
	< 0	Lower extirpation rates
	> 0	Higher extirpation rates

Table 1: Effect of model parameters upon the evolution of sympatric lineages. Trait evolution (ω_x) and extirpation (ω_0) parameters are informed by sympatric differences in traits in the currently inhabited area(s). The colonization parameter (ω_1) is informed by differences in traits between the colonizing lineage and the resident trait distribution in the area to be colonized.

progressing forward until observing the present values at the tips at time T. We denote the entire trait evolutionary history along the phylogenetic tree as X and the entire biogeographic history as Y. As above, let $x_i(t)$ be the trait value, in continuous space, for lineage i at time t. For a set of K discrete areas, $k \in \{1, \ldots, K\}$, let $y_{i,k}(t)$ be 1 if lineage i is present in area k or 0 if it is absent at time t. Thus, the geographic range of lineage i at time t can be represented by the vector $\mathbf{y}_i(t) = \{y_{i,1}(t), \ldots, y_{i,K}(t)\}$. Excluding distributions in which species are absent at all areas (i.e., forbidding lineages from going globally extinct), this yields a biogeographic state space containing $2^K - 1$ possible ranges. We sample n tips at the present, each with trait value, $x_i(T)$, and occurring at a subset of discrete locations, $\mathbf{y}_i(T)$. These observations are the result of trait evolution and of species changing their geographic range either by colonizing (area gain) or going locally extinct (area loss) across time.

We model the effect of competition on the trait evolution of lineage i using the

following SDE

$$
x_i(t+dt) = \begin{cases} x_i(t) + \omega_x \left(\sum_j \mathrm{sgn}(\Delta x_{j,i}(t)) \, \Delta y_{j,i}(t) \, e^{-|\Delta x_{j,i}(t)|} \right) dt + \sigma dW_t & \begin{array}{l} \text{if } \omega_x < 0 \\[4pt] \text{(divergence)} \end{array} \\[20pt] x_i(t) + \sigma dW_t & \begin{array}{l} \text{if } \omega_x = 0 \\[4pt] \text{(no effect)} \end{array} \\[20pt] x_i(t) + \omega_x \left(\sum_j \Delta x_{j,i}(t) \, \Delta y_{j,i}(t) \right) dt + \sigma dW_t & \begin{array}{l} \text{if } \omega_x > 0 \\[4pt] \text{(convergence)} \end{array} \end{cases}
$$

$$(4)$$

259 where $\mathrm{sgn}(x) = \{-1 \text{ if } x < 0, \ 0 \text{ if } x = 0, \text{ and } 1 \text{ if } x > 0\}$, and

$$
\Delta x_{j,i}(t) = x_j(t) - x_i(t),
$$

260 and

$$
\Delta y_{j,i}(t) = \frac{\sum_k y_{j,k}(t) y_{i,k}(t)}{\sum_k y_{i,k}(t)}
$$

261 represent trait and range differences between lineages, respectively. That is, the strength of

262 biotic interactions for the focal lineage i at time t is measured in relation to the weighted

263 sum of trait differences with other species, $\Delta x_{j,i}(t)$, scaled proportionally to the amount of

264 range overlap, $\Delta y_{j,i}(t)$. Figure 2a illustrates the behavior of this SDE. Importantly, it

265 befits the theoretical expectation that competition strength should wither as trait

266 dissimilarity increases. Fortunately, the inference scheme that we use (see below) provides

267 great flexibility in specifying the deterministic part of the SDE, as long as it is a function

268 of the form $x_i(t+dt) = f_x(X(t), Y(t), \omega_x, dt) + \sigma dW_t$.

269   To test the effect of biotic interactions on biogeographic history, we allow for rates

270 of colonization and local extinction for a given lineage i to vary according to the similarity

271 between its phenotype x_i and that amongst all species currently in an area. Specifically, let

272 $\mathbf{u}, \mathbf{v}$ be geographic ranges that differ only on area k, with $u_k = 0$ and $v_k = 1$, and let

273 $\dot{\lambda}_l(i, k, t, \lambda_l, \omega_l, X, Y)$ for $l = \{0, 1\}$ be the instantaneous rates of area gain or loss,

respectively, for area k and lineage i at time t. Then, we define

$$\dot{\lambda}_l(i, k, t, \lambda_l, \omega_l, X, Y) = \begin{cases} \lambda_l \left(1 + e^{-\phi_{i,k}(t)}\right)^{\omega_l} & \text{if } \phi_{i,k} > 0 \\ \lambda_l & \text{if } \phi_{i,k} = 0 \end{cases} \tag{5}$$

where

$$\phi_{i,k}(t) = \min\{y_{1,k}(t)|\Delta x_{1,i}(t)|, \ldots, y_{n,k}(t)|\Delta x_{n,i}(t)|\},$$

λ_l is the "basal" rate of colonization or extinction, ω_l describes the effect of biotic interactions on rates of colonization or extirpation, and $\phi_{i,k}(t)$ is the minimal distance in trait space between lineage i and those in area k.

Equation 5 is a simplified version of the Generalized Logistic function (see Appendix). Note that when ω_1 is negative, these functional forms designate λ_1 as the maximum colonization rate when an area is unoccupied, and the presence of other species induces a penalty on the rates, in turn, when ω_1 is positive colonization rates are enhanced. Similarly, λ_0 is the rate when an area is unoccupied, and sympatric species induce a rate increase with $\omega_0 > 0$ and a decrease when $\omega_0 < 0$. In both cases, the penalty is dependent on the minimum distance between the focal species i and those in the area being considered k (i.e., $\phi_{i,k}$). Thus, the magnitude of ω_1 and ω_0 reflect the relative effect in which biotic interactions affect biogeographic rates (Figure 2b).

A discretized time scheme.— We wish to compute the probability of a single, exact co-evolutionary history of traits and ranges along all branches of a phylogeny. Even for a single trait-range history, we were unable to derive an analytical form of the transition probabilities for trait evolution (Eq. 4) and range evolution (Eq. 5) as functions of continuous time. Thus, following Horvilleur and Lartillot (2014), we represent the continuous-time processes of trait and range evolution in discrete time. This time discretization serves two purposes: first, it lets us derive the discrete-time transition probabilities we need to compute the model probability; and, second, it provides a basis to rapidly query the complete evolutionary state shared across lineages, areas, and traits at

297 regular time intervals, which is essential for computing the transition probabilities.

298   Figures 3a and 3b illustrate an example output of our two-stage discretization

299 procedure, which results in the ordered vector of times, $\boldsymbol{\tau}$. The procedure works as follows.

300 Let $t_0 = 0$ be the crown age of the tree, and let T be the time at which we observe the tip

301 trait values, X_{obs}, and range values, Y_{obs}. Also, let branch b have a start time t_{bs} and end

302 time t_{bf}, such that $t_b = t_{bf} - t_{bs}$. The first stage divides each t_b into $K + 1$ equally spaced

303 time slices (i.e., the number of areas plus one), yielding the vector of sampling times

304 $\boldsymbol{\tau}_b = \{t_{bs} = \tau_{b,1}, \ldots, \tau_{b,K+2}, t_{bf} = \tau_{b,K+3}\}$. Because we only allow one event per time step,

305 the number of slices, $K + 3$, guarantees that lineage i has more than the minimum number

306 of steps possibly needed to evolve from range $\mathbf{y}_i(b_s)$ to $\mathbf{y}_i(b_f)$ in the case where $\mathbf{y}_i(b_s)$ is

307 absolutely different from $\mathbf{y}_i(b_f)$ (for example, it would take at least three events for the

308 range $\mathbf{y}_i(b_s) = \{0, 0, 1\}$ to evolve into range $\mathbf{y}_i(b_f) = \{1, 1, 0\}$). The second stage sets a

309 minimum time step allowed in the analyses, δt_{min}, and proceeds forwards in time to

310 subdivide the remaining periods such that no time step is larger than δt_{min}. In practice, we

311 standardize δt_{min} using the percentage of the tree height for comparability. This procedure

312 results in a sorted vector of sampling times $\boldsymbol{\tau} = \{t_0 = 0, \ldots, T\}$ that are shared among all

313 contemporaneous lineages throughout the clade's history. For each branch b with sampling

314 times $\boldsymbol{\tau}_b \subseteq \boldsymbol{\tau}$, we end up with a time ordered set describing the trait evolution of the

315 lineage, $X_b = \{x_i(\tau_{b,1}), \ldots, x_i(\tau_{b,|\boldsymbol{\tau}_b|})\}$. Likewise, for each branch b, we record an ordered set

316 of vectors describing the biogeographic history of the lineage, $Y_b = \{\mathbf{y}_i(\tau_{b,1}), \ldots, \mathbf{y}_i(\tau_{b,|\boldsymbol{\tau}_b|})\}$.

Likelihood calculation.— We are not aware of an analytical form for the transition

probabilities corresponding to the range-dependent trait evolution model (Eq. 4), so we

approximate the likelihood using the Euler-Maruyama method (see Appendix). The

likelihood for trait evolution for branch b is then

$$L(X_b; \sigma, \omega_x, X, Y) =$$
$$\prod_{j=1}^{|\boldsymbol{\tau}_b|-1} \frac{1}{\sigma\sqrt{2\pi\delta t_j}} \exp\left\{-\frac{(x_i(\tau_{b,j+1}) - f_x(X(\tau_{b,j}), Y(\tau_{b,j}), \omega_x, \delta t_j))^2}{2\sigma^2\delta t_j}\right\}, \quad (6)$$

317 where $\delta t_j = \tau_{b,j+1} - \tau_{b,j}$.

