Online Appendix: Supplemental methods and results

for Evolutionary branching in complex landscapes

Benjamin C. Haller,[1] Rupert Mazzucco,[2] and Ulf Dieckmann[2]

1. Department of Biology and Redpath Museum, McGill University, 859 Sherbrooke Street West, Montreal, Quebec, Canada H3A 0C4;
2. Evolution and Ecology Program, International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria

---

Model description

To test our hypotheses, we constructed an individual-based model, described below following guidelines suggested for the standardized description of such models (Grimm et al. 2006). The model description is also intended to follow closely that of Heinz et al. (2009), itself a descendant of Doebeli and Dieckmann (2003). A brief summary of the model is provided in the main text (see Methods, *Model summary*).

MODEL OVERVIEW

The purpose of this model is to understand the effects of complex spatial environmental heterogeneity on the process of speciation. The model is related to the spatially explicit individual-based stochastic model of Doebeli and Dieckmann (2003), restricted to the asexual case of that model. Our model replaces the linear spatial environmental gradient of Doebeli and Dieckmann (2003) with a complex landscape composed of a nonlinear environmental gradient and continuously varying environmental patchiness. As a consequence of the individual-based approach, stochasticity manifests in several aspects of our model, including variation in the landscape, variation in the initial state, demographic stochasticity, random dispersal, and random mutations during reproduction, as described in the following sections.

ENVIRONMENTAL LANDSCAPE AND STATE VARIABLES

Individuals inhabit a two-dimensional continuous landscape. The landscape is constant in time; its spatial heterogeneity is composed of a gradient in the x-direction, and of patchiness that varies continuously in both the x- and y-directions (see LANDSCAPE HETEROGENEITY). Individuals in this landscape are characterized by their location (x, y), with $0 \le x, y \le 1$, and by their ecological character u, which affects their adaptation to the local environment. The state of the model at any point in time is thus fully specified by triplets (x_i, y_i, u_i) for all individuals i in the population ($i = 1, ..., N$, where N is the current population size). Individuals are affected both by the landscape and by the other individuals in the model through their death rate (see DEATH EVENTS).

LANDSCAPE HETEROGENEITY

The carrying capacity density K at a location (x, y) for ecological character u is defined as

$$K(u, x, y) = K_0 N_{\sigma_K}\left(u - u_0(x, y)\right),$$

where K_0 is the maximum carrying capacity density, $N_\sigma(z) = \exp(-z^2/2\sigma^2)$ denotes a Gaussian function of standard deviation σ, and $u_0(x, y)$ denotes the optimal ecological character value (the value that produces the greatest carrying capacity density) at the given location; $u_0(x, y)$ thus describes the environmental landscape and is defined as

$$u_0(x, y) = \alpha + Sx + \tfrac{1}{2}Cx^2 + \Phi_{L_g, A}(x, y)\,.$$

In this expression, the centering constant α is always selected after landscape generation so as to yield a mean value of 0.5 for $u_0(x, y)$ over the landscape as a whole, while S and C are model parameters defining the slope and curvature, respectively, of the environmental gradient in the x-direction. C represents the maximum environmental gradient due to curvature (at $x = 1$); with a quadratic curvature term, the coefficient $\frac{1}{2}C$ thus represents the average environmental gradient due to curvature. The function $\Phi_{L_g, A}(x, y)$ specifies the autocorrelated noise that describes the environmental patchiness, which is further characterized by the autocorrelation length L_g and the patchiness amplitude A. The function Φ is constructed as the convolution of

Gaussian white noise and a two-dimensional filter function constructed to produce the desired autocorrelation length L_g. The model parameters S, C, L_g, and A thus together govern the generation of the landscape. For further details on the generation procedure, see *Landscape generation*.

A survey of some landscapes generated by this model, and of the effects of parameters on the generated landscape, is shown in Figure 1.

BOUNDARY CONDITIONS

In the x-direction (the direction of the environmental gradient), we tested four boundary conditions: stopping, reflecting, reprising, and absorbing. These boundary conditions differ in the action taken when a natal dispersal location (see BIRTH EVENTS AND DISPERSAL) is drawn outside of the valid range $0 \le x \le 1$. Stopping boundaries replace values $x < 0$ or $x > 1$ by values 0 or 1, respectively, causing dispersers to stop at the edge of the landscape. Reflecting boundaries replace values $x < 0$ or $x > 1$ by values $-x$ or $2 - x$, respectively, as if dispersers bounced off of the edge of the landscape (repeatedly, if necessary, until the location is valid). Reprising boundaries redraw invalid locations until a valid location is obtained, as if dispersers avoid areas beyond the landscape's edge. Absorbing boundaries, finally, remove the individual from the population if an invalid location is drawn; dispersal beyond the edge of the landscape is in this case lethal.

In the y-direction, for which only stochastic environmental heterogeneity exists, periodic boundary conditions were utilized. Periodic boundaries replace values $y < 0$ or $y > 1$ by values $y + 1$ or $y - 1$, respectively (repeatedly, if necessary, until the location is valid). The landscape was generated in such a manner as to guarantee that it meshed seamlessly across the periodic y boundaries (see *Landscape generation*). The landscape thus described is cylindrical in topology for all four boundary conditions used. Periodic boundaries (or cline-periodic boundaries; Heinz et al. 2009) could not be used in the x-direction because the nonlinear environmental gradient (see LANDSCAPE HETEROGENEITY) would cause discontinuity in the landscape.

PHENOTYPE RANGES AND INITIALIZATION

Initially, the model contains N_{init} individuals with state (x_i, y_i, u_i) for each individual i, where x_i and y_i are drawn from a uniform distribution between 0 and 1, and $u_i = 0.5$ (matching the mean value of $u_0(x, y)$ over the landscape). No bounds are enforced by the model on values of the ecological character u thereafter.

This initial state distributes individuals across the entire landscape, although the poorly adapted individuals typically die almost immediately. Biologically, this distribution might represent an unusual event such as a storm that introduces individuals, propagules, or seeds across a landscape, or might represent a previously homogeneous landscape suddenly transformed into heterogeneity by a catastrophic event such as fire, flooding, or volcanism. However, additional test realizations (not shown) suggest that the final outcome of realizations does not depend strongly upon this initial distribution. With all but the shortest dispersal distances, the landscape is quickly colonized and the initial distribution has little effect. With very short-range dispersal, movement across the landscape is very slow and the waiting time to branching can thus be much longer, but the final outcome is similar. Given sufficiently long runtimes, then, we expect that different initial conditions would not substantially affect our results, other than by broadening the branching time distribution.

INTERACTIONS

Individuals interact only through local competition for resources. The strength of competition felt by focal individual i is the sum of the competitive impacts of all other individuals in the system, and is expressed as the effective number of individuals competing with the focal individual,

$$n_{\text{eff}}(u_i, x_i, y_i) = \left(2\pi\sigma_s^2\right)^{-1} \sum_{j=1, j\neq i}^{N} N_{\sigma_c}(u_j - u_i)\, N_{\sigma_s}(x_j - x_i)\, N_{\sigma_s}(y_j - y_i).$$

The strength of competition decreases with increasing spatial distance, as described by a Gaussian function with standard deviation σ_s, the spatial competition radius. The strength of competition also decreases with increasing

ecological character difference between the individuals, as described by a Gaussian function with standard deviation σ_c, the phenotypic competition width. The parameters σ_s and σ_c can therefore be thought of as scaling factors for the impact of spatial differences and phenotypic differences, respectively, on the strength of competition (Doebeli and Dieckmann 2003). The normalization factor $(2\pi\sigma_s^2)^{-1}$ ensures that a spatially uniform population is regulated to a given overall environmental carrying capacity regardless of the value of σ_s. Interactions span the periodic y-boundary; the shortest distance between individuals in the cylindrical landscape is used. As the competition felt by an individual increases, that individual's death rate increases, as described below (DEATH EVENTS).

DEATH EVENTS

The death rate d_i of each individual i is computed as $n_{\text{eff}}(u_i, x_i, y_i)/K(u_i, x_i, y_i)$, where $n_{\text{eff}}(u_i, x_i, y_i)$ denotes the strength of competition felt by individual i (see INTERACTIONS), and $K(u_i, x_i, y_i)$ denotes the carrying capacity density experienced by individual i (see LANDSCAPE HETEROGENEITY). Competition in this model therefore produces soft selection (Wallace 1975), and cannot lead to extinction. When an individual dies, it is removed from the population.

BIRTH EVENTS AND DISPERSAL

The birth rate of every individual i is equal and constant: $b_i = b$. When an individual gives birth, an offspring individual is generated asexually and added to the population. We model natal dispersal as the only form of movement; the spatial coordinates x and y of the offspring are thus drawn from normal distributions with standard deviation V and means equal to the parental values x_i and y_i, and then the boundary conditions are applied as described in BOUNDARY CONDITIONS. The ecological character u of the offspring is nearly faithfully inherited from the parent, except that with probability μ_m a mutational offset, drawn from a normal distribution with mean 0 and standard deviation σ_m, is added to the parental value u_i to represent a mutation event during reproduction.

Since the birth rate b is constant, we scale time by $1/b$. In $1/b$ units of time every individual is expected to produce one offspring, and if the population size is at equilibrium, the same number of individuals is expected to die in that period of time. For this reason, we refer to $1/b$ units of time as a "generation".

PROCESS OVERVIEW AND SCHEDULING

Overlapping generations of individuals are implemented in the model using a continuous-time birth–death process implemented following the so-called Direct Method of Gillespie (1976). First, the current birth rate b_i and current death rate d_i of each individual i are determined by the state of the model, as described above (DEATH EVENTS; BIRTH EVENTS AND DISPERSAL). From these, the population-level birth and death rates are calculated as sums over the individual rates,

$$B = \sum_{i=1}^{N} b_i \text{ and } D = \sum_{i=1}^{N} d_i$$

respectively. The overall event rate is then calculated as $E = B + D$, and the waiting time to the next event is drawn from an exponential distribution with mean $1/E$. With probability B/E this is a birth event, and an individual i is chosen for reproduction with probability b_i/B. Otherwise, with probability D/E, it is a death event, and an individual i is chosen for removal with probability d_i/D. At the end of the event, the model state is updated, and the birth and death rates of all individuals are recalculated.

OBSERVABLES

The $u_0(x, y)$ values that define the landscape were recorded on generation of the landscape, at the start of each realization of the model. Each generated landscape was analyzed to extract metrics describing its salient features: the landscape's scaled standard deviation σ, skewness γ_1, excess kurtosis γ_2, and autocorrelation length l_r (see *Landscape analysis*).

Every 10 generations during each realization a histogram of the ecological character values of the population was recorded. At the end of each realization, the final state (x_i, y_i, u_i) of all individuals was recorded. Following the recommendations of Grimm (2002) regarding the importance of visual debugging, the landscape and the state of all individuals were also observable graphically during model runs. An example of the information available for each realization is presented in Figure 2, and a movie showing an example realization is provided as

Movie 1 (available in the Dryad data repository, http://dx.doi.org/10.5061/dryad.43cj7/1).

Branching was identified by a heuristic analysis of the histograms of ecological character values recorded during the realization (for example, Figure 2D). The initial distribution of ecological character values was unimodal, since we set $u_i = 0.5$ for all individuals. Branching was defined as divergence into a visibly bi- or multi-modal distribution, established not only at the end of the realization, but for a substantial period of time (3000 generations) prior to the end of the realization, to demonstrate the stability of the outcome. An alternative assessment of branching based on cluster analysis was also conducted (see *Alternative assessment of branching*). Finally, for realizations considered to have branched, the branching time was determined (see *Distribution of branching times*).

PARAMETERS

Parameters of the model, including the metrics taken from the realized model landscapes, are shown in Table 1. The parameter σ_s defines a fundamental length scale of the model, because distance in space affects model dynamics through the strength of competition scaled by σ_s. Similarly, the parameter σ_K defines a fundamental ecological phenotype scale of the model, because distance in phenotype affects model dynamics through the carrying capacity density, scaled by σ_K. For this reason, following Doebeli & Dieckmann (2003), we rescaled all other parameters with units involving length or ecological phenotype, using σ_s and σ_K, to derive dimensionless scaled parameters. The unscaled parameters V, S, C, L_g, and A were transformed according to the formulas shown (Table 1) to produce the scaled parameters v, s, c, l_g, and a. These scaled parameters were varied among realizations of the model.

