

Reconstructing the Geography of Speciation from Contemporary Biodiversity Data

Alexander Skeels* and Marcel Cardillo

Macroevolution and Macroecology Group, Research School of Biology, Australian National University, Canberra 0200, Australia

Submitted June 4, 2018; Accepted September 18, 2018; Electronically published December 14, 2018

Online enhancements: appendix, R code. Dryad data: <https://dx.doi.org/10.5061/dryad.d9j09bm>.

ABSTRACT: Inferring the geographic mode of speciation could help reveal the evolutionary and ecological mechanisms that underlie the generation of biodiversity. Comparative methods have sought to reconstruct the geographic speciation history of clades, using data on phylogeny and species geographic ranges. However, inference from comparative methods has been limited by uncertainty over whether contemporary biodiversity data retain the historic signal of speciation. We constructed a process-based simulation model to determine the influence of speciation mode and postspeciation range evolution on current biodiversity patterns. The simulations suggest that the signal of speciation history remains detectable in species distributions and phylogeny, even when species ranges have evolved substantially through time. We extracted this signal by using a combination of summary statistics that had good power to distinguish speciation modes and then used these statistics to infer the speciation history of 30 plant and animal clades. The results point to broad taxonomic patterns in the modes of speciation, with strongest support for founder speciation in mammals and birds and strongest support for sympatric speciation in plants. Our model and analyses show that broad-scale comparative methods can be a powerful complementary approach to more focused genomic analyses in the study of the patterns and mechanisms of speciation.

Keywords: age-range correlation, allopatric speciation, approximate Bayesian computation, biogeography, phylogenetic comparative methods, sympatric speciation.

Introduction

Speciation has an explicitly geographic context, as the spatial proximity of diverging populations will influence gene flow and the development of reproductive isolation (Mayr 1963; Coyne and Orr 2004; Wolf and Ellegren 2016). For example, populations bisected by physical barriers (vicariant speciation) or established by long-distance dispersal (founder speciation) experience immediate reproductive isolation, which may drive speciation (Mayr 1942, 1963; Coyne and Orr 2004). In diverging populations that are broadly overlap-

ping (sympatric speciation), partially overlapping, or adjacent (parapatric speciation), it is expected that ecological or phenotypic differentiation can drive divergence with gene flow by reducing contact between co-occurring populations in time or space or by the development of genetic incompatibilities or decreased hybrid fitness (Turelli et al. 2001; Bolnick and Fitzpatrick 2007). Hence, particular geographic modes of speciation are likely to be associated with particular evolutionary processes, although the association may not always be clear-cut (Bolnick and Fitzpatrick 2007; Fitzpatrick et al. 2009).

Increasingly, detailed molecular and ecological data are being used to reconstruct mechanisms of population divergence within species or species groups, with an emphasis on the genomic architecture underlying the inhibition of gene flow (Savolainen et al. 2006; Seehausen et al. 2008; Luebert et al. 2013; Jónsson et al. 2014; Egan et al. 2015; Li et al. 2015). However, there remains a great deal of interest in a more broad-scale, explicitly geographic approach that aims to reconstruct geographic speciation modes within clades, using comparative methods and readily available, contemporary biodiversity data (Barracough and Vogler 2000; Fitzpatrick and Turelli 2006; Phillimore et al. 2008; Anacker and Strauss 2014; Cardillo and Warren 2016). Comparative methods for reconstructing speciation modes are based on the expectation that speciation history should leave a detectable signature in present-day patterns of phylogeny and species geographic distributions (Lynch 1989; Chesser and Zink 1994; Barracough et al. 1998; fig. 1). For example, under both vicariant and founder speciation, recently diverged sister species ought to show little or no overlap in their geographic ranges, whereas sympatric speciation is expected to produce sister species with complete or near-complete spatial overlap (Chesser and Zink 1994; fig. 1). Founder speciation is expected to result in highly asymmetric range sizes among sister species, because the process of speciation by long-distance dispersal is typically thought of as a daughter species with a very small initial population being seeded by a more widely distributed ancestral species (Gavrilets and Hastings 1996), and we would expect this to also lead to a greater geographic dis-

* Corresponding author; email: alexander.skeels@anu.edu.au.

ORCID: Skeels, <http://orcid.org/0000-0002-9973-4703>.

Am. Nat. 2019. Vol. 193, pp. 000–000. © 2018 by The University of Chicago. 0003-0147/2019/19302-58507\$15.00. All rights reserved.
DOI: 10.1086/701125


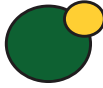
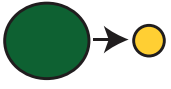

Speciation mode	Range overlap	Range asymmetry	Range isolation
Vicariant 	low	no prediction	high
Parapatric 	moderate	high	low
Founder 	low	high	very high
Sympatric 	high	no prediction	low

Figure 1: Graphical representation of four geographic speciation modes, with a priori predictions of how these might affect key features of sister species range geometries (range overlap, range asymmetry, and range isolation).

tance between sister species ranges at speciation (range isolation). Under vicariant and sympatric modes of speciation, on the other hand, the relative sizes of sister species ranges could range from very similar to highly asymmetric (fig. 1).

Early comparative studies made use of these kinds of expectations to develop summary measures that attempted to provide diagnostic indicators of the prevailing geographic mode of speciation in a clade (Lynch 1989). For example, age-range correlations fit a regression of divergence time against degree of range overlap among all pairs of species in a clade, with the expectation that both the intercept and the slope allow us to distinguish a primarily sympatric from a primarily allopatric speciation mode across the clade (Lynch 1989; Chesser and Zink 1994; Barraclough et al. 1998; Fitzpatrick and Turelli 2006). However, comparative approaches to reconstructing speciation history have often returned results that are equivocal or do not strongly support any particular hypothesis (Fitzpatrick and Turelli 2006). The major criticism leveled at comparative approaches based on contemporary data is that species distributions are dynamic: the spatial configurations of geographic ranges soon after the time of speciation might be erased subsequently by the movement of species range boundaries (Losos and Glor 2003). This may (for example) lead to secondary transitions to sympatry from originally allopatric distributions (Pigot and Tobias 2014) or to more symmetrical sister species range sizes

as the smaller-ranged daughter species expands geographically (Barraclough and Vogler 2000). A critical question, which has remained largely unanswered by previous comparative studies, is, To what extent do present-day geographic ranges of species retain the historic signal of speciation, and to what extent do they reflect postspeciation range movement and evolutionary processes? The answer to this question determines our confidence in the ability of comparative methods to reconstruct speciation history from contemporary biodiversity data.

In large part this will depend on the extent, speed, and predictability of postspeciation range movement. The theoretical expectations are mixed. Closely related species with conserved environmental niches might track changing climates in a similar way, maintaining the spatial relationships of their distributions even as the distributions themselves change. Alternatively, niche evolution might lead species along divergent trajectories in response to climate changes (Ackerly 2003). In most cases, empirical testing of these processes is limited to inferences from indirect evidence, and the evidence to date is also mixed. On the one hand, geographic-range positions tend to show strong phylogenetic signal, suggesting that current species distributions reflect distributions at the time of speciation, modified by gradual drift of range boundaries (Cardillo 2015). On the other hand, it has been shown that range shifts in response to changing climate, such as during Pleistocene glacial cycles, have sometimes been both rapid (Hewitt 2000) and idiosyncratic (Jackson and Overpeck 2000).