The likelihood for the biogeographic history in discrete time can be deconstructed into a series of events and nonevents within small windows of time. An event is defined as either an area colonization or loss, and a nonevent as no change in state. Let $l = \{0, 1\}$, then the likelihood after some time δt for area k is

$$L(\mathbf{y}_i(t) \rightarrow \mathbf{y}_i(t + \delta t); \omega_1, \omega_0, \lambda_1, \lambda_0, \delta t, X, Y) =$$
$$\begin{cases} \exp(-\dot{\lambda}_l(\cdot)\delta t) & \text{if } y_{i,k}(t) = y_{i,k}(t + \delta t) \\ \dot{\lambda}_l(\cdot)\exp(-(\dot{\lambda}_l(\cdot))\delta t) & \text{if } y_{i,k}(t) \neq y_{i,k}(t + \delta t) \end{cases}.$$

318 Then, the likelihood for branch b across all areas is:

$$L(Y_b; \omega_1, \omega_0, \lambda_1, \lambda_0, X, Y) = \prod_{j=1}^{|\boldsymbol{\tau}_b|-1} L(\mathbf{y}_i(\tau_{b,j}) \rightarrow \mathbf{y}_i(\tau_{b,j+1}); \omega_1, \omega_0, \lambda_1, \lambda_0, \delta t_j, X, Y), \quad (7)$$

319 where $\delta t_j = \tau_{b,j+1} - \tau_{b,j}$.

320 The prior probabilities for each state are usually set to the stationary frequencies

321 given by the dispersal rates λ_1 and λ_0. We could not derive an analytical solution for these

322 frequencies, so we add a long branch (twice the tree height by default) at the root and

323 simulate geographic range evolution to approximate geographic range frequencies at the

324 root (Landis et al. 2013). Under the model assumptions, there is no trait data for the stem

325 branch (and, given the tree, there is no competition since only one lineage of the clade is

326 alive), so the likelihood computation can be done in continuous time. Let

327 $L(Y_{\text{root}}; \lambda_1, \lambda_0, \mathcal{M}_c)$ denote this likelihood and $\boldsymbol{\theta} = \{\sigma, \omega_x, \omega_1, \omega_0, \lambda_1, \lambda_0\}$. Then, by

328 incorporating the trait evolution likelihood and multiplying across all branches, we get the

329 following joint likelihood:

$$L(X,Y;\boldsymbol{\theta},\mathcal{M}_c) = L(Y_{\text{root}};\lambda_1,\lambda_0,\mathcal{M}_c) \prod_{b=1}^{2n-2} L(X_b;\sigma,\omega_x,X,Y,\mathcal{M}_c)L(Y_b;\omega_1,\omega_0,\lambda_1,\lambda_0,X,Y,\mathcal{M}_c),$$
(8)

330 where $\mathcal{M}_c$ is the model incorporating biotic interactions.

331 *Collision probability.*— It is possible that a species range gains and then loses an area (or

332 vice versa) so rapidly under the idealized continuous-time model that those events would

333 go undetected by our discrete-time model. Such "collisions" of events within a single

334 discrete time bin might lead to underestimating the area colonization and loss rates. We

335 estimate an upper bound on the collision probability, $\mathbb{P}_c$ that two or more range evolution

336 events occur within a fixed δt, such that our sampling would not detect them. Specifically,

337 let δt be a time interval for which we sample Y and X at the beginning, t_s, and at the end,

338 t_f, where $t_f = t_s + \delta t$. If the lineage is present in area k at time t_s, the lineage could lose

339 this area and regain it before we are able to register such event in t_f. Let $r = (\lambda_1 + \lambda_0)\delta t$,

340 then the probability that two or more events at times occur within δt is

$$\mathbb{P}_c = \mathbb{P}(\text{two or more events} < \delta t)$$
$$= 1 - \mathbb{P}(0 \text{ events in } \delta t) - \mathbb{P}(1 \text{ event in } \delta t)$$
$$= 1 - \frac{r^0 e^{-r}}{0!} - \frac{r^1 e^{-r}}{1!}$$
$$= 1 - e^{-r}(1+r).$$

341 We consider δt to be the largest interval in the analysis, thus providing a somewhat

342 conservative measure of collision probability (given that there are smaller intervals

343 following our discretization procedure). However, since the actual rates rely on the specific

344 interaction between trait value differences and ω_1 and ω_0, this measure does not necessarily

345 reflect the actual collision probability, yet it still is a source of objective information on

346 amount of approximation error. We monitor $\mathbb{P}_c$ during inference to provide a measure of

347 error given the particular parameters and defined δt.

Markov Chain Monte Carlo with data augmentation

349 The main impediment when inferring under such a joint model is the mutual

350 dependence of the trait evolutionary history, X, and the biogeographic history, Y. At any

351 given time, trait evolution for one species depends on the traits of those species it is

352 sympatric with, and the set of species that are able to coexist in sympatry is contingent on

353 the concurrent trait distribution. This, in part, renders common inference procedures such

354 as the derivation of an analytic solution or pseudo-exact likelihood by numerical integration

355 of SDEs infeasible. Rather than analytically integrating over all possible evolutionary

356 histories, we use data augmentation (DA) to numerically sample over those histories

357 (Robinson et al. 2003; Landis et al. 2013). Under DA, one repeatedly simulates otherwise

358 unobservable data to evaluate the probability of the parameters $\boldsymbol{\theta}$ under both the observed

359 data D_{obs} and the augmented data D_{aug}. Among several advantages of using DA is the fact

360 that, for certain problems, simpler and more efficient likelihood functions exist when

361 augmented data is generated. By repeatedly proposing different realizations of D_{aug} across

362 the MCMC, one numerically averages over the augmented data to obtain the joint posterior

363 of evolutionary histories and model parameters, $p(\boldsymbol{\theta}, D_{\mathrm{aug}} \mid D_{\mathrm{obs}}, \mathcal{M})$. In particular, we are

364 interested in computing the posterior probability of all the parameters given the observed

365 data. The posterior probability of one single biogeographic and trait history is

$$p(\boldsymbol{\theta}, X_{\mathrm{aug}}, Y_{\mathrm{aug}} \mid X_{\mathrm{obs}}, Y_{\mathrm{obs}}) \propto L(X_{\mathrm{obs}}, Y_{\mathrm{obs}}, X_{\mathrm{aug}}, Y_{\mathrm{aug}}; \boldsymbol{\theta})\, \pi(\boldsymbol{\theta}),$$

366 where π is the prior distribution of $\boldsymbol{\theta}$. We describe the initialization procedure for X_{aug} and

367 Y_{aug} in the Appendix. Figure 3c,d shows a sample from the marginal posterior for DA trait

368 and biogeographic histories from a simple simulation. We sample augmented evolutionary

369 histories and evolutionary parameters using the Metropolis-Hastings algorithm (Metropolis

370 et al. 1953; Hastings 1970).

371 *Parameter, trait history, and range history proposals.*— Standard slide and scale moves are

372 used to proposed new parameter values for σ, λ_0, λ_1, ω_0, ω_1, and ω_x (see Appendix).

373 We generate proposals for the trait history, X_{aug}, by adding a Gaussian deviation to

374 a uniformly sampled $x_i(t)$, such that $x_i(t)' = x_i(t) + \mathrm{N}(0, s)$, where s represents the tuning

375 parameter. The acceptance ratio for this proposal is

$$\alpha = \min\left\{1, \frac{L(X'_{\mathrm{aug}}, X_{\mathrm{obs}}, Y; \boldsymbol{\theta}, \mathcal{M}_c)}{L(X_{\mathrm{aug}}, X_{\mathrm{obs}}, Y; \boldsymbol{\theta}, \mathcal{M}_c)}\right\}.$$

376 In addition, we generate less conservative updates by proposing branch-wide updates for

377 X_{aug}. We use random samples from an independent distribution for σ^* to generate

378 Brownian bridges for branches in the tree (details for generating a Brownian bridge are

379 given in the Appendix). First, we sample $\sigma^* \sim \mathrm{Lognormal}(0, 1)$ and a branch uniformly

380 and generate a Brownian bridge holding the end nodes constant. Similarly, following

381 Horvilleur and Lartillot (2014), we sample an internal node uniformly and generate a new

382 node state under Brownian motion and generate Brownian bridges for the three adjoining

383 branches. The acceptance ratio for the these proposals is

$$\alpha = \min\left\{1, \frac{L(X'_{\mathrm{aug}}, X_{\mathrm{obs}}, Y; \boldsymbol{\theta}, \mathcal{M}_c)}{L(X_{\mathrm{aug}}, X_{\mathrm{obs}}, Y; \boldsymbol{\theta}, \mathcal{M}_c)} \frac{L(X_{\mathrm{aug}}, X_{\mathrm{obs}}; \sigma^*, \mathcal{M}_{BM})}{L(X'_{\mathrm{aug}}, X_{\mathrm{obs}}; \sigma^*, \mathcal{M}_{BM})}\right\},$$

384 where $\mathcal{M}_{BM}$ denotes the Brownian Motion model.

385 To update the range history, Y_{aug}, we select an internal node uniformly at random,

386 including the root, and sample a new geographic range from the joint density under the

387 mutual-independence model, $\mathcal{M}_0$. We use random samples from an independent

388 distribution for λ_1^* and λ_0^* to generate DA biogeographic histories under $\mathcal{M}_0$. We improve

389 efficiency and acceptance rates of biogeographic histories by disallowing colonization and

390 extirpation rates to be too dissimilar. Therefore, we randomly sample $\upsilon \sim \mathrm{Lognormal}(0, 1)$,

391 and then multiply υ by a Lognormal distribution with expectation of 1 and low variance

392 such that $\frac{\lambda_l^*}{\upsilon} \sim \mathrm{Lognormal}(-0.044, 0.3)$ for $l \in \{0, 1\}$. Using the rejection sampling

393 described in Landis et al. (2013) and the Appendix, we then sample new biogeographic

394 histories along the three adjoining branches such that they are consistent with the new

sampled geographic range at the node and those at the end nodes. The simplified

Metropolis-Hastings acceptance ratio (α) for this proposal is

$$\alpha = \min\left\{1, \frac{L(Y'_{\text{aug}}, Y_{\text{obs}}, X; \boldsymbol{\theta}, \mathcal{M}_c)}{L(Y_{\text{aug}}, Y_{\text{obs}}, X; \boldsymbol{\theta}, \mathcal{M}_c)} \frac{L(Y_{\text{aug}}, Y_{\text{obs}}; \lambda_1^*, \lambda_0^*, \mathcal{M}_0)}{L(Y'_{\text{aug}}, Y_{\text{obs}}; \lambda_1^*, \lambda_0^*, \mathcal{M}_0)}\right\},$$

where the first term is the ratio between the likelihoods of the proposed and current

biogeographic histories under the full model, $\mathcal{M}_c$, and the second term is the proposal

density ratio under the mutual-independence model, $\mathcal{M}_0$. Additionally, we perform more

moderate proposals for range evolution by mapping biogeographic histories on a branch

sampled at random, leaving the end nodes constant. The acceptance ratio for this branch

update is the same as for the node update. As mentioned earlier, daughter lineages inherit

the same geographic range as their parent lineage at speciation times. This mimics a very

particular case of sympatric speciation, a strong assumption for the biogeographic history

of some clades. The intricacies of geographical speciation will be left for future work (e.g.,

Ree et al. 2005; Matzke 2014).