Values for the parameters that were not varied (N_{init}, K_0, σ_K, σ_s, μ_m, σ_m, b) were taken from the equivalent parameters of Doebeli & Dieckmann (2003) so as to allow easy comparison of results between these models. The exception is c_p, chosen here to explore dynamics in the limit of phenotypically indiscriminate competition (large c_p). In this limit, negative frequency-dependent selection would not cause branching in the absence of spatial heterogeneity; this choice thus guarantees that all observed branching is due to the effects of spatial heterogeneity (Doebeli and Dieckmann 2003).

MODEL IMPLEMENTATION

The model was implemented in Objective-C using the Cocoa object-oriented toolkit (Mac OS X 10.6.4; Apple Inc., http://www.apple.com), except for the scheduling module, landscape construction module, and branching analysis module, all of which were written in the `R` programming language (version 2.11.1–2.15.1; `R` Development Core Team, http://www.r-project.org).

Comparison of nested GAM models

The nested models based on the landscape-generating parameters s, c, a, and l_g, in addition to the dispersal distance v, are compared in Table A1 and Figure A1. Similarly, the nested models based on the realized-landscape metrics σ, γ_1, γ_2, and l_r, in addition to v, are compared in Table A2 and Figure A2.

Intercept-only models correctly predicted the outcome of 78.4% of realizations, since that percentage of realizations did not branch (Table A3); these models simply predict that branching will never occur. The prediction rate observed in other models should thus be evaluated relative to this baseline.

A model including only the generating ACL, l_g, only marginally improved on the intercept-only null model (Table A1, model #1 *vs.* #0); similarly, a model with the interaction $l_g * v$ only marginally improved on a model without that interaction (Table A1, model #8 *vs.* #6). However, both terms did improve the BIC score, and were therefore retained. Other terms were more clearly of non-negligible effect size, according to both BIC and Nagelkerke R^2 (change in prediction rate is not a good indicator of effect size with a binary outcome, because a term can improve the certainty of predictions without changing their direction).

Binary predictions (not branched vs. branched) from the best GAM from each set of nested models (GAM_g and GAM_r) agreed in 94.8% of realizations (Figure A5).

The results presented in this section are for reprising boundary conditions; qualitative results were the same for other boundary conditions.

Table A1: Comparison of nested GAM models predicting the outcome of a realization (not branched versus branched) from the scaled landscape-generating parameters s, c, a and l_g and the scaled dispersal distance v, for reprising boundary conditions.

	Model	Terms	EDF	*C*	R^2	*D*	Match	ΔBIC
0.	(Intercept)	0	1.0	.500	.000	.000	.784	159104
1.	l_g	1	3.2	.507	.000	.000	.784	159098
2.	a	1	4.2	.522	.002	.001	.784	158821
3.	s	1	4.9	.625	.052	.034	.784	148851
4.	c	1	4.9	.703	.130	.088	.784	132781
5.	v	1	5.0	.829	.386	.288	.845	72896
6.	$s + c + a + l_g + v$	5	20.0	.922	.600	.493	.882	11585
7.	[6] + $c * v$	6	33.0	.923	.601	.493	.882	11473
8.	[6] + $l_g * v$	6	33.6	.923	.601	.494	.883	11350
9.	[6] + $s * a$	6	33.9	.923	.604	.496	.884	10641
10.	[6] + $s * v$	6	34.8	.924	.604	.496	.884	10629
11.	[6] + $a * v$	6	36.2	.925	.609	.502	.885	8910
12.	[6] + $s * c$	6	36.0	.926	.613	.506	.886	7792
13.	[6] + $c * a$	6	36.4	.928	.617	.511	.887	6368
14.	[7–13]	12	112.7	.933	.637	.533	.894	753
15.	**[14] + $s * c * a$**	**13**	**167.0**	**.935**	**.642**	**.537**	**.895**	**0**

Note: The Terms column shows the number of terms in the model, while the EDF column gives the estimated degrees of freedom used. *C* is the area under the Receiver Operating Characteristic curve (Metz, 1978), R^2 is the Nagelkerke generalized R^2 (Nagelkerke, 1991), *D* is the Discrimination coefficient (Tjur, 2009), Match is the fraction of cases in which the model prediction matched the observed binary outcome, and ΔBIC is the Bayesian Information Criterion relative to the best model (Schwarz, 1978). Better models are indicated by high values, except for ΔBIC, for which a low score is better. Models are sorted by descending ΔBIC; the best model according to BIC (the "GAM_g" model) is shown in bold. All terms in all models are significant with $P < 0.001$. See Figure A1 for a graphical depiction of the best ten models.

Table A2: Comparison of nested GAM models predicting the outcome of a realization (not branched versus branched) from the scaled realized-landscape metrics σ, γ_1, γ_2, and l_r and the scaled dispersal distance v, for reprising boundaries.

	Model	Terms	EDF	*C*	R^2	*D*	Match	ΔBIC
0.	(Intercept)	0	1.0	.500	.000	.000	.784	173875
1.	γ_2	1	5.0	.579	.022	.014	.784	169695
2.	γ_1	1	5.0	.600	.032	.021	.784	167732
3.	l_r	1	5.0	.611	.040	.026	.784	165999
4.	σ	1	5.0	.760	.213	.149	.794	129258
5.	v	1	5.0	.829	.386	.288	.845	87667
6.	$\sigma + \gamma_1 + \gamma_2 + l_r + v$	5	20.8	.943	.666	.564	.899	4795
7.	[6] + $\sigma * v$	6	35.5	.943	.668	.568	.900	4055
8.	[6] + $l_r * v$	6	33.4	.946	.677	.577	.904	1087
9.	**[7–8]**	**7**	**49.3**	**.947**	**.681**	**.582**	**.906**	**0**

Note: The Terms column shows the number of terms in the model, while the EDF column gives the estimated degrees of freedom used. *C* is the area under the Receiver Operating Characteristic curve (Metz, 1978), R^2 is the Nagelkerke generalized R^2 (Nagelkerke, 1991), *D* is the Discrimination coefficient (Tjur, 2009), Match is the fraction of cases in which the model prediction matched the observed binary outcome, and ΔBIC is the Bayesian Information Criterion relative to the best model (Schwarz, 1978). Better models are indicated by high values, except for ΔBIC, for which a low score is better. Models are sorted by descending ΔBIC; the best model according to BIC (the "GAM_r" model) is shown in bold. All terms in all models are significant with $P < 0.001$. See Figure A2 for a graphical depiction of the best four models.

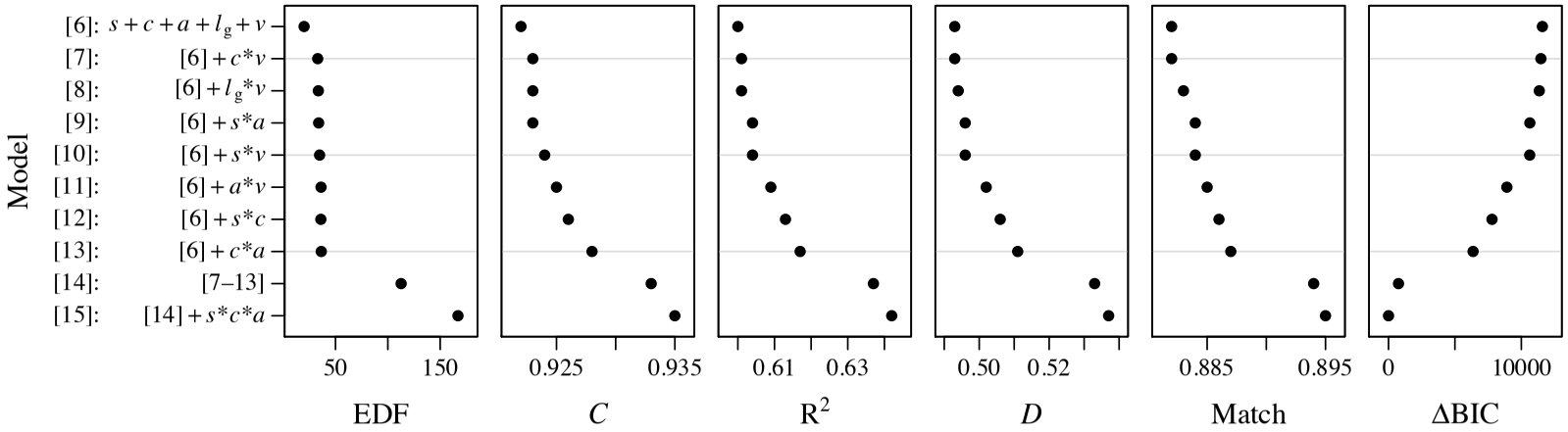

Figure A1: Comparison of the best (ΔBIC < 15000) nested GAMs predicting the outcome of a realization (not branched versus branched) from the scaled landscape-generating parameters s, c, a and l_g and the scaled dispersal distance v, for reprising boundary conditions. This figure is a graphical depiction of models 6–15 in Table A1. EDF: the estimated degrees of freedom used. C: the area under the Receiver Operating Characteristic curve (Metz, 1978). R^2: the Nagelkerke generalized R^2 (Nagelkerke, 1991). D: the Discrimination coefficient (Tjur, 2009). Match: the fraction of cases in which the model prediction matched the observed binary outcome. ΔBIC: the Bayesian Information Criterion, relative to that of the best model (Schwarz, 1978). Better models are indicated by high values, except for ΔBIC, for which a low value is better. All terms in all models are significant with $P < 0.001$. Note that all goodness-of-fit statistics are strongly correlated, and have perfect rank correlation.

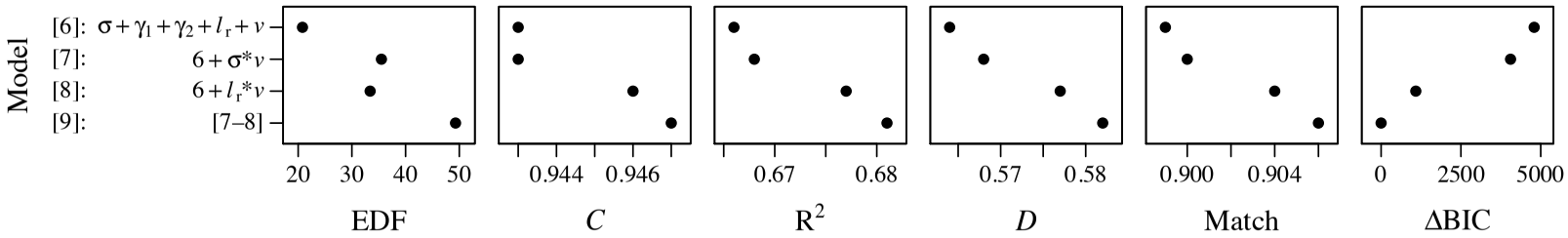

Figure A2: Comparison of the best (ΔBIC < 15000) nested GAMs predicting the outcome of a realization (not branched versus branched) from the scaled realized-landscape metrics σ, γ_1, γ_2, and l_r and the scaled dispersal distance v, for reprising boundary conditions. This figure is a graphical depiction of models 6–9 in Table A2. EDF: the estimated degrees of freedom used. C: the area under the Receiver Operating Characteristic curve (Metz, 1978). R^2: the Nagelkerke generalized R^2 (Nagelkerke, 1991). D: the Discrimination coefficient (Tjur, 2009). Match: the fraction of cases in which the model prediction matched the observed binary outcome. ΔBIC: the Bayesian Information Criterion, relative to that of the best model (Schwarz, 1978). Better models are indicated by high values, except for ΔBIC, for which a low value is better. All terms in all models are significant with $P < 0.001$. Note that all goodness-of-fit statistics are strongly correlated, and have perfect rank correlation.

Table A3: Comparison of results for the four boundary conditions.