The aims of this study are twofold. The first aim is to investigate the degree to which the true speciation history of clades is detectable, and different geographic speciation modes distinguishable, from contemporary data on species distributions and phylogeny. To do this, we construct a process-based simulation model of dynamic range evolution and diversification (the DREaD model), in a way that reflects our expectations of how species ranges evolve and shift following speciation events. Our model incorporates some of the most likely determinants of range movement into the simulation framework: dispersal, niche evolution, and environmental change (Holt 2003; Sexton et al. 2009). By incorporating these processes and exploring the way they interact to determine present-day species ranges, our simulation model extends previous efforts to model the geography of speciation, which have focused primarily on random drift (Phillimore et al. 2008; Cardillo 2015). The second aim of this study is to use simulation-based inference methods to classify the geographic mode of speciation in 30 real clades of plants, vertebrates, and invertebrates to survey patterns in speciation mode across a broad range of taxonomic groups, biogeographic regions, and ecologies. We also ask whether clades show evidence of a predominant geographic mode of speciation (when most divergences in the clades' history can be

classified as a particular speciation mode) or whether clades show evidence of multiple modes in their divergence history. We use simulation-based inference methods that allow us to perform model selection in a statistically rigorous framework (Csilléry et al. 2010) to let the data tell us, in a way that incorporates the complexity and stochasticity of evolutionary and biogeographic processes, whether we can reliably distinguish different geographic modes of speciation in real clades.

Methods

Simulation Model Parameters

The DREaD model simulates the diversification of a clade and the evolution of geographic ranges between speciation events, against the background of a gridded, heterogeneous landscape in which each grid cell contains a value for a single, continuously varying hypothetical environmental variable (for a detailed description of the simulation model, see the appendix, available online). We modeled four distinct speciation modes—vicariant, sympatric, parapatric, and founder speciation (fig. 1)—and one mixed model of speciation including all four modes. The simulation begins by randomly generating the grid values for the environmental variable, with a specified degree of spatial autocorrelation across the whole landscape, using unconditional Gaussian simulation. We then seed the clade with an ancestral species, drawing its attributes (geographic-range boundaries, niche position, and niche breadth) randomly from uniform distributions. At each time step, the species is able to expand its range via dispersal into new grid cells that lie within the range determined by the clade's dispersal capacity (D) and have environmental values that fall within the bounds of the species niche (niche position \pm niche breadth). Dispersal occurs against a background of environmental change through time, modeled in two ways: (1) cyclical, with environmental change modeled as a sine wave with parameters for amplitude, ENV_A , and frequency, ENV_F , and (2) directional, with environmental change modeled as a linear increase with a parameter for the slope, ENV_S . Each environmental-change model was either spatially homogenous, where each grid cell changes by the same amount at each time step, or spatially heterogeneous, where the degree of change in each cell is a linear function of its latitude. Hence, we model four different environmental-change scenarios. Environmental change had the effect of reducing a species geographic range if the environmental values of occupied cells changed so that they no longer fall within the niche of the species.

Concurrently, at each time step, one of six different events can occur: vicariant, sympatric, parapatric, or founder speciation; extinction; or no event. Speciation probability is modeled as a Gaussian function of range size, in which the per-lineage

speciation rate is greatest at intermediate range sizes (Rosenzweig 1978; Gaston 1998). The probability of stochastic extinction decreases logarithmically with range size, so the per-lineage extinction rate is greatest in narrowly distributed species and reaches 0 in species with range sizes equal to or larger than a minimum range size threshold m (Rangel et al. 2007; appendix). The predominance of a particular speciation mode is enforced by setting the values of the speciation rate constant (λ) separately for each speciation mode. A mixed model of speciation is determined by setting equal λ values for all modes.

The basic geographic modes of speciation in the simulation model are sympatric, parapatric, founder, and vicariant. Each of these modes is modeled as follows.

Sympatric. Under the sympatric model, one daughter species maintains the range of the parent and the other occupies a range that lies completely within the boundaries of the parent species range. This is chosen by randomly drawing four coordinates from within the parent species range, which form the boundaries of the daughter species range.

Parapatric. Parapatric speciation occurs via budding at the range periphery. The new species is formed by creating an abutting range that may partially overlap the parent species range. This is done by selecting a cell within dispersal distance, D , from the parent species range boundary and drawing four distances from a uniform distribution (from 1 to D) from this point to be the range boundaries of a new quadrant.

Founder. Founder speciation follows a founder-event model, where dispersal events can found a new species in noncontiguous geographic space. Founder speciation proceeds by selecting a cell within the domain to be colonized with a probability inversely related to the shortest distance from the parent species range. The range boundaries are drawn by selecting four distance values from a uniform distribution (from 1 to D) from the colonized cell.

Vicariant. Vicariant speciation is modeled in two different ways, depending on the geometry of the species range. First, if the species range is a single contiguous area, vicariant speciation will occur via bisection of this range, whereby a line is drawn randomly through the species range, dividing it into two. The bisection is ambiguous with respect to the range size asymmetry of the daughter species ranges. The second method of vicariant division is used if a species range is fragmented. In this case, the parent species range is split so that each daughter species range is composed of a cluster of range fragments that are in closer spatial proximity to each other than to the sister species range. Clustering is performed with a k -means method (Hartigan and Wong 1979) on the X - Y coordinates of the range fragments.

We model phylogenetic signal of the niche at speciation as follows. Under vicariant speciation, immediately follow-

ing speciation both daughter species niche positions are recentered toward the mean environmental value of the new species range. Under founder, parapatric, or sympatric speciation, the niche position of the budded daughter species (the smaller-ranged daughter species) is recentered. The degree to which the species shifts its inherited niche position value toward this new value is modeled with the parameter PS, the proportion of the step between the current niche position and the mean environmental value that the new species will take. A PS value of 1 means that the species will move completely toward the mean, while a value of 0 means that the new species will inherit the same niche position as the parent. In this way, the PS parameter controls the strength of “punctuational” evolution of the niche at speciation events. We also model evolution of the niche along the branches of the phylogeny by allowing a species niche position and niche breadth to drift independently under a modified random-walk model of trait evolution, controlled by the rate parameters NE_p (for niche position) and NE_b (for niche breadth). Species niche positions and breadths drift randomly under the single condition that the environmental value of at least one grid cell within the species range must remain within the species niche (niche position \pm niche breadth).