Finally, to better explore parameter space, we make joint X_{aug} and Y_{aug} proposal

updates. For the first joint update, we uniformly sample a branch and update the trait

history using a Brownian bridge proposal and update biogeographic history using

stochastic mapping as described above. Secondly, we uniformly sample an internal node

and generate a joint proposal for the node and the the three adjoining branches. The

acceptance ratio for these proposals is

$$\alpha = \min\left\{1, \frac{L(Y'_{\text{aug}}, X'_{\text{aug}}, Y_{\text{obs}}, X_{\text{obs}}; \boldsymbol{\theta}, \mathcal{M}_c)}{L(Y_{\text{aug}}, X_{\text{aug}}, Y_{\text{obs}}, X_{\text{obs}}; \boldsymbol{\theta}, \mathcal{M}_c)} \frac{L(Y_{\text{aug}}, Y_{\text{obs}}; \lambda_1^*, \lambda_0^*, \mathcal{M}_0)}{L(Y'_{\text{aug}}, Y_{\text{obs}}; \lambda_1^*, \lambda_0^*, \mathcal{M}_0)} \frac{L(X_{\text{aug}}, X_{\text{obs}}; \sigma^*, \mathcal{M}_{BM})}{L(X'_{\text{aug}}, X_{\text{obs}}; \sigma^*, \mathcal{M}_{BM})}\right\}.$$

Software.— We denote this model as "TRIBE" (which stands for "Trait and Range

Interspecific Biogeographic Evolution") and implement it in a new open source package

named "Tapestree" (`https://github.com/ignacioq/Tapestree.jl`) that we wrote in

Julia (Bezanson et al. 2017). This software makes available the `tribe()` function for

inference and the `simulate_tribe()` for simulations given a fixed tree. We note that, in

418 the software, we allow the user to fix to 0 any or all of the parameters governing the effect

419 of biotic interactions (i.e., ω_x, ω_0, & ω_1).

420 ## *Simulations*

421 We use simulations to explore model behavior. To simulate biogeographic histories

422 under this model, we take advantage of the following approximation. Let V be a random

423 variable denoting the time of an event and $\lambda(t)$ be the event rate at time t, then given a

424 small enough time step δt, we have

$$\mathbb{P}(t \leq V < t + \delta t | t \leq V) \approx \lambda(t)\delta t.$$

425 Thus for a given lineage and timepoint, we use the above time step size for all areas across

426 the geographic range. If there is more than one event within one time step as defined by

427 our time discretization scheme, we reject and sample again. Similarly, to simulate trait

428 evolution under the competition model, we, again, take advantage of the Euler-Murayama

429 method detailed in the Appendix. Simulation code, given a phylogenetic tree, can be found

430 at https://github.com/ignacioq/Tapestree.jl.

431 We simulated phylogenetic trees using a pure-birth process until reaching 25 species

432 and set the MRCA trait value to 0 and the number of areas to 12. Given the relatively

433 large parameter space, we used the same values for λ_1, λ_0 and σ^2 across all simulations,

434 and explored different combinations of the parameters regulating the biotic interactions. In

435 particular we simulated 10 different scenarios with $\lambda_1 = 1.0$, $\lambda_0 = 0.4$ and $\sigma^2 = 0.16$, and

436 the following combinations of $(\omega_x, \omega_1, \omega_0)$: $(0, 0, 0)$, $(-2, -2, 0)$, $(-2, 2, 2)$, $(-2, 0, 0)$,

437 $(2, 0, 0)$, $(2, -2, 2)$, $(2, 0, 0)$, $(0, -2, 2)$, $(0, -2, 0)$, and $(0, 0, 2)$. Each scenario was simulated

438 100 times to yield a total of 1000 simulations. While not exhaustive, these simulations

439 allow us to test the power and bias of our model with regard to each of these three

440 parameters. Further exploration of parameter space is encouraged for the future.

441 We ran MCMC inference on each simulation for $100,000$ iterations, logging every

442 100th iteration, discarding the first $50,000$ samples obtained during the adaptive burn-in

443 phase. We note that each iteration corresponds to $> 55,000$ parameter updates (the user

444 can adjust the weights for each parameter). We used ambiguous priors for all parameters,

445 specifically, we used a normal prior of mean 0 and standard deviation of 10 for ω_x, ω_1, and

446 ω_0, and an exponential prior of mean 10 for σ^2, λ_1 and λ_0. Most of the effective sample

447 sizes (ESS) for all parameter in each simulation were > 300, but in a few cases σ^2 or ω_x

448 had lower ESS; we made sure that the ESS for each parameter was at least > 150.

449 We perform statistical evaluation using highest probability density intervals (HPD)

450 for all the parameters. Overall, our model is able to recover most of the simulated

451 parameter values and associated uncertainty. The posterior median estimates reflect the

452 simulated values (Figure 4) and 95% coverage probability based on HPD for parameters

453 reflecting biotic interactions are over 0.90 for most scenarios (Figure 5). Most importantly,

454 our model is able to reliably discern when there is no effect of biotic interactions for ω_x and

455 ω_0 (Figures 4 & 5). Estimates of the posterior mean of ω_x behave without bias when the

456 true value is negative, yet the have a marginally positive bias towards more positive values

457 when it is ≥ 0; this is most likely because of an increase in skew in the posterior distribution

458 as ω_x increases. The 95% HPD coverage is close to 0.95 for all scenarios (Figure 5).

459 Nonetheless, we find a minor bias in ω_1, the parameter regulating competition on

460 colonization rates. Recovered values for ω_1 are biased toward lower values, however, the

461 coverage remains at least 90% for scenarios for scenarios with $\omega_1 = 0$, yielding acceptable

462 false positive rates for competition (Figures 4 & 5). We find the greatest bias and lowest

463 coverage for scenarios in which $\omega_1 > 0$, and may result in false negatives for facilitation in

464 colonization rates. Finally, we find that posterior estimates of λ_1, λ_0 are somewhat

465 underestimated, and their medians are usually lower that the simulated value. While

466 concerning, this is likely due to the interaction with the phenotypic traits and does not

467 preclude our ability to make inference on the effect of biotic interactions on biogeographic

468 and phenotypic evolution. Overall, most likely increasing the number of taxa and areas will

469 result in higher power.

470 *Impact of δt_{min}.—* To evaluate the impact of different $\delta t_{\min}$ in parameter estimates, we

471 performed inference on the same data with five different $\delta t_{\min} = \{0.99, 0.2, 0.1, 0.01, 0.005\}$.

472 Note that it is often the case that increasing values of $\delta t_{\min}$ to be greater than *ca.* 0.2 gives

473 the same discretization scheme and thereby similar results because our discretization

474 procedure minimally includes times for the start, end, and $K + 1$ intermediate time points

475 along every branch in the tree (clearly, this threshold is relative to the structure of the

476 tree). The simulations were conducted with the same pure-birth tree of 25 species and 4

477 areas, and the following parameter values: $\omega_x = -2$, $\omega_1 = 1$, $\omega_0 = -1$, $\sigma = 0.8$, $\lambda_1 = 4$ and

478 $\lambda_0 = 2$. We ran the analysis with an adaptive burn-in of 50000 iterations and a sampling

479 chain of 100000.

480 We find that the impact of $\delta t_{\min}$ has minor consequences on the parameter

481 estimates in the posterior distributions (Supplementary Figure 1). This is most likely due

482 to the discretization procedure that ensures that each branch will be subdivided into at a

483 number of units greater (by one or more) than the number of areas. Such discretization is

484 thus finer towards the tips, where more branches overlap in time, and where inference is

485 less uncertain (since is more proximate in time to the observed trait and biogeographic

486 data). We find σ^2, ω_1 and ω_0 to be marginally affected by the the choice of $\delta t_{\min}$. The

487 differences are slightly pronounced in ω_x, λ_1, and λ_0, particularly in terms of precision.

488 This is expected as we reiterate that we are approximating the likelihoods, and a finer

489 discretization will be less biased. For instance, a finer discretization allows higher rates of

490 colonization and extinction to be sampled in the posterior (Supplementary Figure 1).

491 Larger δt values between sampling times incur in high collision probabilities, thus ignoring

492 high rates of state changes and setting an upper limit on the inference of rates of state

493 change. Given our simulation results and required computational efficiency, we suggest

494 that a $\delta t_{\min} = 0.01$ yields an acceptable representation of the model likelihood.

495 *Empirical application: Darwin's finches in the Galápagos*

Trait	Posterior median and HPD estimates					
	ω_x		ω_1		ω_0	
Beak size (PC1)	-0.46	[-1.30, 0.54]	-6.38	[-9.89, -2.80]	1.53	[-0.74, 3.90]
Beak shape (PC2)	1.28	[0.12, 3.57]	-4.28	[-8.63, -0.46]	2.2	[-0.16, 4.90]
Tarsus length	2.41	[-0.04, 6.46]	-4.60	[-6.20, -1.50]	2.87	[-0.26, 5.08]
Wing length	0.61	[-0.10, 5.30]	-4.86	[-6.86, -1.23]	0.43	[-0.99, 3.50]

Table 2: Posterior estimates for Darwin's finches analysis.