	Absorbing	Reprising	Reflecting	Stopping
Branched (%)	18.5	21.6	21.9	24.0
Branched $\overline{s}$	0.814	0.804	0.795	0.769
Branched $\overline{c}$	0.358	0.341	0.340	0.336
Branched $\overline{a}$	1.416	1.449	1.473	1.497
Branched $\overline{l_g}$	1.555	1.542	1.525	1.524
Branched $\overline{v}$	0.552	0.732	0.728	0.787
Branched $\overline{\sigma}$	2.856	2.770	2.763	2.714
Branched $\overline{\gamma_1}$	0.268	0.248	0.248	0.246
Branched $\overline{\gamma_2}$	–0.784	–0.759	–0.744	–0.717
Branched $\overline{l_r}$	1.131	1.115	1.104	1.083
Clustered (%)	19.6	23.0	23.2	25.4
Avg. clusters	2.64	2.63	2.65	2.66
GAM_g R^2	0.700	0.642	0.651	0.658
GAM_g pred.	0.918	0.895	0.895	0.887
GAM_r R^2	0.728	0.681	0.689	0.689
GAM_r pred.	0.926	0.906	0.906	0.897

Note: Branched (%) is the percentage of realizations of the 300,000 realizations conducted for the given boundary condition that resulted in evolutionary branching according to our phenotypic history analysis. The Branched $\overline{s}$, $\overline{c}$, $\overline{a}$, $\overline{l_g}$, $\overline{v}$, $\overline{\sigma}$, $\overline{\gamma_1}$, $\overline{\gamma_2}$, and $\overline{l_r}$ are the mean value of that parameter or metric, among only those realizations that branched. Clustered (%) is the percentage of realizations that resulted in evolutionary branching according to alternative cluster analysis method of detecting branching; "Avg. clusters" is the average number of clusters produced, among only those realizations that produced two or more clusters. GAM_g results are for the full GAM model fitted using the model parameters, as in Table A1 model #15; the Nagelkerke R^2 and the proportion of realization outcomes correctly predicted are given. Similarly, GAM_r results are for the full GAM fitted using the landscape metrics, as in Table A2 model #9.

Landscape-generating parameters versus realized-landscape metrics

Although the generating ACL l_g and its interaction $l_g * v$ had very little effect (Table A1 and Figure A3B, and previous section), the realized ACL l_r and its interaction $l_r * v$ had a substantial effect (Table A2 and Figure 5). The generating and realized ACLs are correlated when l_g is very small, because very small-scale patchiness can be accurately reproduced with little distortion due to stochasticity, but when l_g is larger the correlation breaks down, because large-scale patchiness is stochastic at a scale commensurate with the scale of the landscape itself; the realized ACL of the landscape may therefore be quite different from the generating ACL (Figure A3A). Since the environmental gradient slope and curvature also affect the realized ACL, that is another reason why l_g and l_r are generally poorly correlated; for clarity, however, Figure A3A shows only values for landscapes with no slope or curvature, so that it depicts solely the loss of correlation for large values of l_g due to stochasticity in the landscape generation and system size effects.

The superiority of l_r to l_g in predicting branching is one aspect of the larger phenomenon that GAM_r, the best GAM based on the realized-landscape metrics, substantially outperformed GAM_g, the best model based on the landscape-generating parameters (Table A1 model #15 *vs*. Table A2 model #9). This was the case even though GAM_g contained 13 terms and used 167.0 effective degrees of freedom, compared to 7 terms and 49.3 effective degrees of freedom for GAM_r. For large l_g, stochasticity caused considerable variation not only in l_r, but also in the realized-landscape metrics σ, γ_1, and γ_2, for given values of the landscape-generating parameters s, c,

a, and l_g (Figure A4). Unsurprisingly, given this, the realized-landscape metrics are a better predictor of the effect of the landscape than are the landscape-generating parameters; the realized landscape is what the individuals in the model actually occupy, and so its properties are what matters. However, it is perhaps surprising that the simple metrics we used proved so effective; one might ponder whether different metrics could be even better. In any case, this result suggests that caution must be used whenever modeling heterogeneous landscapes, to be sure that the effect of the realized pattern of heterogeneity, and not solely the effect of the landscape-generating parameters, is analyzed, since the two may be quite different – an issue that has not, to our knowledge, been raised previously.

The realized-landscape metrics spanned a wide range of types of heterogeneity, although negative skewness, positive kurtosis, and short ACLs were relatively infrequent (Figure A6). Although this means that there is less data in some regions of the parameter space, the very large number of realizations analyzed mitigates this problem, except where specifically noted.

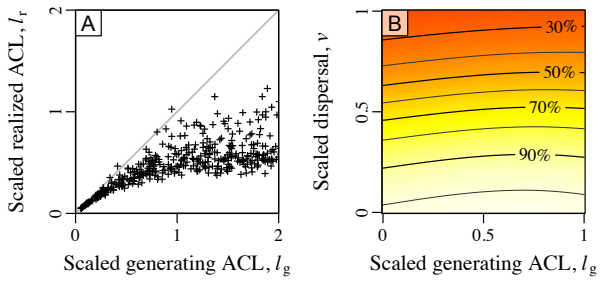

Figure A3: Comparison of the scaled generating autocorrelation length (ACL) l_g to the scaled realized ACL l_r. A: Scatter plot of l_g versus l_r for 500 randomly chosen landscapes; the gray line shows the 1:1 line at which $l_g = l_r$. The deviation of l_g from l_r illustrates that autocorrelation lengths that are large relative to the spatial scale of the landscape can be unrepresentable because they correspond to spatial features approaching or exceeding the scale of the landscape itself. B: Predicted branching probabilities from GAM_g, showing the effects of l_g and the scaled dispersal distance v; compare the weak relationships shown here (little effect of l_g and little interaction between l_g and v) to the stronger effects from GAM_r shown in Figure 5. The color scale is as shown in Figure 5. Results shown are for reprising boundaries; results for other boundary conditions are qualitatively similar (not shown). Other parameters: $s = 0$, $c = 0$, $a = 1.5$. Contour lines above 90% are placed to best show the contours of the data.

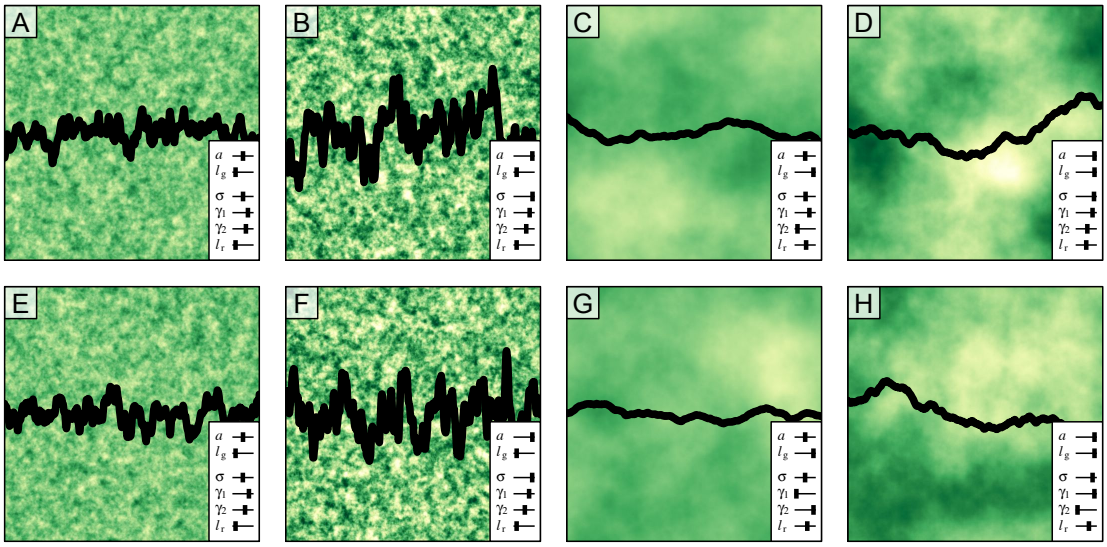

Figure A4: Examples of generated landscapes. These examples provide an illustration of the variability in realized-landscape metrics resulting from different types of spatially continuous patchiness. Vertical pairs of panels used the same landscape-generating parameters (scaled amplitude a and scaled generating ACL l_g; $s = 0$ and $c = 0$ for all panels), but used different random number generator seeds and thus differ. Sliders in each panel show landscape-generating parameters (a, l_g) and realized-landscape metrics (σ, γ_1, γ_2, l_r) on an arbitrary scale. Colors indicate locally optimal ecological trait values, ranging from white (low) to dark green (high). Black lines show locally optimal ecological trait values across one horizontal transect of each landscape. Panels A, B, E, F used short l_g while panels C, D, G, H used long l_g. Panels A, C, E, G used intermediate a while panels B, D, F, H used large a. These results illustrate that landscapes that share generating parameters will also share similar realized metrics if l_g is small, but when l_g is large the realized metrics can be quite divergent because of the stochastic patterns at large spatial scales in the landscapes generated.

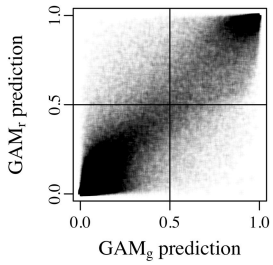

Figure A5: Correlation between the predictions of the two GAM models, GAM_g and GAM_r. A total of 300,000 points are plotted; each point represents one realization, showing the predicted probability of branching from the GAM_g model (*x*-axis) and the GAM_r model (*y*-axis). The lower left quadrant, in which both models predict no branching, contains 78.9% of realizations, and the upper right quadrant, in which both models predict branching, contains 15.9% of realizations; the two models thus made the same prediction in 94.8% of realizations. The upper left quadrant, in which GAM_r predicts branching but GAM_g does not, contains 2.8% of realizations, while the lower right quadrant, in which GAM_r does not predict branching but GAM_g does, contains 2.4% of realizations. Note also that for the quadrants in which the models agree, both models tend to be relatively certain about their predictions (predicted probabilities near 0.0 or 1.0), while predictions are generally much less certain for the quadrants in which the models disagree (predicted probabilities near 0.5). Results shown are for reprising boundaries; results are very similar for other boundary conditions, with small differences in the proportion of realizations predicted to branch (see Table A3).

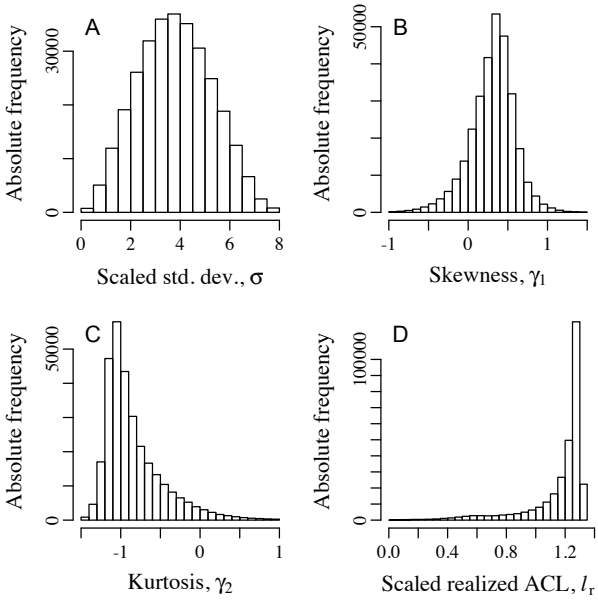

Figure A6: Distributions of the realized-landscape metrics. A: Scaled standard deviation, σ. B: Skewness, γ_1. C: Excess kurtosis, γ_2. D: Scaled realized ACL, l_r. A small number of outliers are omitted in panels A–C to allow the remaining data to be better displayed. Distributions shown encompass all landscapes generated in realizations utilizing reprising boundary conditions; since landscape generation is independent of the boundary conditions chosen, this is a representative sample. Note that although the landscape-generating parameters were drawn from uniform distributions (Methods: *Model realizations*), the realized-landscape metrics were not uniformly distributed, and so the realized metric parameter space was not uniformly sampled. Nevertheless coverage was extensive due to the large number of realizations conducted; more than 1000 realizations had $l_r < 0.2$, for example.