Running the Simulation Model

We simulated range evolution and speciation under different scenarios of environmental change, niche evolution, dispersal rate, clade size, and geographic speciation modes. Parameters were sampled from uniform prior distributions; Sobol sequences were used to efficiently explore parameter space (Burhenne et al. 2011; appendix), reducing the number of simulation replicates required to effectively explore parameter space. We ran the simulation 36,000 times, until a clade of size n was generated (table A1; tables A1–A3 are available online), discarding 269 simulations that could not be completed because the parameter combination led to total clade extinction more than five times successively. This led to roughly 7,200 replicates for each speciation scenario, which is considered adequate for model selection in an approximate Bayesian computation (ABC) framework (Pudlo et al. 2016). At the completion of each simulation, we retained the resulting phylogeny, each tip species final geographic range, the environmental grid, and a data frame containing information on the final niche position, niche breadth, range size, and mode of speciation at each node in the phylogeny. From this output we generated 30 summary statistics (table A2) that describe aspects of species geographic-range overlap, range size asymmetry, range isolation, range size, and phylogenetic tree shape, within each clade. We reduced these 30 summary statistics to a set of 14, using a variable se-

lection procedure (appendix; table 1). Our simulations were written and implemented in R, version 3.4.2.

Analysis of Simulation Outputs

We compared the distributions of each summary statistic for simulated clades generated under different geographic speciation modes, using pairwise Kolmogorov-Smirnov tests (KSTs). Our simulations generated large sample sizes ($\sim 7,200$ replicates for each speciation scenario), which may return significant P values even when the effect sizes are small. To reduce the chance of misinterpreting significant differences, we ran the analysis across 500 randomly subsampled simulation replicates 100 times and took the mean values of the test statistic and P value. Next, we asked whether the signal of speciation is stronger than the signal of geographic-range evolution in present-day (simulated) data, by exploring which model parameters explained the greatest amount of variation in the summary statistics, independent of all other model parameters. To do this, we used a hierarchical partitioning protocol (Chevan and Sutherland 2017) that assesses all possible combinations of independent variables (model parameters) on the response (summary statistics) in a generalized linear modeling framework and partitions the variance according to a goodness-of-fit statistic (R^2). This allows for the independent assessment of each parameter’s contribution while removing the effects of multicollinearity (MacNally 2000). Hierarchical partitioning was implemented in the R package *hier.part* (Walsh and MacNally 2013).

Empirical Data Collection

We used the outcomes of the simulation model to infer speciation modes in a selection of empirical data sets. We collected spatial and phylogenetic data for 30 clades (six plant, two fish, one invertebrate, four amphibian, four reptile, five mammal, and eight bird clades), selected to cover a range of taxonomic groups and levels, clade sizes, and geographic regions (appendix; table A3). All selected clades were monophyletic and densely sampled, with all clades having $>80\%$ of known species included and most having $>90\%$. Spatial data for species distributions were obtained, where possible, as spatial polygons from the International Union for the Conservation of Nature (IUCN; <http://www.iucnredlist.org>) or BirdLife (2016), which depict species range extents based on both occurrence data and expert assessment of species contemporary distributions. Where range polygons were not available, we used point occurrence records downloaded from the Global Biodiversity Information Facility (GBIF; <http://www.gbif.org>) and cleaned of obvious outliers or supplied from the supplemental materials of an associated article relating to the clade. Spatial polygons were estimated from occurrence points with a fixed- k convex-hull method (Getz

Table 1: The 14 summary metrics used for the study of the geographic mode of speciation with description and supporting references

Summary metric	Abbreviation	Description
Age-range correlation (ARC)	ARC _{slope} , ARC _{intercept}	Slope and intercept of regression between phylogenetic node age and geographic-range overlap (RO) among nodal descendants (Fitzpatrick and Turelli 2006)
Sister species RO × divergence times	RO _{intercept}	Intercept of regression between sister species RO and divergence times; similar to ARC but uses only sister species
Mean sister species RO	RO _{mean}	RO = proportion of range of the smaller-ranged sister species found within the larger-ranged sister species
Sister species sympatry proportions	RO ₉₀ , RO ₁₀₀	Proportion of species that have RO ≥ .9 or RO = 1.0
Difference between sister species RO and sister species–outgroup RO	TO _{mean}	Mean of the difference in the RO between sister species and the RO of each sister species with an outgroup species; scaled between −1, where sister species ranges overlap completely with outgroup ranges and not at all with each other, and 1, where sister species ranges completely overlap with each other and not at all with outgroup ranges (Cardillo 2015)
Range asymmetry × divergence times	Asym _{slope} , Asym _{intercept}	Slope and intercept of regression between range asymmetry and divergence times
Mean and SD of standardized range size	RS _{mean} , RS _{SD}	Standardized range size for each species = (range size)/(largest range size in clade); measured across all tip species in the phylogeny
SD of standardized distance between sister species ranges	RD _{SD}	Standardized range distance = (minimum distance between sister species ranges)/(maximum distance between two species in the clade)
Range isolation × divergence times	RD _{intercept}	Intercept of regression between standardized range distance and divergence times
Sackin’s index	SI	Phylogenetic tree imbalance (Blum and François 2005)

and Wilmers 2004). From the spatial and phylogenetic data we obtained the same summary statistics as from the simulated data.

Model Selection

To infer the predominant speciation mode in the empirical data sets, we used three likelihood-free model selection and model classification techniques: a machine-learning linear discriminant analysis (LDA) and two ABC approaches. We tested the discriminatory ability of our candidate summary statistics to distinguish speciation modes for each method by using a leave-one-out cross-validation procedure, calculating the rate of model misspecification for each geographic speciation mode (the reclassification accuracy). We then inferred the geographic mode of speciation in the 30 empirical data sets, using LDA, implemented with the caret package in R (Kuhn 2016), and two ABC methods (multinomial logistic regression [mnl], and neural net [NN]), implemented with the abc package (Csilléry et al. 2012). An illustrated example of the simulation and model selection pipeline can be seen in figure 2. Derived summary statistics for both empirical and simulated data are deposited in the Dryad Digital Repository, <https://dx.doi.org/10.5061/dryad.d9j09bm> (Skeels and Cardillo 2019; data tables A and B, respectively).

Results

Simulation Results: Speciation Mode, Range Evolution, and Biodiversity Summary Statistics

Across our simulated data set, many of the summary statistics exhibited enough variation between speciation modes to provide discriminating power. Variance partitioning showed that, of the parameters included in the simulations, the geographic mode of speciation had the largest independent effect on 12 of the 14 summary statistics (fig. 3A), while the rate of niche position evolution explained more of the variation in two summary statistics, RD_{intercept} and TO_{mean} (see table 1 for abbreviations). Pairwise KSTs confirmed that distributions of the 14 retained summary statistics were significantly different between speciation modes (fig. 4). The KSTs identified two summary statistics (RO_{intercept} and Asym_{intercept}) that were able to distinguish between the five speciation models in 9 of the 10 pairwise comparisons and a further two (ARC_{slope} and TO_{mean}) that were able to distinguish between models in at least eight comparisons.