We use our model to study how biotic interactions have shaped the biogeographic and trait evolution of Darwin's finches on the Galápagos islands (Grant 1999). We used the species phylogenetic tree from (Lamichhaney et al. 2015) for 14 species and obtained corresponding breeding distributions across the major Galápagos islands (19 islands, including Cocos island), following Table 1.2 in Grant and Grant (2011) and phenotypic measurements from Clarke et al. (2017), originally compiled in Harmon et al. (2010) from which we obtained the data for *Certhidea olivacea*. Specifically, we used three beak measurements: length (culmen), width and depth (gonys), and tarsus and wing length, all with natural logarithmic transformations. Given the high correlation between the three beak measurements, we used the first and second principal components (which together explained $> 99.6\%$ of the variance). The first component mostly corresponds to size, while the second corresponds to overall shape (Supplementary Figure 2; Grant and Grant 2002). The finch data used in this study can be found in the Supplementary Table 1. We ran separate models for these four trait values, for 500 thousand iterations with an adaptive burn-in phase of 50 thousand.

We find that *in situ* trait evolution behaves very differently across the four traits studied here (Figure 6). Overall, we do detect a signal of competitive exclusion ($\omega_1 < 0$), with varied levels of strength. Beak morphometrics (the first and second PCA components relating to size and shape, respectively) display different results (Figure 6e). Beak size

515 shows divergence in sympatry (median $\omega_x = -0.46$, 95% HPD $= [-1.3, 0.54]$); on the other

516 hand beak shape shows convergence (median $\omega_x = 1.28$, 95% HPD $= [0.12, 3.57]$). These

517 traits display values of $\omega_1 < 0$, suggesting strong signals of competitive exclusion,

518 particularly for beak size (median for size $= -6.38$ $[-9.89, -2.8]$; median for shape

519 $= -4.28$ $[-8.63, -0.46]$). Finally we find a weak effect of biotic interactions on the

520 influence of beak size and shape on local extirpation (median ω_0 for size

521 $= 1.53$, $[-0.74, 3.9]$, for shape $= 2.2$, $[-0.16, 4.9]$).

522   Figure 6b focuses on just two finch species that share similar beak sizes at one

523 moment time (present-day), but do not overlap on their geographic distributions.

524 Evidently, *Certhidea fusca* is expected to suffer from lower colonization rates into areas

525 that are occupied by *C. olivacea*, with reciprocal effect for the latter species attempting to

526 colonize areas occupied by the former. This example highlights how our approach may

527 identify whether the allopatric (or sympatric) distribution between species is a product of

528 biotic interactions or independent of them.

529   We find a signal of *in situ* convergence for tarsus length (Figure 6e), (median $\omega_x =$

530 2.41, 95% HPD $= [-0.04, 6.46]$). We observe a strong effect of competition in colonization

531 rates (median $\omega_1 = -4.6$, 95% HPD $= [-6.2, -1.5]$) and no effect of biotic interactions on

532 extirpation rates (median $\omega_0 = 2.87$, 95% HPD $= [-0.26, 5.08]$). Biotic interactions had no

533 effect for wing length when in sympatry (median $\omega_x = 0.61$, 95% HPD $= [-0.1, 5.3]$), but

534 instead find strong competitive exclusion (median $\omega_1 = -4.86$, 95% HPD $= [-6.86, -1.23]$).

535 We find no evidence for an effect of competition in driving local extinction (median $\omega_0 =$

536 0.43, 95% HPD $= [-0.99, 3.5]$). Together, these results suggest that there is strong

537 evidence for competitive exclusion in Darwin's finches in beak morphology and,

538 particularly, in wing length (Figure 6e). Overall, beak size seems to reflect a key

539 competition axis that has driven trait divergence and shaped biogeographic history.

DISCUSSION

540

541   Ever since Darwin (1859), biologists have strived to understand the extent and

542 generality of different biological processes in driving current patterns of diversity (Simpson

543 1953; Mayr 1970; Schluter 2000). Building on previous developments, we introduce a

544 simple but extensible model that integrates discrete biogeographic processes with

545 continuous phenotypic evolution, enabling direct tests on those processes underlying trait

546 evolution, biogeographic history, biotic interactions and community assembly.

547 *Darwin's finches.*— We show how biotic interactions influence trait and biogeographic

548 evolution using the radiation of Darwin's Finches and find that competition has played a

549 role in beak size divergence when different species come into sympatry (Figure 6). While

550 this is in accordance with previous findings (e.g., Lack 1947; Grant and Grant 2006;

551 Clarke et al. 2017), we only find evidence for trait divergence in beak size but not in shape.

552 Instead, our results suggest that bill shape and tarsus length have converged among

553 coexisting species. Presumably, the harsh and unpredictable environmental conditions in

554 the archipelago give rise to strong selection against variants (Price et al. 1984), leading to

555 long term morphological convergence in some traits across the different islands. Indeed,

556 character displacement presupposes that there exists niche space to be displaced into, but

557 extreme events such as droughts severely reduce the number of available sources within an

558 area (Grant and Grant 2011), removing accumulated trait variance. Thus, our results

559 suggest that there is character displacement in beak size but other traits might be

560 phenotypically constrained given the available environment. Future model enhancements

561 could incorporate environmental information to distinguish biotic from abiotic effects.

562 Similarly, persistent introgression during the clade's evolutionary history could lead to

563 some the observed convergence in morphology (Grant et al. 2004; Grant and Grant 2008;

564 Farrington et al. 2014; Lamichhaney et al. 2018).

565   Notably, by allowing trait-mediated biotic interactions to directly influence

566 biogeographic evolution, we are able to recover evidence for competitive exclusion during

567 the radiation of Darwin's Finches (Figure 6). That is, niche dissimilarity facilitated the

568 colonization of new areas during the finch radiation. We observe that all four traits shaped

569 the rates of colonization, to different extents, among the different islands in the

570 archipelago. This is in accordance to theoretical and other empirical evidence suggesting

571 that coexistence can only be tenable with some degree of niche divergence (Elton 1946;

572 Hardin 1960; Macarthur and Levins 1967; Diamond 1978; Godoy et al. 2014). Furthermore,

573 since successful colonization is a necessary step to increase an area's biodiversity, our

574 results hint at the mechanism in which microevolutionary processes might lead up to

575 macroevolutionary patterns, such as the generation of spatial variation in richness. More

576 detailed inspection of per lineage-area effects of biotic interactions during the clade's

577 evolutionary history allows us to disentangle between biogeographic events that involved

578 biotic interactions against those that did not (Figure 6).

579 *Inferring trait-range histories.*— The development of phylogenetic models has allowed

580 researchers to reconstruct historical processes, even when restricted to only extant

581 information, and to test central hypotheses regarding the tempo and mode of evolutionary

582 dynamics (Garamszegi 2014). Such models are valuable, in part, because they require

583 hypotheses about the mode by which lineages evolve and diversify (e.g., Butler and King

584 2004) to be defined in formal terms (e.g., in an SDE). Understanding what features are and

585 are not formally modeled determines what one may prudently conclude from analyses

586 under the method, which we aim to make explicit below. While our model entails several

587 simplifying assumptions, future work may relax these assumptions to incorporate

588 additional features important to modeling trait and range co-evolution.

589 The simple biogeographic model used here assumes that at the moment of

590 speciation the daughter lineages inherit identical ranges as their ancestor lineage, a

591 particular case of sympatric speciation. Given that the great majority of speciation events

592 involve a phase of geographical isolation (Mayr 1970; Rundell and Price 2009), we

593 acknowledge that this assumption does not hold in most empirical systems. Importantly,

594 by not allowing allopatric cladogenesis *sensu* Ree et al. (2005), the inferred parameters

595 governing biotic interactions can be equivocal on a clade with a history of allopatric

596 speciation. For instance, the effect of competitive exclusion (ω_1) is presumed to be large

597 between recently diverged species, yet, these are forced to coexist instantly after speciation,

598 probably underestimating the effect of similarity in colonization rates (e.g., secondary

599 contact times) by overestimating the period of sympatry and bearing upon *in situ* biotic

600 interactions (ω_x) to explain the trait variance. Consequently, an important next step is to

601 incorporate models that allow for different modes of geographical speciation, such as the

602 Dispersal-Extinction-Cladogenesis (DEC) model and relatives (Ree et al. 2008; Matzke

603 2014). This requires designing efficient data augmentation proposals and their associated

604 Metropolis-Hastings ratios, which we are currently working to solve. Other relevant

605 biogeographic processes are not being considered, but are relatively straightforward to

606 incorporate in future versions of the model. Characteristics of the delimited geographical

607 regions, such as distance from each other (Landis et al. 2013), geographical area

608 (Tagliacollo et al. 2015), connectivity (Kadmon and Allouche 2007), age of area availability

609 (e.g., on volcanic islands, Landis et al. 2018), and resource availability (Tilman 1985) will

610 provide key information when inferring biotic interactions. Furthermore, incorporating

611 abiotic optima, as determined by the different regional environments, could be used to

612 distinguish abiotic from biotic forces acting upon trait and range evolution. Research in

613 these directions would further demonstrate the potential of inferring trait and

614 biogeographic evolution as interacting processes (Sukumaran and Knowles 2018).

615   Assuming that interspecific competition acts upon only a single axis of niche

616 evolution, as we assume, may be problematic (Connell 1980). Species niches are better

617 thought of as multidimensional hypervolumes (Hutchinson 1957), and so viewing this

618 complexity through a single, univariate trait must misrepresent the true nature of biotic

619 interactions between species (Diamond 1978; Grether et al. 2009). In some cases, fitting

620 the model separately to each trait or asserting independence on the traits by multivariate

621 transformations (such as PCA) can unduly influence parameter estimates (Uyeda et al.