Effects of boundary conditions

Some basic metrics comparing the four tested boundary conditions are shown in Table A3. Boundary condition effects generally followed the findings of previous research (Mazzucco et al., unpublished manuscript). In particular, the stopping boundary condition produced the highest branching propensity due to its effect of generating disruptive selection, while the absorbing boundary condition produced the lowest branching propensity due to its generation of stabilizing selection. Similarly, absorbing boundaries produced branching only for relatively short dispersal distances, because long-range dispersal was more likely to be lethal, while stopping boundaries produced branching even for large dispersal distances, because long-range dispersal predictably stopped individuals at the edge of the landscape. Effects of the boundary condition on the distribution of branching times followed the same pattern, with the modal time to branching being longest with absorbing boundaries and shortest with stopping boundaries, although the differences were slight (see *Distribution of branching times* and *Correlations with branching times*).

Reprising boundaries were chosen for our main results as a middle ground, minimizing the influence of the landscape boundary and thus best approximating the effect expected on a landscape of infinite extent. Stopping boundaries cause an accumulation of individuals near the boundary, whereas absorbing boundaries cause a deficit of individuals near the boundary, and these density effects influence the branching propensity. Reprising boundaries do not bias the branching propensity through such density effects, and thus insights from

this case are likely to apply generally – even when the density effects of other boundary conditions are added in. Reflecting boundaries would have been a similarly reasonable choice, and showed very similar effects in our analysis. Minimization of boundary effects is preferable for theoretical purposes, to maximally reveal the effects of model parameters without any obscuring influence from the model's assumptions; however, when comparing our model results to a particular biological system, the boundary condition should be chosen that best matches the empirical dispersal characteristics of the system.

The effects of different types of heterogeneity on branching are qualitatively similar under different boundary conditions (Figure A7). Interestingly, the intermediate optimum patchiness amplitude is most apparent for absorbing boundaries and least apparent for stopping boundaries. A secondary, weaker optimal magnitude of curvature is also observed for all boundary conditions except absorbing; this may reflect the degree of curvature that produces an optimal difference between the environments at the extreme left and right edges of the landscape, since the landscape edges can be exploited as niches with non-absorbing boundary conditions (particularly stopping boundaries, for which this secondary optimum is strongest).

The interactions between different types of heterogeneity are also qualitatively similar under different boundary conditions (Figure A8). The particular contour line shape that indicates the refugium effect (Results: *Additivity and the refugium effect*) appears strongest with reflecting boundaries for the interaction of slope and amplitude, and strongest with reprising boundaries for the interaction of curvature and amplitude, but it is visible for all boundary conditions; this suggests that the refugium effect is general.

The interaction between realized ACL and dispersal distance is shown for different boundary conditions in Figure A9. The insensitivity of branching to the dispersal distance for short realized ACL is most apparent for stopping boundaries, and least apparent for absorbing boundaries; this may be a simple result of dispersal distance being least important for stopping boundaries (since dispersal often stops at the edge regardless of the "intended" dispersal distance), and most important for absorbing boundaries (since dispersing too far is lethal). In any case, the qualitative pattern of interaction between the realized ACL and the dispersal distance is similar for all boundary conditions.

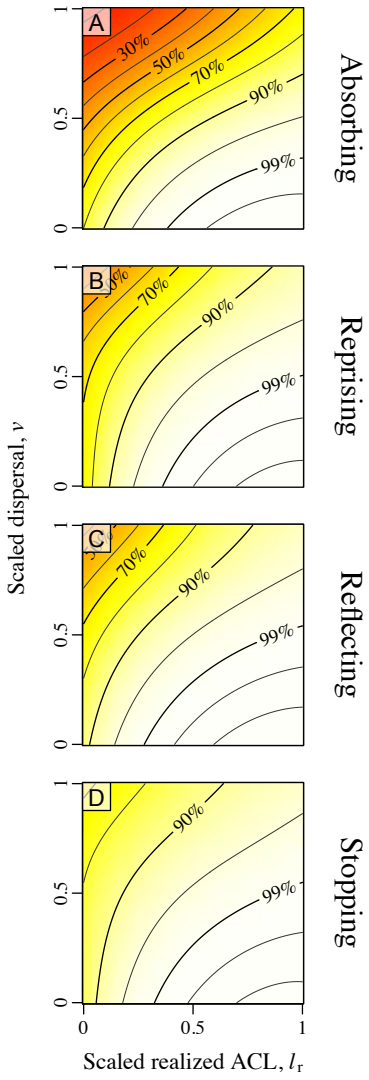

Figure A9: Predicted branching probabilities from fitted GAMs, showing a comparison among boundary conditions in their interaction between realized autocorrelation length (ACL) and dispersal distance. Each row of this figure follows Figure 5, but is plotted for a particular boundary condition; the second row, for reprising boundaries, thus matches Figure 5 exactly. Contour lines show the predicted branching probability; the color scale is as shown in Figure 5. Branching is always promoted by large realized ACL and by short dispersal distance. An interaction between v and l_r exists for all boundary conditions, causing a decrease in the effect of v for small l_r, but this interaction varies in strength among boundary conditions.

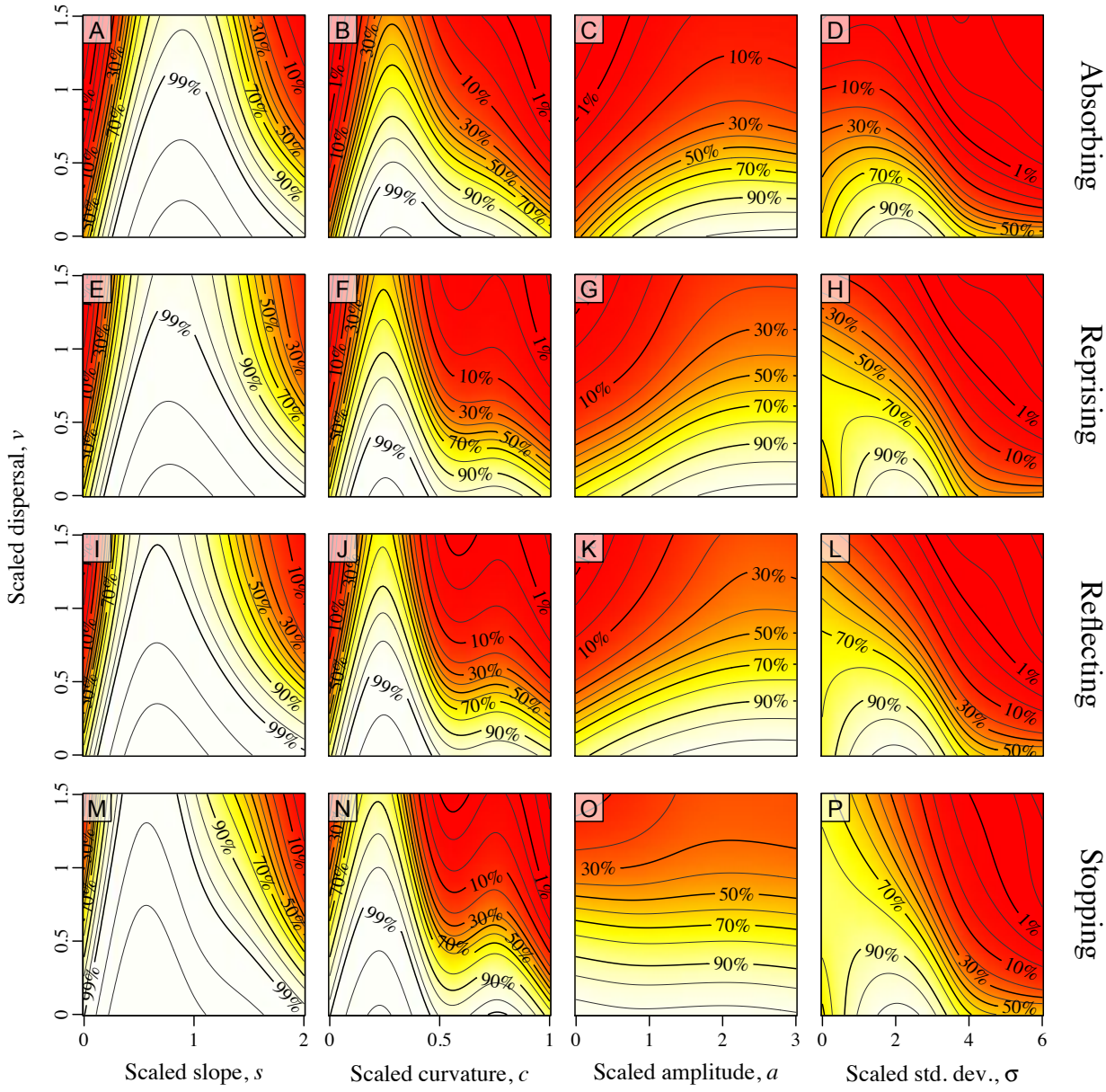

Figure A7: Predicted branching probabilities from fitted GAMs, showing a comparison among boundary conditions in the effects of scaled dispersal distance and different heterogeneity types. Each row of this figure follows Figure 3, but is plotted for a particular boundary condition; the second row, for reprising boundaries, thus matches Figure 3 exactly. Contour lines show the predicted branching probability; the color scale is as shown in Figure 3. The probability of branching is maximized at an intermediate value of s and c for all boundary conditions, although this optimum value varies among boundary conditions. Similarly, an intermediate optimum for σ exists for all boundary conditions, but typically only for smaller values of v. For a, an intermediate optimum that maximizes branching clearly exists for absorbing boundaries, and may also exist for reflecting and reprising boundary conditions (but since the optimum in these cases is close to the edge of the parameter space explored, this is uncertain); no well-defined intermediate optimum for a appears to exist for stopping boundaries.