There were a number of broad patterns evident in the summary statistics generated from our simulated data set. (1) Founder speciation led to the highest degree of range isolation and lowest range overlap. (2) Vicariant speciation resulted in larger absolute range sizes, low rates of range

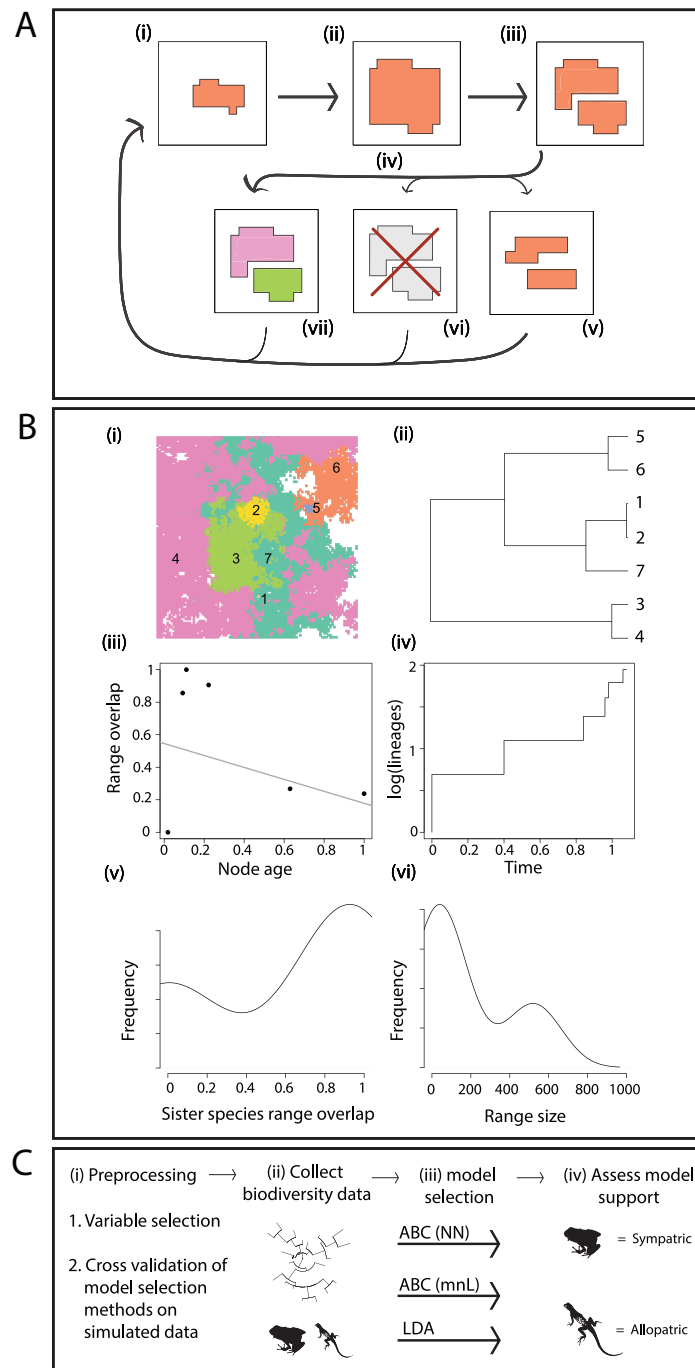


Figure 2: Schematic of the simulation procedure (A), simulation output (B), and model selection procedure (C). A, Simulation workflow: at each time, step each species geographic range (i) can expand via dispersal (ii) or change as a result of changes in the distribution of suitable habitat (iii); the probabilities of each event are assessed, and one event is selected (iv) from niche evolution (v), extinction (vi), or speciation (vii). This algorithm is repeated until a clade of a given number of species is generated. B, Simulation output: the simulation generates a set of species distributions (i) and a phylogenetic tree (ii). Summary metrics are then derived from the final output, including range overlap \times node age (iii), temporal distribution of nodes (γ ; iv), and distribution of sister species range overlap and range size (v). C, Model selection procedure: data preprocessing and model validation (i), generating summary statistics for clades of interest (ii), implementing model selection procedures (ABC = approximate Bayesian computation; NN = neural net; mnL = multinomial logistic regression; LDA = linear discriminant analysis; iii), and identifying support for different geographic speciation modes based on posterior probabilities (iv).

asymmetry, and more balanced phylogenetic branching. (3) Sympatric speciation resulted in the highest degree of range overlap and the lowest range isolation. (4) Parapatric speciation resulted in highly imbalanced phylogenetic trees and asymmetrical range sizes (fig. 3B).

The relative difference between summary statistics for different speciation modes was generally consistent for different values of key model parameters (e.g., the degree of range overlap was higher in sympatric than in vicariant speciation modes at different rates of niche position evolution; fig. 3B). However, variance partitioning showed that both the rate of niche position evolution (NE_p) and the rate of dispersal (D) explained a considerable amount of variance in a number of summary statistics (fig. 3A). Other parameters, including the rate of niche breadth evolution and the environmental-change model, also explained a considerable amount of variance in some parameters, although this was less consistent across summary statistics (fig. 3A). Investigating relationships between estimated summary statistics and parameter values for key range evolution parameters for each speciation mode separately, we found that the same values for several summary statistics could be produced by different speciation modes at different values of niche position evolution (NE_p) and dispersal (D). This shows the interactive effect of speciation mode and range/niche evolution (e.g., range size asymmetry is similar when speciation is sympatric and dispersal rates are low, compared to when speciation is parapatric and dispersal rates are high; fig. 3B). Higher rates of niche evolution led to increased range isolation, decreased range size and range overlap, and greater asymmetry of range sizes. Dispersal rate, on the other hand, had an opposing effect, increasing range overlap and range size while decreasing range isolation and range asymmetry. Furthermore, some summary statistics did not show clear directional trends with model parameters or did not shift in parallel across speciation modes. For example, phylogenetic tree shape (SI) showed an increasing trend with niche evolution rate for the vicariant speciation model but no trend for other speciation models (fig. 3B).

Empirical Results: the Geography of Speciation across Plants and Animals

We asked whether biodiversity summary statistics for 30 different clades, spanning a broad range of plant and animal taxa, are consistent with a single predominant mode of speciation within clades. We used three alternative model selection methods to infer the support for different speciation modes. We found that mammal clades tended to show strongest support for a founder mode of speciation, with two clades, the Madagascar lemurs (Lemuridae) and the Australian diprotodont marsupials (Diprotodontia), also showing strong support for vicariant speciation, especially under the two ABC