2015; Cadena et al. 2018). For example, a lack of evidence for biotic interactions within a given axis does not rule out competition from occurring along other unmeasured resource utilization axes (Connell 1980). We advise the researcher to select a trait of study that has been suggested as relevant to niche partitioning (e.g., bill size and shape in the Darwin's finches; Grant and Grant 2002). Measuring species niche overlap between partitions, however, is a general problem pervasive across ecology (Diamond 1978; Petraitis 1979). Species usually occupy ranges of values along niche axes (e.g., the range of temperature where the species can persist) or have considerable intraspecific variation; these features warrant modeling in future methods (e.g., as in Quintero et al. 2015). Moreover, niche similarity might differ between univariate and multivariate spaces, and improved phylogenetic models of competition should account for the multivariate distances between value ranges in niches (Huelsenbeck and Rannala 2003). Despite complications in identifying and representing which traits may be involved in competition, competitive forces are thought to be stronger among recently diverged species because of their overall similarity in resource use (Darwin 1859). Likewise, we assume that biotic interactions have had the same directionality and magnitude (relative to phenotypic dissimilarity) across all lineages throughout the clade's evolutionary history, even though the magnitude and sign of competitive effects probably varies within and between clades, contingent on measured, unmeasured, and unknown factors. While our current model tests for the constant effect of a clade-wide competitive process influencing a univariate trait, it may be extended to accommodate multivariate traits, trait value ranges, and branch-heterogeneous competitive effects.

Finally, our model assumes that biotic interactions only occur between lineages modeled by the phylogenetic tree, which we take to be the reconstructed tree—a tree that only represents lineages corresponding to the set of most recent common ancestors shared among the sampled taxa. Modeling competition while naively taking the reconstructed tree to represent the true evolutionary history among all lineages overlooks any historical

649 contribution from lineages left absent in the reconstructed tree, namely absent lineages

650 representing the ancestors of excluded, unsequenced, or extinct lineages. While, in

651 principle, we can improve representation among extant lineages, that is not always the case

652 with extinct lineages, yet disregarding the influence of extinct lineages is known to mislead

653 some evolutionary inferences (Schindel and Gould 1977; Slater et al. 2012). Being blind to

654 paleobiological interactions may be particularly troublesome in our case, since the

655 geographic and phenotypic evolution of any one ancestral lineage should depend on that of

656 all other contemporaneous lineages, independent on their survival to the present. Provided

657 the data are available, spatial and morphological information from paleontology could be

658 incorporated in our model to attain more biological realism and broaden applicability to

659 clades were extinction rates have been presumed to be high (Mitchell 2015). Correctly

660 modeling the influence of competitive effects with extinct or unsampled ghost lineages that

661 are not represented in the model will require the the introduction of features from

662 birth-death processes.

663   At first glance, developing such a model appears mathematically and

664 methodologically challenging, but progress here would be rewarding. Modeling interactions

665 between trait evolution, competition, biogeography, and diversification processes in a

666 phylogenetic context would represent a major advance towards how we understand the

667 generation and maintenance of biodiversity. As phylogenetic models of competition

668 continue to mature, we must strive to incorporate trait-diversification dynamics that are

669 thought to underlie well-studied macroevolutionary phenomena, such as the Great

670 American Biotic Interchange (GABI; Simpson 1950; Benton 1987). The biogeographic

671 exchange of lineages during GABI is considered to be the result of competition between

672 distantly related clades (Diamond 1978), and classic macroevolutionary hypotheses, such as

673 the "Red Queen" (Van Valen 1973), suggest that temporal and spatial turnover in taxa

674 results mostly from biotic interactions.

675 *Bayesian data augmentation.*— In our work, we provide a framework to test the effect of

ecological processes on phenotypic and biogeographical distribution of species across evolutionary time. The Bayesian data augmentation framework we present here is robust yet flexible, making it adaptable to similar inference problems of associated discrete and continuous character co-evolution. For instance, similar models were developed for processes of correlated nucleotide substitution rates and Brownian motion evolution (Lartillot and Poujol 2011; Horvilleur and Lartillot 2014; Lartillot et al. 2016), and it is conceivable that nucleotide substitution patterns should in some way reciprocally influence how molecular phenotypic traits, such as protein function, evolves (Robinson et al. 2003; Rodrigue et al. 2006). We hope that our algorithmic framework encourages and allows other researchers to develop phylogenetic models that study the interdependent effects of continuous and discrete trait evolution within and between lineages.

Conclusion

Given the ubiquity of character displacement, it might be tempting to assume that phenotypic divergence is the direct result of natural selection acting to avoid competition on sympatric populations (Grant 1972). But it is also plausible that those populations were only able to spread into sympatry because their niche was sufficiently different in the first place (Schluter and McPhail 1992). Lack (1954) pointedly outlined this difference over half a century ago when discussing a case of the bird genus *Sitta*: "...the two species show no overlap in beak measurements [where they occur in sympatry], a difference presumably evolved through the need for avoiding competition for food; or rather, it is only where such a difference has been evolved that the two forms can live alongside each other". Jointly examining distinct mechanisms in trait and biogeographic evolution allows testing core evolutionary theories on how biodiversity is brought about. Clearly, the process by which species diversify phenotypically and attain coexistence is fundamentally important to the generation of spatial gradients of diversity, and thus further understanding the underlying mechanisms is a paramount goal of evolutionary biology.

FUNDING

ACKNOWLEDGMENTS

*

References

Anacker, B. L. and S. Y. Strauss. 2014. The geography and ecology of plant speciation: range overlap and niche divergence in sister species. Proceedings of the Royal Society of London B: Biological Sciences 281.

Benton, M. J. 1987. Progress and competition in macroevolution. Biological Reviews 62:305–338.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. 2017. Julia: A fresh approach to numerical computing. SIAM Review 59:65–98.

Brown, W. L. and E. O. Wilson. 1956. Character displacement. Systematic Zoology 5:49–64.

723 Butler, M. and A. King. 2004. Phylogenetic comparative analysis: A modeling approach for

724   adaptive evolution. The American Naturalist 164:683–695.

725 Cadena, C. D., K. H. Kozak, J. P. Gómez, J. L. Parra, C. M. McCain, R. C. K. Bowie,

726   A. C. Carnaval, C. Moritz, C. Rahbek, T. E. Roberts, N. J. Sanders, C. J. Schneider,

727   J. VanDerWal, K. R. Zamudio, and C. H. Graham. 2011. Latitude, elevational climatic

728   zonation and speciation in new world vertebrates. Proceedings of the Royal Society of

729   London B: Biological Sciences .

730 Cadena, C. D., F. Zapata, and I. Jimnez. 2018. Issues and perspectives in species

731   delimitation using phenotypic data: Atlantean evolution in darwins finches. Systematic

732   Biology 67:181–194.

733 Cadotte, M. W., T. J. Davies, and P. R. PeresNeto. 2017. Why phylogenies do not always

734   predict ecological differences. Ecological Monographs 87:535–551.

735 Cavender-Bares, J., K. H. Kozak, P. V. A. Fine, and S. W. Kembel. 2009. The merging of

736   community ecology and phylogenetic biology. Ecology Letters 12:693–715.

737 Clarke, M., G. H. Thomas, and R. P. Freckleton. 2017. Trait evolution in adaptive

738   radiations: Modeling and measuring interspecific competition on phylogenies. The

739   American Naturalist 189:121–137 pMID: 28107052.

740 Connell, J. H. 1980. Diversity and the coevolution of competitors, or the ghost of

741   competition past. Oikos 35:131–138.

742 Coyne, J. A. and H. A. Orr. 2004. Speciation. sinauer. Sunderland, MA .

743 Darwin, C. 1859. On the origin of species. 1 ed. Harvard University Press.

744 Davies, T. J., S. Meiri, T. G. Barraclough, and J. L. Gittleman. 2007. Species coexistence

745   and character divergence across carnivores. Ecology Letters 10:146–152.

Diamond, J. M. 1978. Niche shifts and the rediscovery of interspecific competition: Why did field biologists so long overlook the widespread evidence for interspecific competition that had already impressed darwin? American Scientist 66:322–331.

Drury, J., J. Clavel, M. Manceau, and H. Morlon. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. Systematic Biology 65:700.

Elredge, N. 1974. Character displacement in evolutionary time. American Zoologist 14:1083–1097.

Elton, C. 1946. Competition and the structure of ecological communities. Journal of Animal Ecology 15:54–68.

Farrington, H. L., L. P. Lawson, C. M. Clark, and K. Petren. 2014. The evolutionary history of darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. Evolution 68:2932–2944.

Freeman, B. G. 2015. Competitive interactions upon secondary contact drive elevational divergence in tropical birds. The American Naturalist 186:470–479 pMID: 26655571.

Garamszegi, L. Z. 2014. Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Springer, Berlin, Heidelberg.

Gause, G. F. 1934. The Struggle for Existence. Hadner Publishing Company, Inc., New York.

Gill, M. S., L. S. Tung Ho, G. Baele, P. Lemey, and M. A. Suchard. 2017. A relaxed directional random walk model for phylogenetic trait evolution. Systematic Biology 66:299–319.

Godoy, O., N. J. B. Kraft, J. M. Levine, and J. Chave. 2014. Phylogenetic relatedness and the determinants of competitive outcomes. Ecology Letters 17:836–844.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Systematic Biology 60:451–465.

Grant, B. R. and P. R. Grant. 2008. Fission and fusion of darwin's finches populations. Philosophical Transactions: Biological Sciences 363:2821–2829.

Grant, P. R. 1972. Convergent and divergent character displacement. Biological Journal of the Linnean Society 4:39–68.

Grant, P. R. 1999. Ecology and evolution of Darwin's finches. Princeton University Press.

Grant, P. R. and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of darwin's finches. Science 296:707–711.

Grant, P. R. and B. R. Grant. 2006. Evolution of character displacement in darwin's finches. Science 313:224–226.

Grant, P. R. and B. R. Grant. 2011. How and why species multiply: the radiation of Darwin's finches. Princeton University Press.