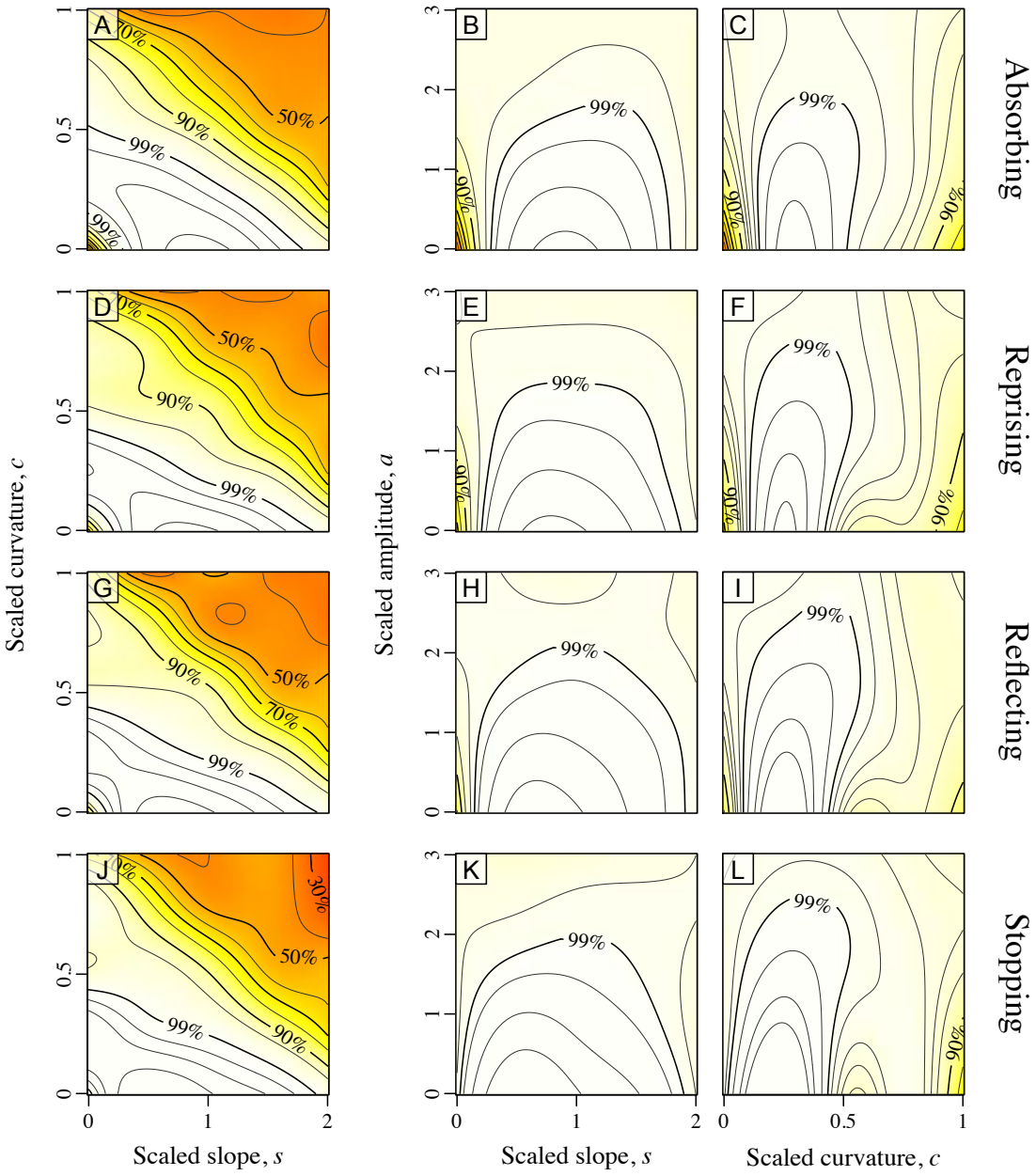

Figure A8: Predicted branching probabilities from fitted GAMs, showing a comparison among boundary conditions in their interactions between different heterogeneity types. Each row of this figure follows Figure 4, but is plotted for a particular boundary condition; the second row, for reprising boundaries, thus matches Figure 4 exactly. Contour lines show the predicted branching probability; the color scale is as shown in Figure 4. Additivity of the effects of s and c is apparent for all boundary conditions, as is the lack of additivity of s and a, and of c and a, due to the refugium effect. The refugium effect, which causes mitigation of the negative effect on branching of large s and large c due to increasing amplitude a, is generally visible; however, the effect varies in strength, and is always stronger with c than with s.

Effects of skewness and kurtosis

The bias towards positive skewness values among our generated landscapes (Figure A6B) is due to the fact that curvature, as implemented in our model as increasing gradient slope from left to right, leads to positive skewness. For this reason, the causes of positive skewness and negative skewness in our model are not symmetric; positive skewness is caused by both patchiness and curvature, while negative skewness is caused by patchiness alone. These causes would not segregate in this manner for real environments; empirically, the meaning of the sign of the skewness depends simply on the ordination of environmental values (whether a cold environment is given a value less than, or greater than, a hot environment, for example). Since the effect of skewness was found to be reasonably symmetric around zero (Figure A10), however, this concern can likely be ignored. In general, then, our results indicate that large skewness, whether caused by gradient nonlinearity or by an uneven pattern of patchiness, hinders branching; the optimal level of skewness appears to be close to zero (Figure A10).

Most kurtosis values were negative, with a peak near −1 (Figure A6C), indicating that platykurtic distributions of environmental conditions were much more common than leptokurtic distributions. This is unsurprising, since an environmental gradient is inherently platykurtic, resembling a uniform distribution (kurtosis −1.2) much more than a normal distribution (kurtosis 0), not to mention a more leptokurtic distribution such as a Laplace distribution (kurtosis 3). To a first approximation, then, the landscape kurtosis may be thought of as representing the degree to which the pattern of heterogeneity of the landscape is dominated by a linear gradient that produces a uniform distribution of environmental values. It is interesting, therefore, that more negative values of kurtosis hindered branching (Figure A10). We tentatively interpret this as a different angle on the same refugium effect discussed previously (Discussion: *Additivity and the refugium effect*), since it shows that a linear gradient with no other heterogeneity (more negative kurtosis values) is less effective in promoting branching than more complexly heterogeneous landscapes.

Interestingly, then, our results suggest that a normal distribution of environmental values (with both skewness and kurtosis close to zero) may be close to optimal for evolutionary branching. The spatial distribution of those values is, of course, also important, however, and so broad conclusions regarding the optimal pattern of environmental heterogeneity for the promotion of evolutionary branching will require further investigation.

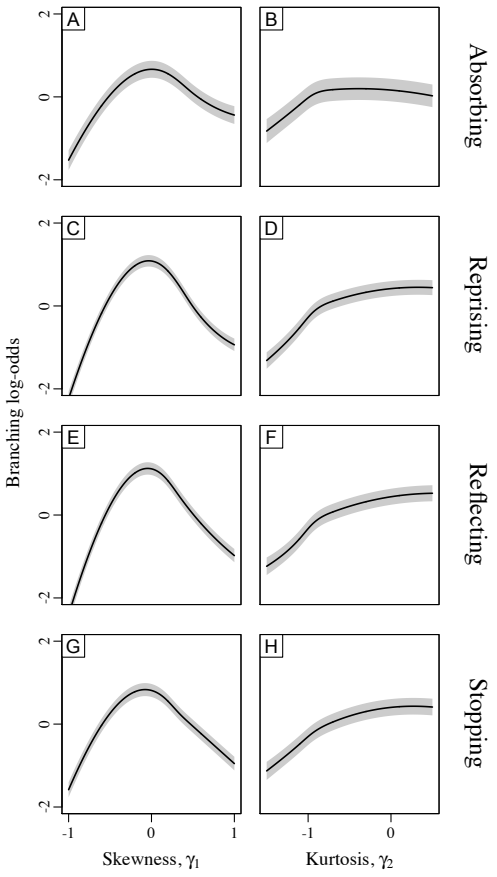

Figure A10: Predicted branching propensities from fitted GAMs, showing a comparison among boundary conditions in the effects of skewness and excess kurtosis. Branching propensity is shown as the branching log-odds (the logit of the branching probability); positive log-odds indicate relative promotion of branching, while negative log-odds indicate relative hindrance of branching. The curve in each panel is re-centered on a mean y value of zero, removing the absolute differences in branching propensity among boundary conditions, for ease of interpretation. Gray shaded bands indicate standard error confidence intervals. An intermediate skewness maximally promotes branching for all boundary conditions. The branching propensity generally increases with increasing kurtosis, although this effect saturates and even, for absorbing boundary conditions, declines at high kurtosis values. Other parameter values are unspecified, since no interaction terms involving skewness and kurtosis are included in the GAM models.

Alternative assessment of branching

As described previously (*Model description*, OBSERVABLES), a heuristic analysis of the distribution of ecological trait values over the final 3000 generations was used to determine whether evolutionary branching had occurred. An alternative assessment of branching was also conducted using a Gaussian mixture model cluster analysis performed on the ecological character values from the end-of-realization census data, with the `Mclust()` function of the `R` package `mclust` (Fraley and Raftery 2003). A Bayesian prior was used on the variance of the cluster analysis, with a `scale` parameter value of 0.5 for the `priorControl()`, determined by trial and error to provide a good detection of clustering as compared to a visual determination.

For the detection of branching, the heuristic method and the cluster analysis method produced similar results, except that cluster analysis, since it considered only the pattern of phenotypes in the final generation of a realization, overestimated the branching rate slightly (Table A3). Despite their very different mathematical bases, the level of agreement between the two methods was quite high, giving some assurance that our results were not a consequence, in some way, of our method of assessing branching. Furthermore, GAM models fit using the cluster analysis branching assessment provided a close match to the GAM models presented.

One advantage of the cluster analysis method is that it produces not just a binary assessment of whether branching did or did not occur, but a count of the number of clusters. This ranged from 1 to 9, for the realizations conducted; typically, however, even those realizations that did branch produced only 2 or 3 clusters (Table A3), with the number of clusters produced following roughly a negative exponential distribution across all realizations (not shown). Interestingly, the number of clusters observed was negatively correlated with branching times, suggesting that landscapes that can support a larger number of species also produce more rapid evolutionary branching (see *Correlations with branching times*). Ordinal multinomial GLM fits (not shown) indicated that the factors causing a high branching propensity are the same as the factors causing a large number of clusters, suggesting that adaptive radiation in our model is simply adaptive branching "writ large"; the two metrics are simply different views on the propensity for adaptive diversification.

Distribution of branching times

For each realization that was classified as branched, we also determined the branching time using a heuristic algorithm that looked backwards from the end of the realization until finding a generation without a bimodal (or multimodal) phenotypic distribution. Since the phenotypic distribution was saved every ten generations (see *Model description*, OBSERVABLES), the branching time was then estimated as the generation found, plus five (intermediate between the last unbranched and the first branched generations). This determination was generally robust for a wide variety of phenotypic histories, although a small number (~2%) could not be assigned a well-defined branching time.

The distribution of branching times for branched realizations using reprising boundary conditions is shown in Figure A11A. Most realizations that branched did so quite early – within the first ~1000 generations – as indicated by the strong modal peak of the distribution. The distribution also has a long tail that asymptotically approaches a non-zero constant value. This non-zero offset probably represents realizations that branched only after establishing a favorable spatial configuration; however, there is currently no good theory for spatial branching processes that would allow us to quantify this contribution (see Discussion: *Future directions*). Finally, the bump at the end of the distribution, comprised of realizations evaluated to have branched at close to 5000 generations, represents noise in the detection of branching times caused by stochastic demographic effects. We fitted lognormal and exponential curves with a vertical offset added, using branching times less than 4500 generations to exclude the final bump (Figure A11B). The exponential fit is poor, suggesting that branching is not a Poisson process; the lognormal fit is clearly better, and indeed, fits the distribution quite closely. In fact this result is expected, since branching is the result of many independent random events that combine multiplicatively (mutation,

dispersal, survival). Wakano and Iwasa (2013) showed similar results with a substantially different model (their Figure 9).

The predicted probability of branching had no effect on the modal branching time (Figures A11C and A11D). This indicates that branching is essentially a binary process once a favorable spatial configuration is reached: it happens early or does not happen at all. However, the mean and median branching times were negatively correlated with the predicted probability of branching (Figures A11C and A11D). This was due to a decreased frequency of early branching for realizations with low predicted branching probabilities, and thus an increased influence of the late-branching realizations over the mean and median branching time (see the next section, *Correlations with branching times*).

Figure A11 shows results only for reprising boundary conditions; results varied slightly for other boundary conditions, but showed qualitatively identical patterns. In particular, the modal time to branching estimated by the lognormal fits was slightly longer for absorbing boundaries (153 generations), slightly shorter for stopping boundaries (126 generations), and intermediate for reprising and reflecting boundaries (136 and 134, respectively). These results are in accord with the other effects of the boundary conditions (see *Effects of boundary conditions*). However, the general pattern that most realizations branched early or did not branch at all, discussed above, held across boundary conditions; the difference among boundary conditions in branching propensity is thus not an effect of these differences in the modal waiting time.