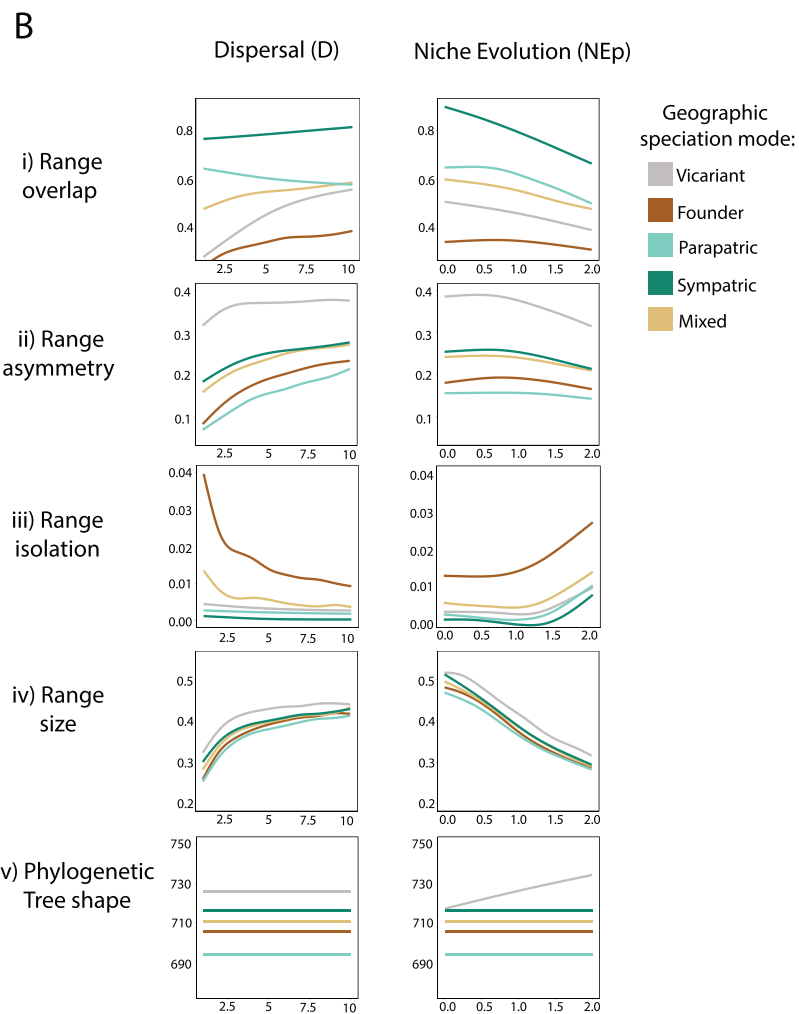
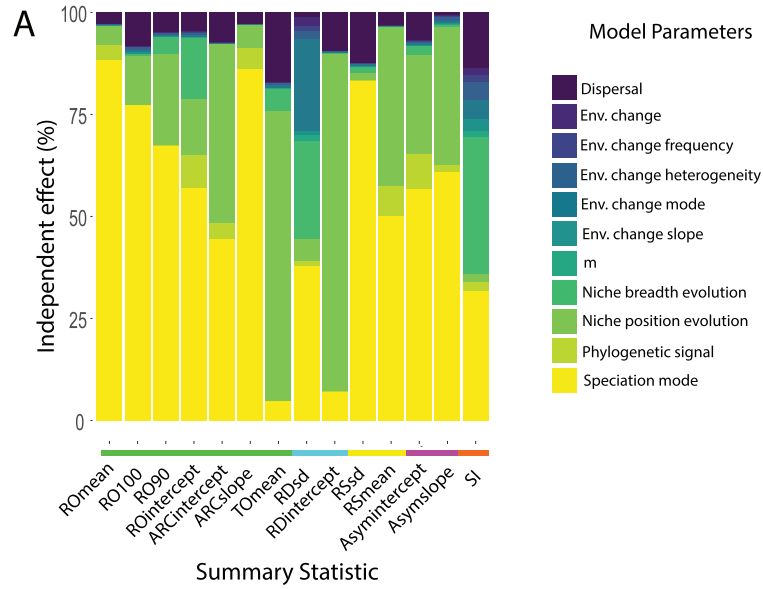
methods, mnL and NN (fig. 5). Similarly, bird clades showed strongest support for founder speciation, except for the Australian robins (Petroicidae), which showed equivalent levels of support for vicariant, mixed, and founder speciation, and the indigobirds/whydahs (*Vidua*), which showed strongest support for vicariant speciation with mnL and for both mixed and vicariant speciation modes with LDA and NN. Founder speciation was also strongly supported in three of the four reptile clades, while support for different speciation modes was more variable among the amphibian, fish, and insect clades. Among plant genera, three (*Protea*, *Hakea*, and *Bursera*) showed support for sympatric speciation under all three methods, one (*Sidalcea*) for parapatric speciation, and two (*Banksia* and *Mimulus*) for a mixed model of speciation, although LDA strongly supported sympatric speciation in these two genera. In only one of the 30 clades we tested (Myobatrachidae) was there strong support for a vicariant speciation mode across all three methods. Relative levels of support for different speciation modes tended to be consistent among the three methods (LDA, mnL, and NN), although in general, the LDA method more frequently attributed strong support to a single speciation mode, while NN was more likely to apportion support more evenly among several modes.

Model classification accuracy was assessed with a cross-validation procedure for each method (LDA, mnL, and NN). We found that reclassification accuracy was significantly better than random ($\kappa > 0.6$, $P < .001$ across three methods), and we were able to predict the correct model of speciation geography in 69%–71% of simulations. Vicariant speciation had the highest reclassification accuracy of all speciation models (>84%), followed by sympatric (>78%), founder (>73%), parapatric (>61%), and finally the mixed model (<46%). Model adequacy was assessed graphically by plotting the predicted values for the first two axes of an LDA for the empirical data sets onto the simulated data sets (fig. 6). The empirical data sets fall within the range of our simulated data in all cases, so we can conclude that our simulations generated realistic biogeographic patterns (Pudlo et al. 2016).

Discussion

Detecting the Geographic Mode of Speciation

Comparative analyses to recover the prevailing geographic mode of speciation in clades have been limited by uncertainty over how much of the historic signal of speciation is retained in contemporary biodiversity data and how much is eroded by postspeciation geographic-range movement (Losos and Glor 2003). Previous attempts to model these processes have gone some way to reducing the uncertainty, but their interpretation has remained rather ambiguous because models have been based on a limited set of processes, have typically applied only one or two summary statistics to



recover signal from contemporary data, and have focused on hypothesis testing rather than model comparison and selection (Barracough and Vogler 2000; Fitzpatrick and Turelli 2006; Phillimore et al. 2008; Cardillo and Warren 2016). Figure 3B illustrates why previous studies based on one or two biodiversity summary statistics have often lacked power to discriminate geographic speciation modes. Interactive effects between speciation mode and range evolution mean that models with different speciation modes can return similar values for summary statistics, depending on the range of dispersal and niche evolution parameters.

Our simulation model (DREaD), together with model selection methods based on 14 summary statistics that capture aspects of species range sizes, proximity, and overlap and their phylogeny, show that the signal of speciation mode can indeed be detected in contemporary biodiversity data. Several of our analysis results point to this conclusion. First, the geographic mode of speciation had the strongest independent effect on the majority of summary statistics, compared to other simulation model parameters (fig. 3A). Second, simulations under different speciation modes produced summary statistics with consistently different distributions (e.g., figs. 3B, 4). Third, the cross-validation procedures under both LDA and ABC methods had a considerably higher-than-random reclassification accuracy. This allowed us to infer the geographic modes of speciation in 30 empirical clades across a broad range of plant and animal taxa. The results point to a predominance of founder speciation in animals and of sympatric speciation in plants.