Grant, P. R., B. R. Grant, J. A. Markert, L. F. Keller, and K. Petren. 2004. Convergent evolution of darwin's finches caused by introgressive hybridization and selection. Evolution 58:1588–1599.

Grether, G. F., N. Losin, C. N. Anderson, and K. Okamoto. 2009. The role of interspecific interference competition in character displacement and the evolution of competitor recognition. Biological Reviews 84:617–635.

Hardin, G. 1960. The competitive exclusion principle. Science 131:1292–1297.

Harmon, L. J., J. B. Losos, D. T. Jonathan, R. G. Gillespie, J. L. Gittleman, J. W. Bryan, K. H. Kozak, M. A. McPeek, F. MorenoRoark, T. J. Near, A. Purvis, R. E. Ricklefs,

D. Schluter, J. A. Schulte II, O. Seehausen, B. L. Sidlauskas, O. TorresCarvajal, J. T. Weir, and A. . Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Hastings, W. K. 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57:97.

Horvilleur, B. and N. Lartillot. 2014. Monte carlo algorithms for brownian phylogenetic models. Bioinformatics 30:3020–3028.

Huelsenbeck, J. P. and B. Rannala. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. Evolution 57:1237–1247.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415–427.

Jordan, D. S. 1905. The origin of species through isolation. Science 22:545–562.

Kadmon, R. and O. Allouche. 2007. Integrating the effects of area, isolation, and habitat heterogeneity on species diversity: A unification of island biogeography and niche theory. The American Naturalist 170:443–454 pMID: 17879194.

Kozak, K. H. and J. Wiens. 2006. Does niche conservatism promote speciation? a case study in north american salamanders. Evolution 60:2604–2621.

Lack, D. 1947. The significance of clutchsize. Ibis 89:302–352.

Lack, D. 1954. The Natural Regulation of Animal Numbers. Oxford University Press, Ely House, London W. I.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, et al. 2015. Evolution of darwins finches and their beaks revealed by genome sequencing. Nature 518:371.

816 Lamichhaney, S., F. Han, M. T. Webster, L. Andersson, B. R. Grant, and P. R. Grant.
817   2018. Rapid hybrid speciation in darwin's finches. Science 359:224–228.

818 Landis, M. J., W. A. Freyman, and B. G. Baldwin. 2018. Retracing the hawaiian
819   silversword radiation despite phylogenetic, biogeographic, and paleogeographic
820   uncertainty. bioRxiv .

821 Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis
822   of biogeography when the number of areas is large. Systematic Biology 62:789.

823 Lartillot, N., M. J. Phillips, and F. Ronquist. 2016. A mixed relaxed clock model.
824   Philosophical Transactions of the Royal Society of London B: Biological Sciences 371.

825 Lartillot, N. and R. Poujol. 2011. A phylogenetic model for investigating correlated
826   evolution of substitution rates and continuous phenotypic characters. Molecular Biology
827   and Evolution 28:729–744.

828 Lemey, P., A. Rambaut, J. J. Welch, and M. A. Suchard. 2010. Phylogeography takes a
829   relaxed random walk in continuous space and time. Molecular Biology and Evolution
830   27:1877–1885.

831 Lotka, A. J. 1924. Elements of Physical Biology, reprinted 1956 as Elements of
832   Mathematical Biology. New York: Dover Publications.

833 Macarthur, R. and R. Levins. 1967. The limiting similarity, convergence, and divergence of
834   coexisting species. The American Naturalist 101:377–385.

835 MacColl, A. D. 2011. The ecological causes of evolution. Trends in Ecology & Evolution
836   26:514 – 522.

837 Manceau, M., A. Lambert, and H. Morlon. 2017. A unifying comparative phylogenetic
838   framework including traits coevolving across interacting lineages. Systematic Biology
839   66:551–568.

840 Matzke, N. J. 2014. Model selection in historical biogeography reveals that founder-event

841   speciation is a crucial process in island clades. Systematic Biology 63:951–970.

842 Mayr, E. 1970. Populations, species, and evolution: an abridgment of animal species and

843   evolution. Harvard University Press.

844 McEntee, J. P., J. A. Tobias, C. Sheard, and J. G. Burleigh. 2018. Tempo and timing of

845   ecological trait divergence in bird speciation. Nature ecology & evolution Page 1.

846 Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953.

847   Equation of state calculations by fast computing machines. The Journal of Chemical

848   Physics 21:1087–1092.

849 Mitchell, J. S. 2015. Extant-only comparative methods fail to recover the disparity

850   preserved in the bird fossil record. Evolution 69:2414–2424.

851 Neuhauser, C. and S. W. Pacala. 1999. An explicitly spatial version of the lotka-volterra

852   model with interspecific competition. The Annals of Applied Probability 9:1226–1259.

853 Nuismer, S. L. and L. J. Harmon. 2015. Predicting rates of interspecific interaction from

854   phylogenetic trees. Ecology Letters 18:17–27.

855 Petraitis, P. S. 1979. Likelihood measures of niche breadth and overlap. Ecology

856   60:703–710.

857 Pfennig, D. W. and K. S. Pfennig. 2012. Evolution's wedge: competition and the origins of

858   diversity. 12 Univ of California Press.

859 Pigot, A. L., W. Jetz, C. Sheard, and J. A. Tobias. 2018. The macroecological dynamics of

860   species coexistence in birds. Nature ecology & evolution Page 1.

861 Pigot, A. L. and J. A. Tobias. 2013. Species interactions constrain geographic range

862   expansion over evolutionary time. Ecology Letters 16:330–338.

863 Price, T. D., P. R. Grant, H. L. Gibbs, and P. T. Boag. 1984. Recurrent patterns of natural
864   selection in a population of darwin's finches. Nature 309:787.

865 Quintero, I., P. Keil, W. Jetz, and F. W. Crawford. 2015. Historical biogeography using
866   species geographical ranges. Systematic Biology 64:1059–1073.

867 Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue. 2005. A likelihood framework
868   for inferring the evolution of geographic range on phylogenetic trees. Evolution
869   59:2299–2311.

870 Ree, R. H., S. A. Smith, and A. Baker. 2008. Maximum likelihood inference of geographic
871   range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology
872   57:4–14.

873 Ricklefs, R. E. 1987. Community diversity: Relative roles of local and regional processes.
874   Science 235:167–171.

875 Ricklefs, R. E. 2010. Host–pathogen coevolution, secondary sympatry and species
876   diversification. Philosophical Transactions of the Royal Society B: Biological Sciences
877   365:1139–1147.

878 Robinson, D. M., D. T. Jones, H. Kishino, N. Goldman, and J. L. Thorne. 2003. Protein
879   evolution with dependence among codons due to tertiary structure. Molecular Biology
880   and Evolution 20:1692–1704.

881 Rodrigue, N., H. Philippe, and N. Lartillot. 2006. Assessing site-interdependent
882   phylogenetic models of sequence evolution. Molecular biology and evolution
883   23:1762–1775.

884 Rohwer, S. A. 1973. Significance of sympatry to behavior and evolution of great plains
885   meadowlarks. Evolution 27:44–57.

886 Rundell, R. J. and T. D. Price. 2009. Adaptive radiation, nonadaptive radiation, ecological
887   speciation and nonecological speciation. Trends in Ecology & Evolution 24:394 – 399.

888 Scheffer, M. and E. H. van Nes. 2006. Self-organized similarity, the evolutionary emergence
889   of groups of similar species. Proceedings of the National Academy of Sciences
890   103:6230–6235.

891 Schindel, D. E. and S. J. Gould. 1977. Biological interaction between fossil species:
892   Character displacement in bermudian land snails. Paleobiology 3:259–269.

893 Schluter, D. 2000. Ecological character displacement in adaptive radiation. The American
894   Naturalist 156:S4–S16.

895 Schluter, D. and J. D. McPhail. 1992. Ecological character displacement and speciation in
896   sticklebacks. The American Naturalist 140:85–108.

897 Schluter, D., T. D. Price, and P. R. Grant. 1985. Ecological character displacement in
898   darwin's finches. Science 227:1056–1059.

899 Simberloff, D. and W. Boecklen. 1991. Patterns of extinction in the introduced hawaiian
900   avifauna: A reexamination of the role of competition. The American Naturalist
901   138:300–327.

902 Simpson, G. G. 1950. History of the fauna of latin america. American Scientist 38:361–389.

903 Simpson, G. G. 1953. Evolution and Geography: An Essay on Historical Biogeography,
904   with Special Reference to Mammals. Oregon State System of Higher Education.

905 Slater, G. J., L. J. Harmon, and M. E. Alfaro. 2012. Integrating fossils with molecular
906   phylogenies improves inference of trait evolution. Evolution 66:3931–3944.

907 Slatkin, M. 1974. Competition and regional coexistence. Ecology 55:128–134.

Smith, B. T., J. E. McCormack, A. M. Cuervo, M. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-Emán, C. W. Burney, X. Xie, M. G. Harvey, et al. 2014. The drivers of tropical speciation. Nature 515:406–409.

Stroud, J. T. and J. B. Losos. 2016. Ecological opportunity and adaptive radiation. Annual Review of Ecology, Evolution, and Systematics 47:507–532.

Sukumaran, J. and L. L. Knowles. 2018. Trait-dependent biogeography: (re)integrating biology into probabilistic historical biogeographical models. Trends in Ecology & Evolution 33:390 – 398.

Tagliacollo, V. A., S. M. Duke-Sylvester, W. A. Matamoros, P. Chakrabarty, and J. S. Albert. 2015. Coordinated dispersal and Pre-Isthmian assembly of the Central American ichthyofauna. Systematic biology 66:183–196.

Tilman, D. 1985. The resource-ratio hypothesis of plant succession. The American Naturalist 125:827–852.

Tobias, J. A., C. K. Cornwallis, E. P. Derryberry, S. Claramunt, R. T. Brumfield, and N. Seddon. 2014. Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. Nature 506:359.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2015. Comparative analysis of principal components can be misleading. Systematic Biology 64:677–689.

Uyeda, J. C. and L. J. Harmon. 2014. A novel bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Systematic Biology 63:902–918.