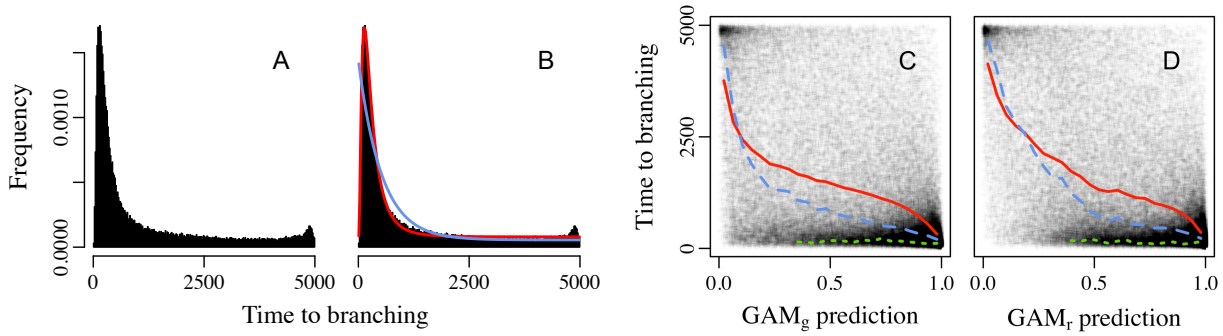

Figure A11: Distribution of branching times for branched realizations, and its relationship to the predicted probability of branching. A: The distribution of branching times for branched realizations, showing a strong peak indicating that branching typically occurs in the first ~1000 generations, with a long tail asymptotically approaching a non-zero limit, and a final bump (see text). B: The same distribution, overlaid with best fit lognormal (red) and exponential (blue) curves that include a height offset parameter. Curves were fitted to the branching times < 4500, to exclude the final bump (since it is an analysis artifact; see text). C–D: Scatter plots showing the prediction of GAM$_g$ (C) and GAM$_r$ (D) versus the branching time for branched realizations. Red solid, blue dashed, and green dotted curves respectively show the mean, median, and mode of 25 bins of equal x span; the mode is shown only when the early-branching peak is well-defined. The negative relationship between predicted branching probability and branching time shown by the mean and median curves is due to a progressive shift from early branching with a consistent modal branching time unrelated to the predicted probability (predicted branching probabilities > ~0.4) to less frequent early branching and thus a higher proportion of realizations with longer branching times (the long tail; predicted branching probabilities < ~0.4), and then into a regime in which the rare detection of branching is usually an analysis artifact (the final bump; predicted branching probabilities < ~0.1). Results shown are for reprising boundaries; qualitative results were the same for other boundary conditions (see text).

Correlations with branching times

Mean and median branching times correlated with some parameters and metrics; the modal branching time, however, was nearly invariant (Figure A12). These qualitative differences between the mean and median compared to the mode are due to changes in the proportion of realizations that branched early (the modal peak of the branching time distribution) versus realizations that branched later (the long tail of the distribution; see *Distribution of branching times*). Large mean/median branching times may thus perhaps be understood as indicating that a high proportion of realizations had to establish a favorable spatial configuration prior to branching (see *Distribution of branching times*). The effects

presented are for reprising boundaries; results were qualitatively identical for other boundary conditions with the exception of an interaction between the boundary condition and the effect of dispersal distance, discussed below.

Increasing slope increased the mean/median branching time (Figure A12A), as did increasing curvature (Figure A12B) and increasing realized-landscape heterogeneity (Figure A12E). It is worth noting that the mean/median branching time was not minimized at an intermediate value of these parameters (in contrast to Results: *Intermediate heterogeneity maximizes branching propensity*); this is another view on the result that branching time and branching probability are not closely related phenomena in our model. Interestingly, increasing patchiness amplitude had no effect on mean/median branching time, unlike the other types of heterogeneity (Figure A12C).

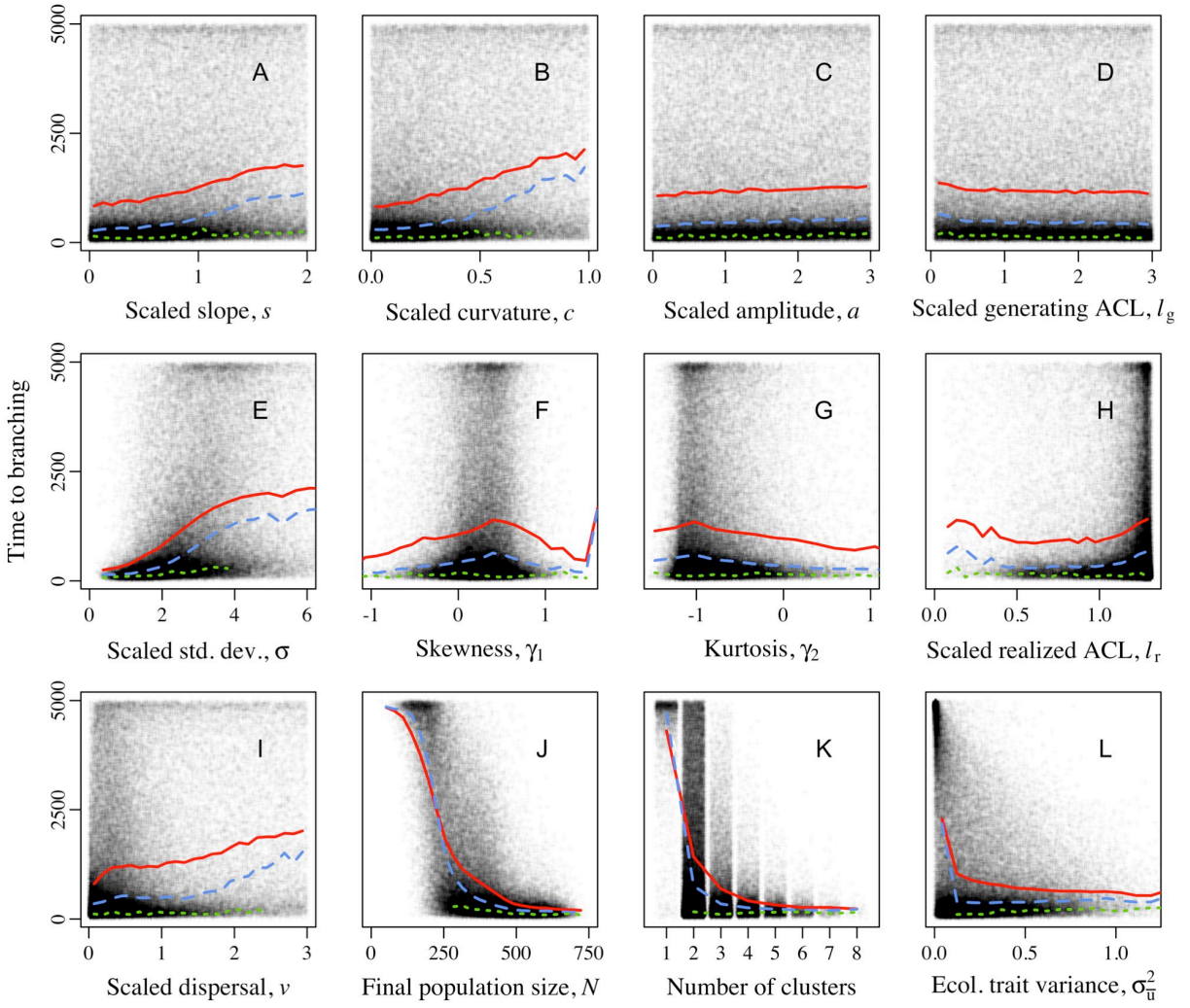

Figure A12: Correlations between branching times for branched realizations and landscape-generating parameters (top row), realized-landscape metrics (middle row), and other parameters and metrics (bottom row). Red solid, blue dashed, and green dotted curves respectively show the mean, median, and mode of 25 bins of equal x span (8 bins, for panel K); the mode is shown only when the early-branching peak is well-defined. The position of the early-branching peak was nearly constant in all panels, but positive relationships were observed between the mean/median branching time and several parameters and metrics: slope (A), curvature (B), realized-landscape standard deviation (E) and autocorrelation length (H), and dispersal distance (I). Horizontal jitter was added to panel K to separate points. Results shown are for reprising boundaries; qualitative results were the same for other boundary conditions, except that the strength of the effect of dispersal (panel I) varied (see text).

Large realized ACL increased the mean/median branching time (Figure A12H), presumably because the scale of heterogeneity was then so large that it often took populations a long time simply to reach a part of the environment that was ecologically different enough to produce branching. The generating ACL had no effect (Figure A12D), as might be expected (see *Landscape-generating parameters versus realized-landscape metrics*).

The mean/median branching time increased with increasing dispersal (Figure A12I). This effect was weak with reprising boundaries (Figure A12I) and reflecting boundaries, stronger with absorbing boundaries, and entirely absent with stopping boundaries (not shown). This variation appears to be driven by the interaction between dispersal and the boundary condition (see *Effects of boundary conditions*). In particular, long-distance dispersal with stopping boundaries does not hinder branching; whether dispersal is short or long, realizations that are going to branch typically do so early (see *Distribution of branching times*), and so the relationship between dispersal distance and mean/median branching time is flat. With absorbing boundaries, long-distance dispersal is lethal and thus early branching is hindered; late branching, however, can still (very rarely) occur.

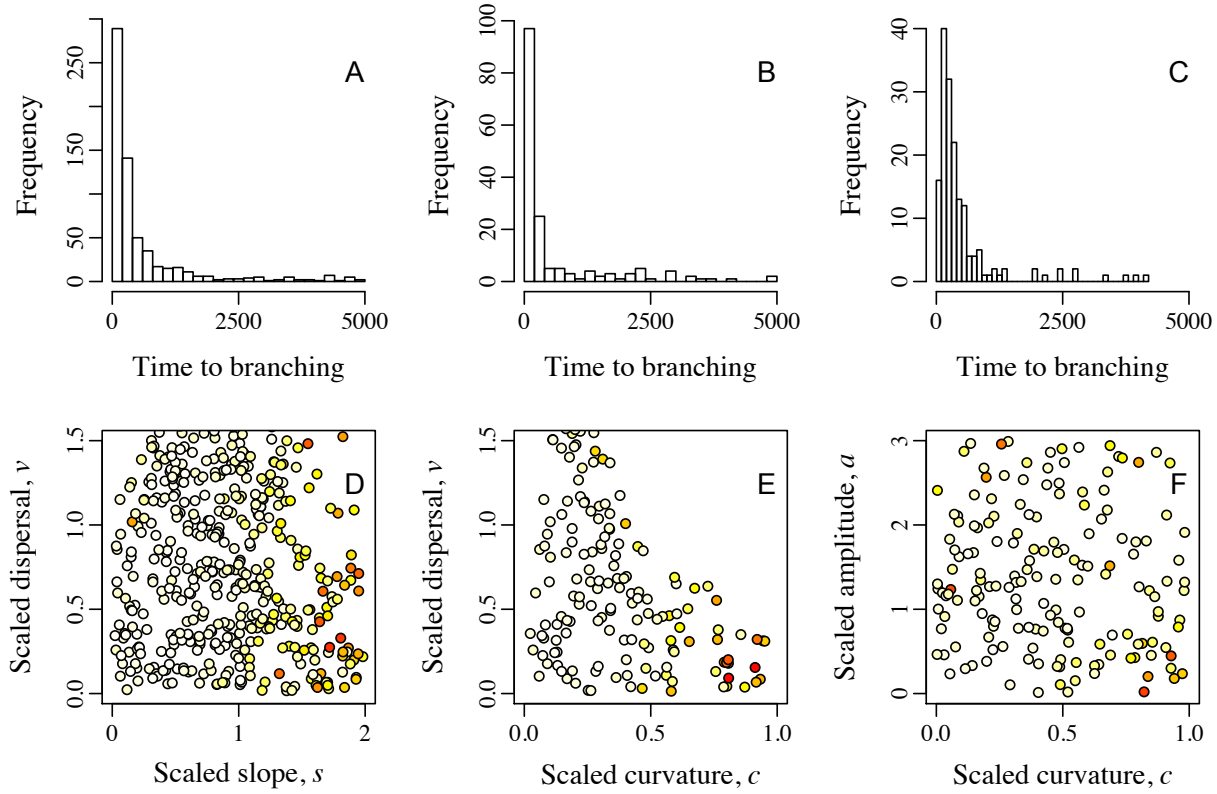

Figure A13: Alternate view of the parameter-region slices shown in Figures 3A (left column), 3B (middle column), and 4C (right column), showing branching times for branched realizations near those slices. A–C: distributions of branching times for each set of realizations, showing similarity to the overall distribution shown in Figure A11A. D–F: each set of realizations plotted following the axes of Figures 3A, 3B, and 4C; colors indicate branching time, from white (0 generations) to red (5000 generations). Comparison to Figures 3A, 3B, and 4C shows that intermediate probabilities of branching do not generally correspond to long branching times, although a weak association may exist in panel F. Instead, large slope and/or large curvature appear to cause long branching times (see also Figure A12). Parameter ranges for these plots: A/D: $a < 0.1$, $c < 0.1$, 631 realizations; B/E: $a < 0.1$, $s < 0.1$, 166 realizations; C/F: $s < 0.2$, $l_g < 0.3$, $v < 0.225$, 166 realizations (parameters not listed were unconstrained to maximize the number of realizations plotted). Results shown are for reprising boundaries; qualitative results are the same for other boundary conditions.