Broad-Scale Patterns in Speciation Mode

Our reconstructions of geographic speciation modes for plant and animal clades reveal broad-scale taxonomic patterns in speciation mode across a wide range of organisms. For much of the twentieth century, allopatric speciation was widely considered the most likely mode of speciation, at least in animals, because of the theoretical requirement for disruption to gene flow in order to generate reproductive isolation (Mayr 1963). This idea has been supported by several comparative studies of speciation mode in mammals and birds (Fitzpatrick and Turelli 2006; Phillimore et al. 2008) and is supported for many vertebrate clades in our study. Despite the importance of allopatry in the formation of incipient

species, debate over the relative importance of vicariance versus dispersal has been of interest in biogeography, as these different mechanisms relate to fundamentally different processes—geographic isolation is environmental in the case of vicariant speciation (e.g., mountains, rivers, coastlines) and essentially biological in the case of founder speciation, since dispersal is associated with the mobility of individual organisms. Our result is consistent with recent evidence for the frequency of founder-event speciation in model-based biogeographic analyses (Cowie and Holland 2006; Matzke 2014), suggesting that the preponderance of allopatric speciation might be explained by the continual dispersal “attempts” being made by individual organisms driven beyond their range limits (propagule pressure; Levin 2006), compared to the relative infrequency of the geological events that cause a population to be subdivided (Gaston 1998).

More recent theoretical developments have shown that speciation can occur with geographic-range overlap and the opportunity for gene flow (Gavrilets et al. 2000; Baack et al. 2015). A number of studies have demonstrated this empirically (e.g., Barluenga et al. 2006; Savolainen et al. 2006; Seehausen et al. 2008; Peakall et al. 2010) and suggest that it may be common, especially in higher taxa such as plants (Anacker and Strauss 2014; Grossenbacher et al. 2014) and, under certain conditions, in animals (Bush 1994; Via 2001; Nosil 2008; Rosser et al. 2015). Our results support this suggestion by inferring a prevailing sympatric mode of speciation for many of the plant clades we examined as well as for several herpetile clades (*Anolis*, *Litoria*, and *Plethodon*) and the rockfish genus *Sebastes*. Many of these groups are hypothesized to have diversified by mechanisms of ecological divergence, a key step in establishing a unique species identity in the face of gene flow. The Caribbean lizard genus *Anolis* is, in fact, a textbook case study for repeated speciation by the divergence of co-occurring island populations into habitat-specific ecomorphs (Mahler et al. 2010). Sympatric speciation has always been considered to be more prevalent in plants than in animals, because of strong divergent or disruptive selection that can be exerted by ecological mechanisms such as differences in flowering times or soil type specialization (Savolainen et al. 2006) and because plants are considered prone to speciation by genome duplication leading to polyploidy (Rieseberg and Willis 2007).

Figure 3: A, Independent effect (% of variance explained) of each simulation model parameter on 14 summary statistics (table 1) describing spatial and phylogenetic biodiversity patterns; B, relationship between five summary statistics and two key model parameters, niche position evolution (NE_p) and dispersal rate (D), for each of the five speciation modes. Summary statistics can be classified as describing either measures of range overlap (horizontal green line in A), range isolation (blue line), range size (yellow line), range asymmetry (purple line), or phylogenetic tree shape (orange line). B depicts the loess curves for one summary statistic from each of these categories to give a graphical impression of the general trends: $RO_{intercept}$, a measure of range overlap (i), $Asym_{intercept}$, a measure of range size asymmetry (ii), $RD_{intercept}$, a measure of the distance between sister species ranges (iii), RS_{means} , a measure of the average range size (iv), and Sackin's index (SI in A), a measure of phylogenetic tree imbalance (v). Env. change = environmental change.

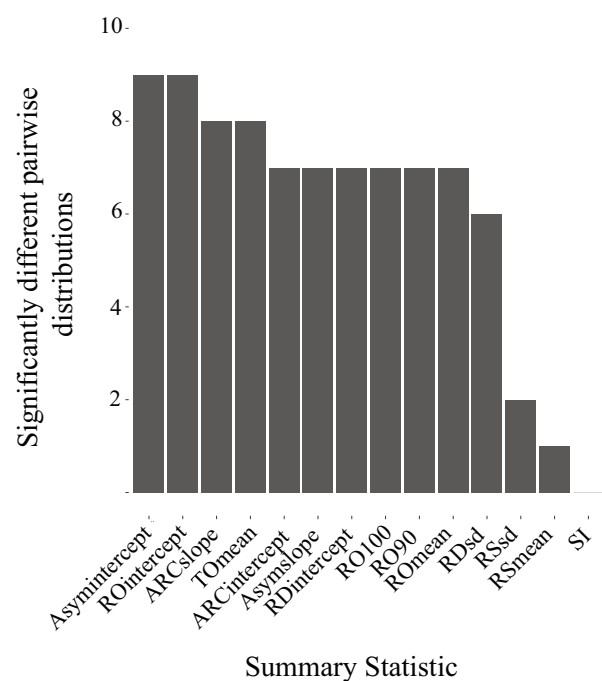


Figure 4: Number of significantly different comparisons of the distribution of 14 summary statistics (table 1) between each pair of speciation modes in the simulated data set; 10 pairs in total. Significance was tested with Kolmogorov-Smirnov tests.

Geographic-Range Evolution

Although the mode of speciation was the major driver of spatial and phylogenetic patterns in our simulated data, there were strong interactive effects of niche evolution and dispersal with the mode of speciation. Niche position evolution was negatively correlated, and dispersal positively correlated, with range overlap, such that low rates of niche evolution (i.e., niche conservatism: Wiens 2004) and a high rate of dispersal were both associated with a higher degree of range overlap between closely related species. The explanation of this effect in our simulation model also suggests a plausible biological scenario in the real world, as follows: if allopatric sister species inherit a similar environmental niche from their common ancestor, then in a spatially autocorrelated landscape the most suitable habitat for a species is likely to be found within the range of its sister species. A high dispersal capacity will offer many opportunities for sister species ranges to move back into sympatry (Pigot and Tobias 2014). This suggests that the degree of spatial autocorrelation in key environmental-niche variables for a given clade may be important: if niches are conserved, stronger spatial autocorrelation should exert a pressure for distributions of sister species to overlap. A further process that could counteract this pressure for sister species to overlap is biotic interaction. We did not incorporate biotic interactions (such as interspecific competition) into

our model because they add a substantial layer of complexity and are probably best tackled as a separate question, but it is conceivable that in some cases, competition serves to minimize range overlap and maintain allopatric distributions (Pigot and Tobias 2013; Wisz et al. 2013; Pigot et al. 2018). We suggest that this should be an important avenue for further development of our model.

Environmental-change parameters did not exert a strong influence on the distribution of the summary statistics, possibly because of a filtering effect. If a species cannot disperse rapidly enough to track environmental change, there will be a high probability of its extinction, leaving no record in contemporary phylogenetic and geographic data. Furthermore, species that are severely affected by climatic change may be a phylogenetically biased sample, if closely related species inherit from their common ancestors (1) a similar capacity to disperse, (2) similar ecological traits that put them at risk of extinction, or (3) the occupation of similar habitats that are differentially threatened (Parmesan 2006). Depending on the phylogenetic conservatism of key environmental-niche traits, we might therefore expect closely related species to show similar responses to environmental fluctuations. In some cases, this could elevate the probability of extinction of the entire clade, while in others it may serve to maintain the relative spatial relationships of species ranges.