Valone, T. and J. Brown. 1995. Effects of competition, colonization, and extinction on rodent species diversity. Science 267:880–883.

Van Valen, L. 1973. A new evolutionary law. Evolutionary Theory 1:1–30.

932 Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and

933 community ecology. Annual Review of Ecology and Systematics 33:475–505.

934 Weir, J. T. and T. D. Price. 2011. Limits to speciation inferred from times to secondary

935 sympatry and ages of hybridizing species along a latitudinal gradient. The American

936 Naturalist 177:462–469 pMID: 21460568.

937 Wiens, J. J. and M. J. Donoghue. 2004. Historical biogeography, ecology and species

938 richness. Trends in Ecology and Evolution 19:639–644.

939 # FIGURES
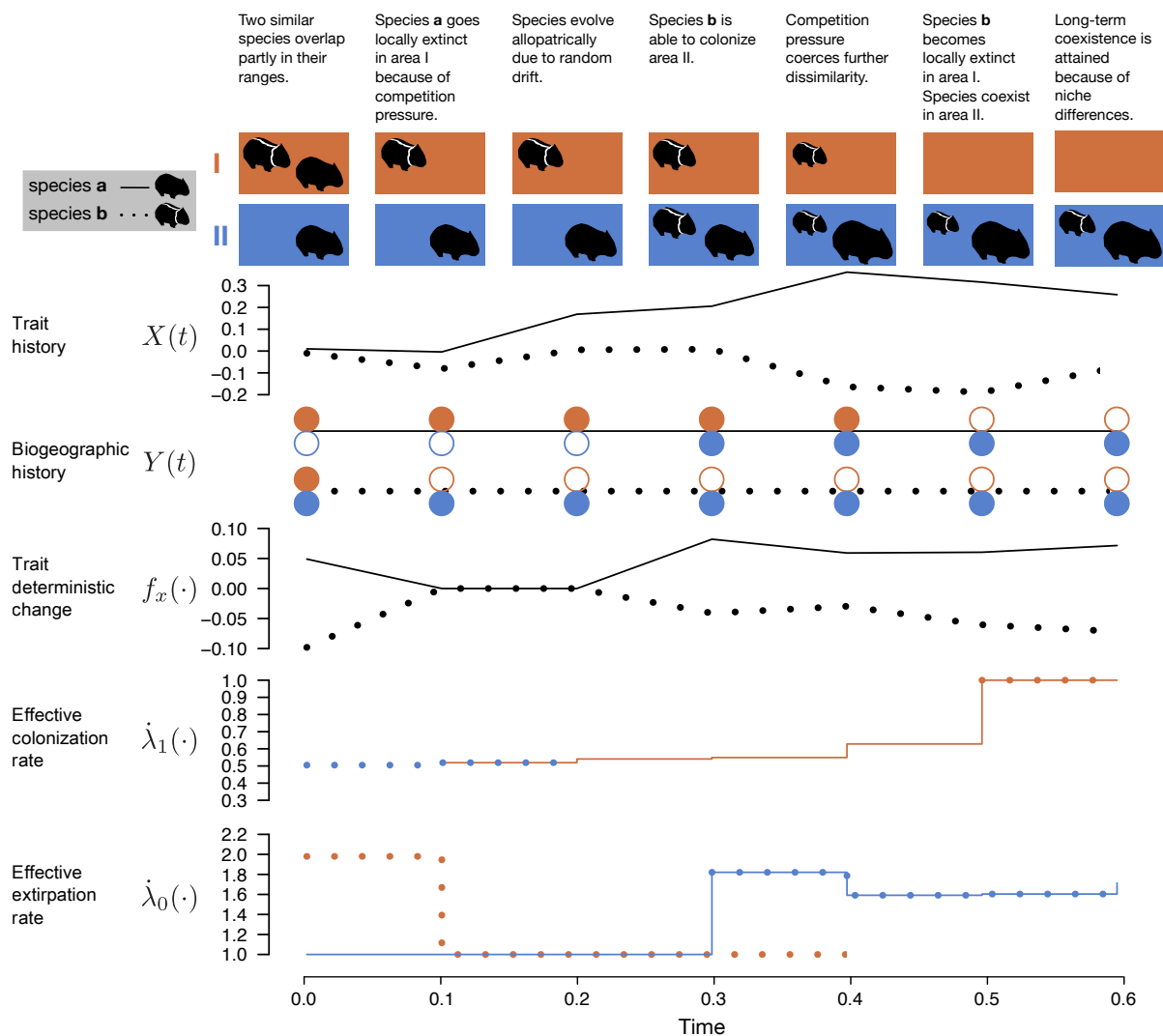
Figure 1: Hypothetical example of a time-discrete history with interdependence between biogeographic and trait evolution for two species, **a** (no stripes and solid lines) and **b** (white stripes and dotted lines), across two areas, **I** (orange) and **II** (blue). We assume that there is *in situ* competition, fixing $\omega_x = -1$, that there is competitive exclusion by fixing $\omega_1 = -1$, and that there is extinction mediated competition by fixing $\omega_0 = 1$. Furthermore, we assume that the random drift $\sigma^2 = 0.1$, the base rate of colonization $\lambda_1 = 1$ and the base rate of extinction $\lambda_0 = 1$. The trait under consideration is the standardized size, specified by $X(t)$. $Y(t)$ conveys the specific biogeographic history for each species; filled circles represent the species occupies the area while empty ones that it is absent. The deterministic component of our Stochastic Differential Equation is given by $f_x(\cdot)$ and determines the directionality of trait change when in sympatry (Equation 4). Effective rates of colonization per species per area is given by $\dot{\lambda}_1(\cdot)$; the highest rate of colonization is λ_1 and is given when an area is empty (e.g., last two time steps for area **II**; Equation 5). Effective rates of extinction per species per area is given by $\dot{\lambda}_0(\cdot)$; the lowest rate of local extinction is λ_0 and is given when the species is alone in an area (Equation 5). Drawings and values are mathematically consistent following our model.
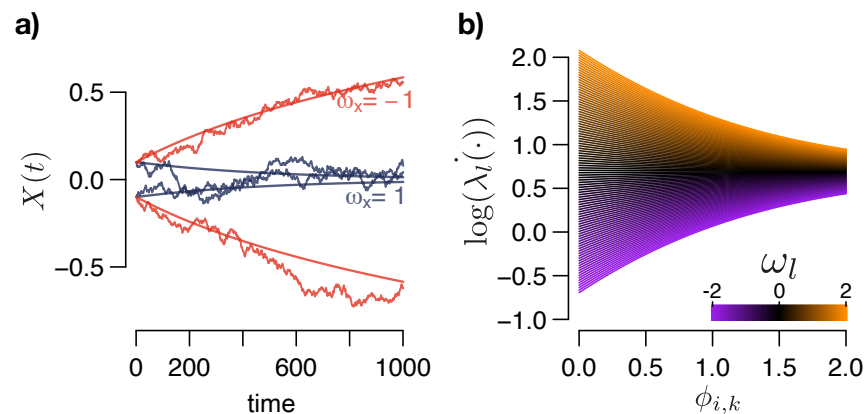
Figure 2: Functional forms for the joint evolution of trait and ranges. **a)** An illustration of the Stochastic Differential Equation (SDE) used to model the role of biotic interactions in trait evolution. We plot trait evolution as the stochastic (diffusion) component superimposed upon the deterministic (interspecific) component. At time $t = 0$, the phenotypic values of two lineages, $X_a(t) = -0.1$ and $X_b(t) = 0.1$, evolve according to the *in situ* biotic interations parameter, ω_x. If $\omega_x < 0$, the lineages repel each other, if $\omega_x = 0$, the lineage evolves by random drift, and if $\omega_x > 0$, they attract each other. **b)** Functional form relating trait differences for lineage i and those in area k, $\phi_{i,k}$, and the logarithm of the effective rates of colonization or extinction, $\log(\dot{\lambda}_l(\cdot))$. Here, l indicates a gain (1) or loss (0) event, for different values of ω_l. Purple colors represent ω_l values close to -2 and orange colors close to 2. If $\omega_l < 0$, lower trait differences between lineages suffer higher penalties in rates of colonization or extirpation relative to larger differences, if $\omega_l = 0$, then $\dot{\lambda}_l(\cdot) = \lambda_l = 2$, and finally if $\omega_l < 0$, larger trait differences between lineages enhance the rates of colonization or extirpation.
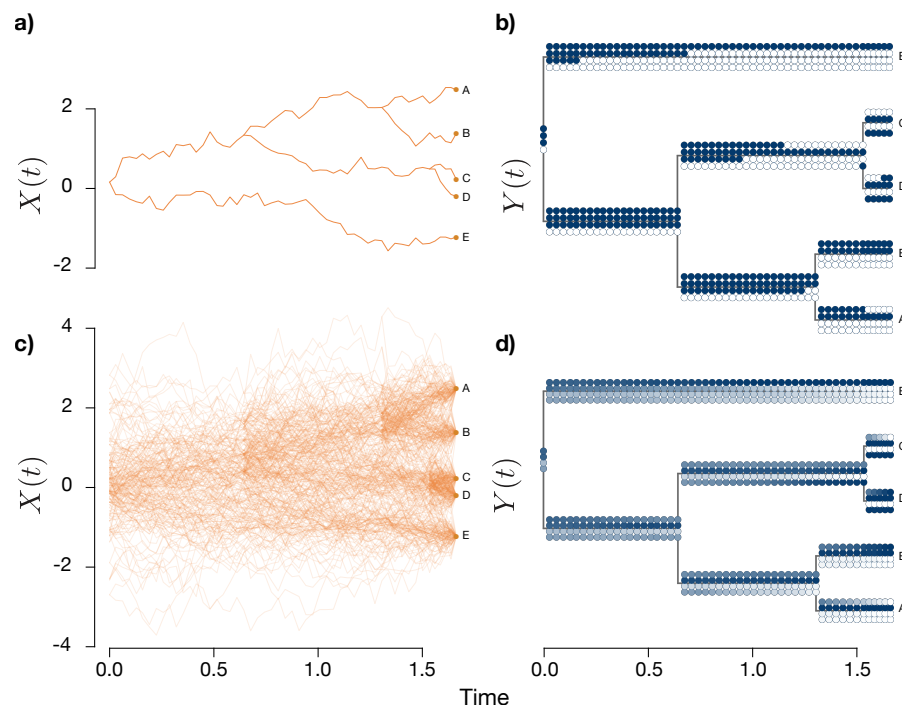
Figure 3: Illustration of the discretized data augmentation used from a simulation performed on an ultrametric tree of 5 tips and four areas with in situ competition (i.e., $\omega_x = -1$). **a)** One random sample trait history, $X(t)$, from the posterior. **b)** One random sample of biogeographic range history, $Y(t)$, from the posterior across four areas. Each time sample has four circles in vertical orientation, each representing one of the areas. Filled circles represent occupied areas while empty circles represent absence. Note that all branches have at least five internal discrete sampling times, that is, one more than the number of areas in the current system. We set the minimum time interval here to be 2% for the tree height for illustration purposes. **c)** Marginal posterior data augmented histories based on 100 samples in trait with translucency. **d)** Corresponding marginal biogeographic histories. Darker tones represent higher marginal probabilities of area occupancy.
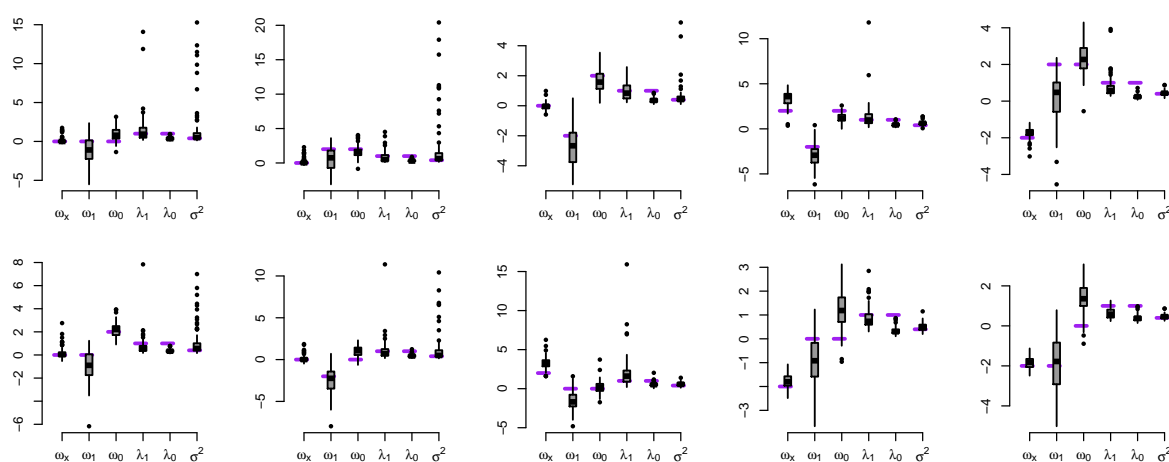
Figure 4: Boxplots of median posterior estimates from the different simulation scenarios. Each panel represents 100 different simulations in pure-birth trees with 25 tips and 10 areas. The true values used for the simulations are represented in horizontal dotted purple lines.
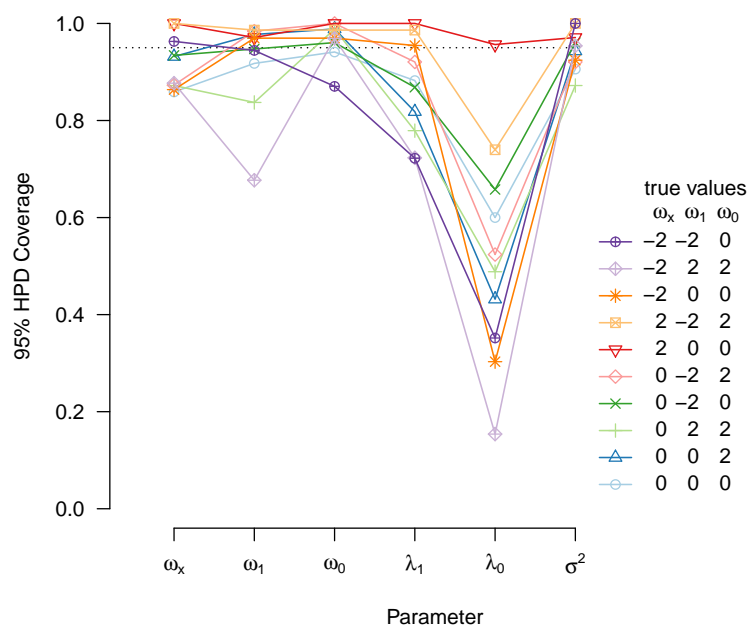