Finally, short mean/median branching times were associated with large population size (Figure A12J), a large number of phenotypic clusters (Figure A12K), and a large ecological trait variance (Figure A12L). These associations might appear to be almost tautological: the faster branching is, the more clusters are produced at equilibrium, the more variance there is in the ecological trait, and the larger population size can be (due to relaxed competition resulting from greater phenotypic variance). However, it could have been otherwise; if it were true here that "slow and steady wins the race", then realizations with long branching times might have ultimately produced the most variation. Our result suggests instead that radiations producing a large final clade size should be expected to proceed more rapidly than smaller radiations. This may fit with empirical results, since some of the largest adaptive radiations known are believed to have occurred in a short time, such as the radiations of cichlids in African lakes (Seehausen 2002), of plants in the Andes (Hughes and Eastwood 2006) and in the Cape flora of South Africa (Richardson et al. 2001), and other cases (Schluter 2000); however, the generality of this pattern is not clear.

Figure A13 shows these results regarding branching times in the light of three particular cases, following three figures in the main text (Figures 3A, 3B, and 4C). Comparison of Figure A13 to those figures shows again that branching time and branching probability were unrelated in our model, but that branching time increased with increasing environmental gradient slope and curvature.

Complex landscapes

Some previous work has been done regarding the generation of complexly heterogeneous landscapes. A method similar to ours was briefly outlined by Cliff and Ord (1973; Appendix 1) but was not fully specified. Keitt (2000) describes a variety of methods for landscape generation; our method is essentially a special case of one of these. Similar ideas have also been developed in other fields, such as materials science (Hu and Tonder 1992; Wu 2000; Uchidate et al. 2010). However, these papers are not generally accessible to non-mathematicians; we thus felt there would be value in presenting our landscape generation algorithm explicitly.

A few previous models have looked at ecological and/or evolutionary dynamics on complex two-dimensional landscapes. However, the landscapes of these studies differ from ours in several important ways. In some models, the landscape consists of only two types of terrain, habitable and uninhabitable (e.g., Hiebeler 2000; Bonte et al. 2010), or two discrete resource types distributed heterogeneously (e.g., North et al. 2011b; Karonen 2012). In others, the environment does not change smoothly from one type to another, but instead exhibits abrupt transitions between types (e.g., Birand et al. 2012). Commonly, the landscapes modeled are considered to vary in suitability in some absolute sense, producing areas that are more habitable and areas that are less habitable, rather than allowing local adaptation to the different environment types to occur (e.g., North and Ovaskainen 2007; Stoddard 2010; North et al. 2011a). As far as we are aware, our model is unique in its incorporation of both spatially and ecologically continuous complex spatial heterogeneity that determines the locally optimal phenotype rather than the habitability *per se*.

Landscape generation

Given parameters S, C, L_g, and A, the landscapes used in our model are constructed in a six-step process. These steps are explained here.

First, a patchiness matrix with resolution sufficient to represent patchiness with autocorrelation length L_g (confirmed by post-analysis of the generated landscape) is filled with random values drawn from a Gaussian distribution with mean 0 and standard deviation A.

Second, this patchiness matrix is convoluted with a filter function (a modified Bessel function of the second kind, order 0) using Fourier transforms. This yields the desired periodic properties for the y-direction; in the x-direction the patchiness matrix is padded prior to convolution to remove the correlation that would otherwise result between the left and right edges.

Third, an environmental gradient matrix is constructed, of the same resolution as the patchiness matrix, expressing the linear and quadratic gradients as a function $e(x) = Sx + \frac{1}{2}Cx^2$.

Fourth, the final patchiness matrix and the environmental gradient matrix are added together to produce the preliminary landscape matrix.

Fifth, the preliminary landscape matrix is adjusted by adding a constant $\alpha = 0.5 - \mu_{env}$, where μ_{env} is the mean of the preliminary landscape matrix; this re-centers the matrix to have a mean value of 0.5, so that the initial population is adapted to the mean environmental value (see *Model description*, PHENOTYPE RANGES AND INITIALIZATION).

Sixth and last, bilinear interpolation is used to construct the canonical landscape matrix. The size of this matrix is constant, 512 by 512, for ease of implementation. This is a sufficiently high resolution to represent with high fidelity all landscapes generated by the values of S, C, L_g, and A used in this study. Values of $u_0(x, y)$ were obtained from the canonical landscape matrix, scaled to exactly span the dimensions $0 \le x, y \le 1$ of the landscape.

Landscape analysis

The canonical landscape matrix (see previous section, *Landscape generation*) was analyzed to derive metrics regarding the heterogeneity of the landscape. The scaled standard deviation σ, skewness γ_1, and excess kurtosis γ_2 were computed across all the values in the canonical landscape matrix, thus characterizing the distribution of environmental values represented in the landscape without regard to their spatial distribution. The standard deviation was scaled by σ_K (Table 1) in order to yield a dimensionless metric of the environmental heterogeneity, σ.

Additionally, a metric of the characteristic spatial scale of heterogeneity in the landscape was computed. First, the autocorrelation function (ACF) for the canonical landscape matrix was computed via the Wiener-Khinchin Theorem, using Fast Fourier Transforms (FFTs); specifically, the autocorrelation function is $F^{-1}\left(F(E)\overline{F}(E)\right)$, where F denotes the Fourier transform, F^{-1} the inverse transform, $\overline{F}$ the complex conjugate, and E the canonical landscape matrix. To eliminate the periodicity in the x-direction implicitly imposed by the Fourier transform, flipped copies of the landscape were pasted to its left and right edge. Given the ACF (normalized to a maximum value of 1), the autocorrelation length (ACL) was estimated as the lag at which the ACF first crossed below the threshold value $1/e$ (using linear interpolation to estimate the exact position of crossing, between the last lag above the threshold and the first lag below it). The lag was converted to the model coordinate system by dividing by 512, and was then divided by σ_s to yield the scaled realized ACL, l_r, of the landscape. This scaled realized ACL is a dimensionless metric comparable to the scaled generating ACL, l_g.

Several other ACL estimation methods were tried, including fitting an exponential curve to the ACF, and using estimates of Moran's I and Geary's c. Although these metrics were all strongly correlated with l_r, they produced less accurate predictions of the branching propensity of the landscape, indicating that they did not capture as much useful information as the $1/e$ crossing point. The $1/e$ crossing point was therefore used to estimate the realized ACL, l_r, in all subsequent analyses.

Patterns of evolutionary branching

In the analysis presented, we have focused on the effects of model parameters on the probability of branching during a realization. We have also presented data regarding the timing of branching (see *Distribution of branching times* and *Correlations with branching times*), and have briefly considered the number of branches generated (see *Alternative assessment of branching* and *Correlations with branching times*). Except for the example realization shown in Figure 2, however, we have thus far left unaddressed more qualitative questions: What does the typical pattern of branching in a realization look like? What do landscapes that cause branching, or do not cause branching, look like? How stable is the branching in a typical realization? What does the spatial distribution of branched and unbranched populations look like? What, if anything, differentiates landscapes that allow a single branching event from landscapes that allow larger adaptive radiations? To answer these questions, we have chosen a small selection of unbranched (Figure A14) and branched (Figure A15) realizations that illustrate different types of observed model dynamics.

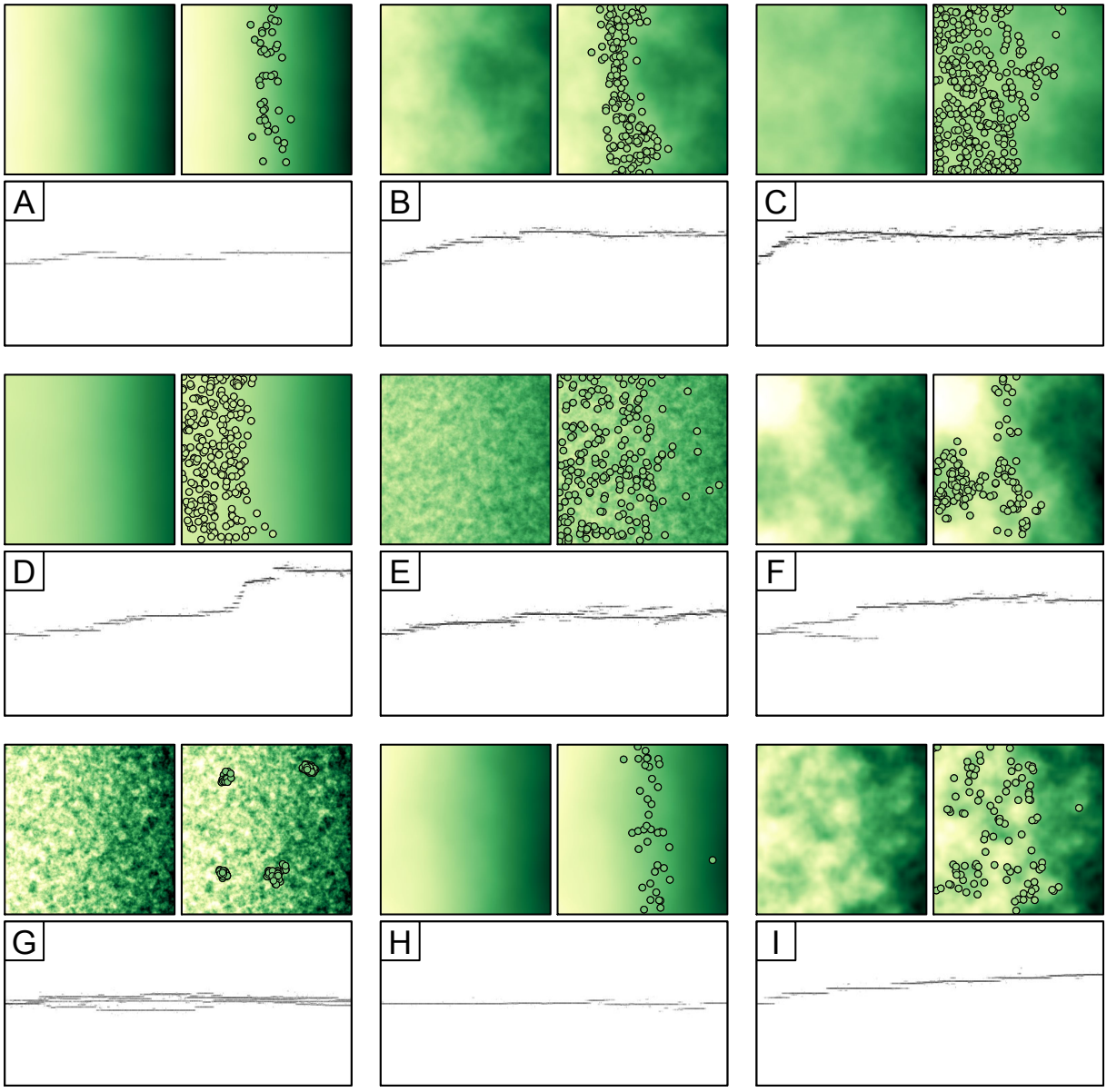

Figure A14: A selection of realizations that did not branch. Each panel is composed of three subpanels: the landscape (upper left), the end-of-realization population census (upper right), and the evolutionary history (bottom). Landscape colors indicate locally optimal ecological trait values, while circle colors indicate the actual trait values of individuals, both ranging from white (low) to dark green (high). The evolutionary history shows ecological trait values (y-axis) through time (from the beginning of the realization, on the left, to the end at 5000 generations, on the right). A: a steep gradient can inhibit branching; note the small population size due to low mean fitness. B: a shallower gradient allows a larger population size and higher genetic variability. C: a relatively flat landscape with mild patchiness allows even larger population size and genetic variability. D: curvature pushes the population to the left, toward the shallower end of the gradient. E: fine-grained heterogeneity promotes branching, but branches may be ephemeral. F: strong patchiness promotes branching, but it can be difficult to persist; note how the population's distribution fits to the shape of the landscape. G: very short-range dispersal creates highly localized clusters, but with extreme heterogeneity even this may be insufficient to allow branching. H: A steep gradient hinders branching; note the lone colonist on the right, which will surely die. I: Strong patchiness also hinders branching, but refugia may mitigate this; note the lone colonist on the right, which might establish a foothold and found a new branch.