Model Assumptions and Caveats

Like any model of complex systems, the DREAd model necessarily makes a number of simplifying assumptions. As mentioned above, we omit the possible role of biotic interactions in limiting species range boundaries. Additionally, our modeling approach follows, a priori, the simplifying assumption that speciation can be categorized into a set of discrete geographic modes. The geography of speciation, however, could instead be described by a continuum from complete spatial separation to complete overlap (Fitzpatrick et al. 2009). There is empirical evidence to suggest that many speciation events may be best explained with varying degrees of range overlap along this continuum (Pinho and Hey 2010). Furthermore, each individual speciation event may involve a protracted process in which populations pass through various stages of allopatry and sympatry. For example, gene flow between populations may recur at different stages of the speciation process (Rundle and Nosil 2005; Nosil 2008), with genetic variation accumulating in allopatry, then becoming sorted into sympatric populations through introgression (Feder et al. 2005; Poelstra et al. 2018). In such cases, whether we can identify the one predominant geographic mode of speciation is unclear. Our model is phenomenological and simulates divergence at the lineage level rather than the population level, so we cannot rule out more complex speciation histories of the clades we have investigated. We believe that there is some

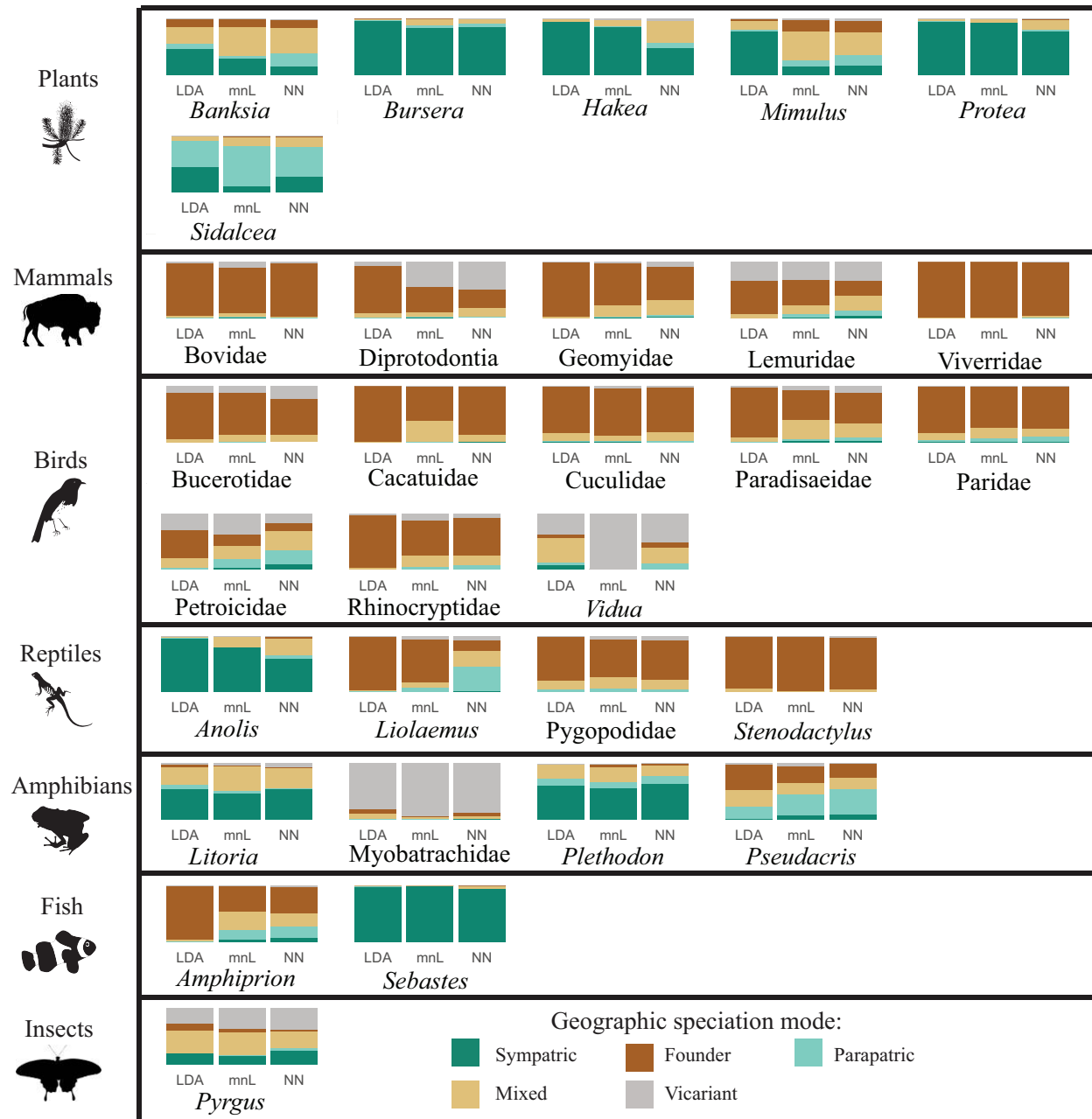


Figure 5: Posterior probabilities for inferred geographic speciation mode for 30 plant and animal clades. The height of the colored bars indicate the relative support for each speciation mode under each of three model selection methods: linear discriminant analysis (LDA), multinomial logistic regression (mnL), and neural net (NN). Clades are grouped by higher taxa: plants, mammals, birds, reptiles, amphibians, fish, and invertebrates. Founder speciation shows the strongest support in birds, mammals, and reptiles, whereas sympatric, parapatric, and mixed speciation modes tend to prevail in amphibians and plants.

value in classifying speciation modes into discrete categories at this level of investigation. For example, if speciation is completed in sympatry, then a spatial model of sympatric speciation may adequately capture the spatial patterns of divergence, even if the processes by which those patterns emerge are not explicitly considered. Nevertheless, we acknowledge

that definitions of allopatry and sympatry often fall in a gray area that defies simple categorization.