Figure 5: Posterior statistical 95% Highest Posterior Density (HPD) coverage for the 10 simulation scenarios for each parameter. Each symbol and color represents a different set of true values used for the simulation, corresponding to those used in Figure 4. The dotted line corresponds to 95% of HPDs across simulations covering the true simulated parameter.
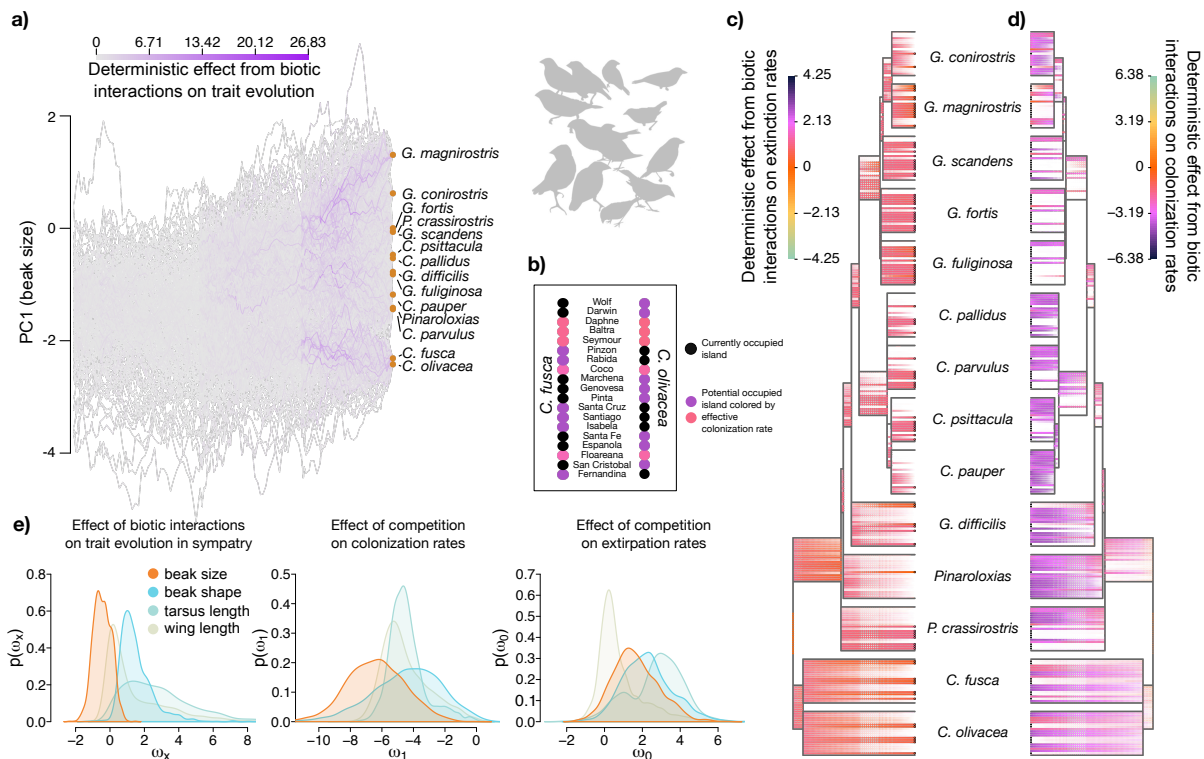
Figure 6: Empirical results for the effect of biotic interactions on the trait and biogeographic evolution of Darwin's finches. **a)** 100 data augmented trait histories for PC1 (beak size). Absolute deterministic effects of biotic interactions on trait evolution for sympatric lineages are colored from grey (isolated evolution under Brownian motion) to purple (strongest effect of biotic interactions). **b)** Example of present-day effect of biotic interactions in colonization rates between two species that are phenotypically similar, *Certhidea fusca* and *C. olivacea*. The areas are displayed as circles arranged in a column, with currently occupied areas (islands) in black and unoccupied areas colored according to effective colonization rates following the color scale in Figure 6d (below). Note that areas occupied by the sister species suffer a colonization penalty and reflect competitive exclusion in beak size as given by our model. **c)** Marginal data augmented biogeographic histories for the same 19 areas shown in Figure 6b. Alpha opacity denotes the marginal probability of presence at a given time for a given lineage-area. The color scale represents the average effect of biotic interactions on local extinction rates (purple denoting higher rates of local extinction and orange, no influence). Currently occupied areas are shown with black unfilled circles at the tips. **d)** As in Figure 6c, but alpha opacity denote the marginal probabilities of absences at a given time for a given lineage-area, and the color scale represent the average effect of biotic interactions on colonization rates (purple denoting lower rates of colonization and orange no influence). Currently occupied areas are shown with black filled circles at the tips. **e)** Posterior marginal densities for the parameter governing biotic interactions (*left: ω_x, middle: ω_1, right: ω_0*) for each of the four phenotypic traits analyzed separately. The results suggest *in situ* competition for beak size and strong convergence for tarsus and beak shape. All traits show strong penalization for colonization when similar. See text for further details. Finch silhouettes from Caroline O'Donnell, redrawn from Biological Sciences Curriculum Study, *Biological Science: Molecules to Man*, Houghton Mifflin (1963).