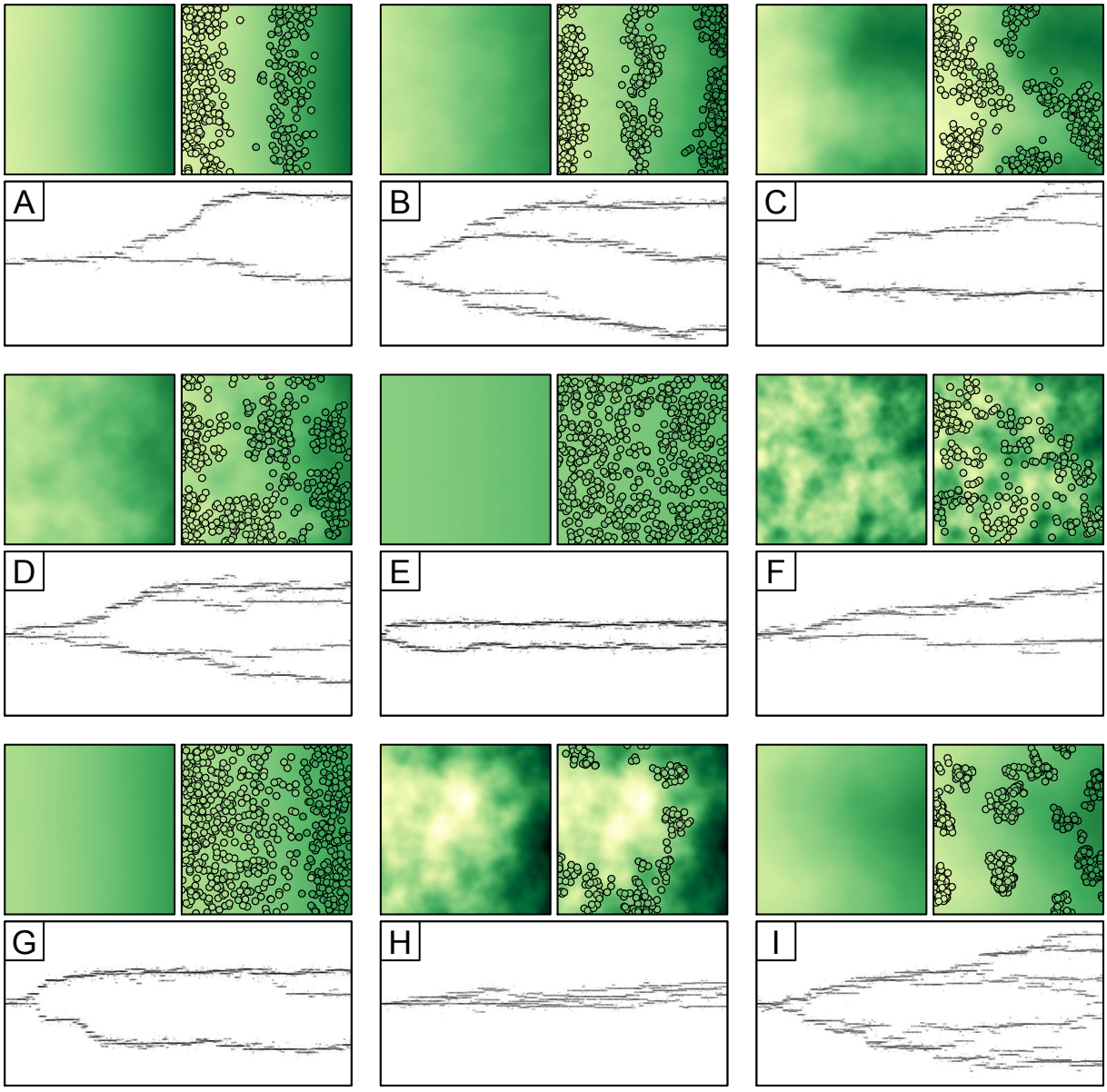

Figure A15: A selection of realizations that branched. Each panel is composed of three subpanels: the landscape (upper left), the end-of-realization population census (upper right), and the evolutionary history (bottom). Landscape colors indicate locally optimal ecological trait values, while circle colors indicate the actual trait values of individuals, both ranging from white (low) to dark green (high). The evolutionary history shows ecological trait values (y-axis) through time (from the beginning of the realization, on the left, to the end at 5000 generations, on the right). A: simple branching on a gradient; note that branching was somewhat delayed. B: sometimes several branches can become established; however, note the transient branch that went extinct. C: large-scale patchiness creates local gradients that can promote branching. D: medium-scale patchiness can support several locally-adapted groups. E: close to pure competitive branching (subdivision of a homogeneous resource maintained by character displacement due to competition); a very slight gradient is present, however. F: strong patchiness can subdivide the landscape into interleaved, fragmented habitat types. G: branching with long-range dispersal on a curved gradient; note that the flatter part of the gradient supports a larger population, allowing further branching to occur (although it may not persist). H: very strong patchiness can restrict the population to a single zone, just as with a very steep gradient. I: A rare example of extreme adaptive radiation, the product of an optimally intermediate degree of heterogeneity, a short dispersal distance, and luck.

Beyond remarks on individual realizations (see captions, Figure A14 and A15), some general observations can be made. Perhaps the broadest observation is that branching is a one-way process in this model; "reverse speciation" (Seehausen 2006) does not occur, although branches occasionally go extinct. This is likely for two reasons. First, the collapse of incipient branching due to non-assortative mating does not occur since this model is asexual; this might be expected to occur in a sexual version of the model, until the strength of assortative mating grew sufficiently strong to prevent hybridization (Felsenstein 1981; Dieckmann and Doebeli 1999; Doebeli and Dieckmann 2003). Second, the heterogeneity in this model is temporally invariant, and thus causes a constant selective pressure towards diversification; temporal variation in environmental heterogeneity, on the other hand, might be expected to lead to more ephemeral divergence (Seehausen 2006).

Another common pattern is that of evolutionary stasis (Estes and Arnold 2007) apart from branching and extinction ("punctuated equilibrium", Eldredge and Gould 1972; Gould and Eldredge 1977; Gould and Eldredge 1993). The phenotype of unbranched lineages often wanders over time due to drift (e.g., Figures A14B, A14D, A14I), particularly on landscapes with weak heterogeneity that provide some variety in local conditions, but not so much variation (e.g., Figures A14A, A14H) as to constrain evolution through strong selection. Branched lineages, on the other hand, more typically reach an equilibrium and exhibit stasis thenceforth (Figure A15) – although reaching equilibrium may take a long time, as populations jockey for position on both the physical landscape and on an adaptive landscape with peaks that shift due to frequency-dependent selection. Once equilibrium is reached, each branch exerts selection on every other branch, due to competition, and this stabilizes the system. Destabilization occurs, however, when a new branch arises, causing other branches to evolve away to avoid competitive pressure (e.g., Figures A15C, A15D). Similarly, when a branch goes extinct this may remove previously existing competitive pressures, allowing other branches to stop diverging (e.g., Figure A14F) or even evolve toward the now-unoccupied niche (e.g., Figure A15B). Nevertheless, after character displacement has equilibrated, stasis is generally again observed.

These qualitative patterns are not here developed to the point of rigorously shown results; rather we offer them as interesting observations, in the spirit of "natural philosophy", to spur further inquiry into less-studied aspects of the process of diversification in biological systems.

Literature Cited

Birand, A., A. Vose, and S. Gavrilets. 2012. Patterns of species ranges, speciation, and extinction. American Naturalist 179:1–21.

Bonte, D., T. Hovestadt, and H. J. Poethke. 2010. Evolution of dispersal polymorphism and local adaptation of dispersal distance in spatially structured landscapes. Oikos 119:560–566.

Cliff, A. D., and J. K. Ord. 1973. Spatial autocorrelation: Monographs in spatial and environmental systems analysis. Pion Limited, London.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354–357.

Doebeli, M., and U. Dieckmann. 2003. Speciation along environmental gradients. Nature 421:259–264.

Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: An alternative to phyletic gradualism, Pages 82–115 *in* T. J. M. Schopf, ed. Models in Paleobiology. Freeman, Cooper & Co., San Francisco.

Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. American Naturalist 169:227–244.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or Why are there so few kinds of animals? Evolution 35:124–138.

Fraley, C., and A. E. Raftery. 2003. Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. Journal of Classification 20:263–286.

Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22:403–434.

Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: The tempo and mode of evolution reconsidered. Paleobiology 3:115–151.

Gould, S. J., and N. Eldredge. 1993. Punctuated equilibrium comes of age. Nature 366:223–227.

Grimm, V. 2002. Visual debugging: A way of analyzing, understanding and communicating bottom-up simulation models in ecology. Natural Resource Modeling 15:23–38.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard et al. 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198:115–126.

Heinz, S. K., R. Mazzucco, and U. Dieckmann. 2009. Speciation and the evolution of dispersal along environmental gradients. Evolutionary Ecology 23:53–70.

Hiebeler, D. 2000. Populations on fragmented landscapes with spatially structured heterogeneities: Landscape generation and local dispersal. Ecology 81:1629–1641.

Hu, Y. Z., and K. Tonder. 1992. Simulation of 3-D random rough surface by 2-D digital filter and Fourier analysis. International Journal of Machine Tools & Manufacture 32:83–90.

Hughes, C., and R. Eastwood. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Sciences of the United States of America 103:10334–10339.

Karonen, I. 2012. Stable trimorphic coexistence in a lattice model of spatial competition with two site types. Journal of Theoretical Biology 295:77–85.

Keitt, T. H. 2000. Spectral representation of neutral landscapes. Landscape Ecology 15:479–493.

Mazzucco, R., M. Doebeli, and U. Dieckmann. (unpublished manuscript). The influence of habitat boundaries on evolutionary branching along environmental gradients.

North, A., S. Cornell, and O. Ovaskainen. 2011a. Evolutionary responses of dispersal distance to landscape structure and habitat loss. Evolution 65:1739–1751.

North, A., and O. Ovaskainen. 2007. Interactions between dispersal, competition, and landscape heterogeneity. Oikos 116:1106–1119.

North, A., J. Pennanen, O. Ovaskainen, and A. L. Laine. 2011b. Local adaptation in a changing world: The roles of gene-flow, mutation, and sexual reproduction. Evolution 65:79–89.

Richardson, J. E., F. M. Weitz, M. F. Fay, Q. C. B. Cronk, H. P. Linder, G. Reeves, and M. W. Chase. 2001. Rapid and recent origin of species richness in the Cape flora of South Africa. Nature 412:181–183.

Schluter, D. 2000. The Ecology of Adaptive Radiation: Oxford Series in Ecology and Evolution. Oxford University Press, Oxford.

Seehausen, O. 2002. Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history for its cichlid species flock. Proceedings of the Royal Society B-Biological Sciences 269:491–497.

—. 2006. Conservation: Losing biodiversity by reverse speciation. Current Biology 16:R334–R337.

Stoddard, S. T. 2010. Continuous versus binary representations of landscape heterogeneity in spatially-explicit models of mobile populations. Ecological Modelling 221:2409–2414.

Uchidate, M., T. Shimizu, and A. Iwabuchi. 2010. Data generation techniques of three-dimensional surface roughness for tribology simulations. Journal of Japanese Society of Tribologists 55:375–380.

Wakano, J. Y., and Y. Iwasa. 2013. Evolutionary Branching in a Finite Population: Deterministic Branching vs. Stochastic Branching. Genetics 193:229–241.

Wallace, B. 1975. Hard and soft selection revisited. Evolution 29:465–473.

Wu, J. J. 2000. Simulation of rough surfaces with FFT. Tribology International 33:47–58.