We treat speciation mode as a categorical parameter in our model to compare whether one particular speciation mode predominates in the evolutionary history of each clade or a mixed model of speciation better explains the

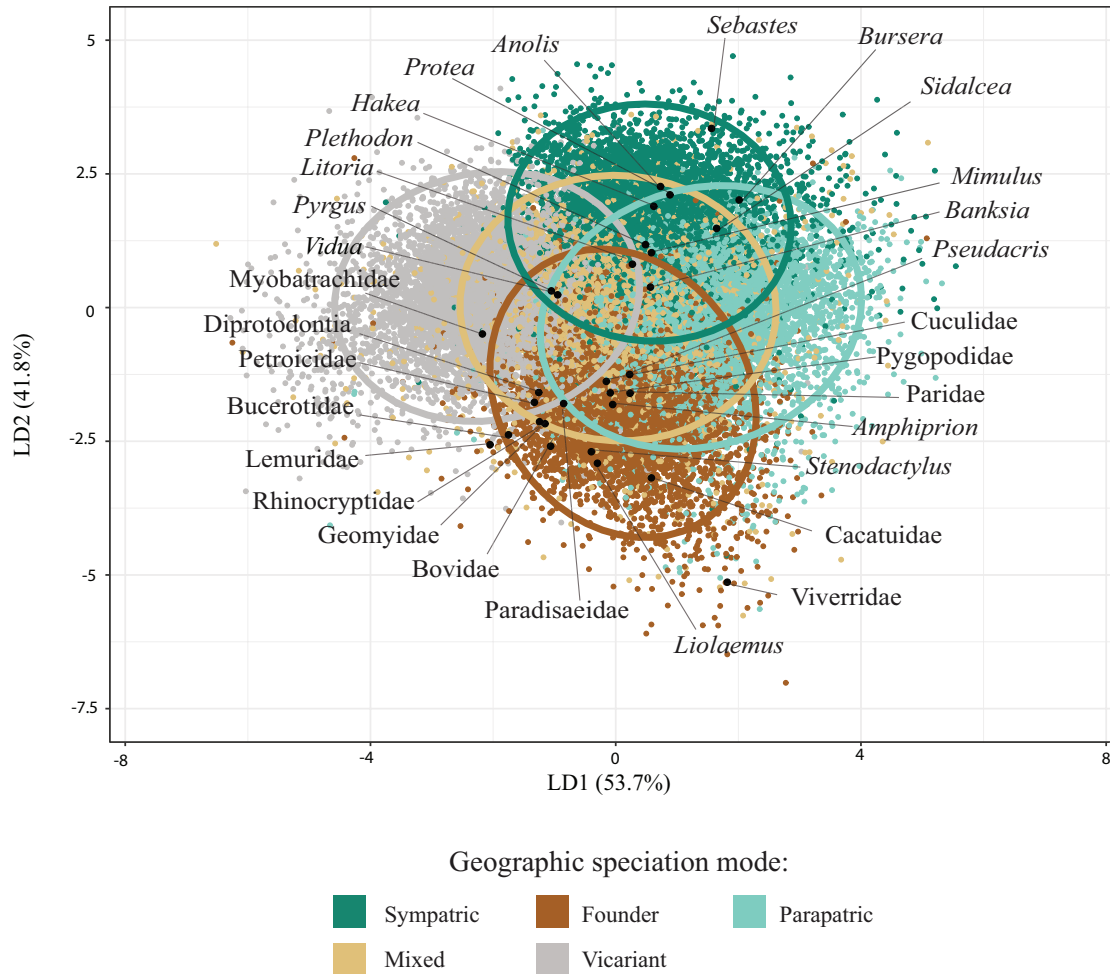


Figure 6: Projection of the first two axes of a linear discriminant analysis (LDA) performed on simulated and empirical data sets. Colored points show bivariate distributions of the first two axes of an LDA constructed from 14 biodiversity summary statistics (table 1). The summary statistics are derived from clades simulated under five geographic modes of speciation. Colored ellipses contain 90% of all simulated data sets for a given speciation mode. Black points represent predicted values of 30 empirical plant and animal clades.

data in a model selection framework. However, as the mixed model had poor reclassification accuracy in the simulation analysis, there may be a bias toward underrepresentation in the posterior probabilities of our model selection procedure. This may partially explain the tendency for greater support for single speciation modes in the clades we investigated. We suggest that a future avenue of research should be to allow the frequencies of different speciation modes (λ parameter) within a clade to be estimated from the data. In fact, several model-based methods for reconstructing ancestral ranges already do this (e.g., DIVA: Ronquist 1997; DEC: Ree and Smith 2008). However, these methods differ from DREaD in that they treat geographic ranges as the occupancy of discrete biogeographical regions; they do not explicitly model speciation as a function of the configuration of species ranges in continuous space and are aimed at

reconstructing large-scale biogeographic shifts rather than the geographic mode of speciation.

As well as the relative values of λ between modes, the absolute rate of speciation and the function linking range size to the probability of speciation require further exploration. The justification of a peaked relationship between range size and speciation is largely based on theory regarding the placement of barriers and the likelihood of vicariant speciation (Rosenzweig 1978; Gaston 2003); however, whether we expect the same relationship to hold with other speciation modes is unclear, as there is little empirical evidence on which to base the parameterization of this model component. It is possible that different relationships between speciation mode, range size (and other features of species biogeography), and speciation rate may lead to patterns in summary statistics different from those we observed. The rate of speciation relative to other

rates in DREaD (e.g., niche evolution, dispersal, or environmental change) may also affect the relative impact of these parameters on biogeographic patterns, although we did not explore this in depth. There are still relatively few spatially explicit studies that have integrated multiple macroevolutionary and biogeographic processes (but see, e.g., Qiao et al. 2016; Rangel et al. 2018), so there is much work to be done to understand how these dynamics interact to influence present-day patterns in species distributions (Weber et al. 2017).

Another area of uncertainty is the sensitivity of our results to the way species geographic ranges are defined. In particular, the use of polygons that define a species extent of occurrence may overestimate the degree of range overlap between two species and hide the microallopatry that may occur at much finer spatial scales (Cardillo and Warren 2016). For example, although we found strong support for a sympatric model of speciation in *Sebastes* rockfish, previous studies have suggested that species in this group have partitioned geographic space in three dimensions along a depth gradient and that their distributions may be better explained by a parapatric model of speciation (Ingram 2011). Furthermore, there may be a difference between using polygons that have been determined by expert assessment (e.g., IUCN) and using those that are modeled from species occurrence records (e.g., convex hulls), as these may infer range boundaries at different resolutions, potentially biasing hull methods toward nonallopatric speciation modes. However, it is far from clear what the appropriate spatial resolution is for the measurement of geographic-range overlap to infer speciation mode, and we suspect that the answer is context dependent, depending on the patterns of habitat use and dispersal capabilities of the species involved. The application of simulation models such as DREaD may allow this to be tested *in silico* to inform future studies.

Conclusions

Losos and Glor (2003, p. 221) asked the question, “Can a null hypothesis that speciation was not sympatric be rejected if sympatric species are found to be sister taxa?” As allopatric speciation is typically considered the “null model” of speciation (Coyne and Orr 2004), the burden of evidence has been on finding strong support for nonallopatric speciation (Bolnick and Fitzpatrick 2007). Comparative approaches to understanding speciation geography have stalled in the past decade, partly because existing methods have had relatively little discriminatory power. We have shown that, by adopting an approach based on simulation and model selection increasingly advocated in biogeography (Goldberg et al. 2011; Matzke 2014; Qiao et al. 2016; Sukumaran et al. 2016; Cabral et al. 2017), Losos and Glor’s question can be rephrased as “Given the observed data and a model of geographic-range

evolution, what is the support for allopatric speciation relative to other models?” Our results suggest that a broad-scale comparative approach to understanding speciation processes can indeed be powerful and informative.

In recent years, much of the research on speciation has shifted away from tests with an explicitly geographic focus toward questions about regions of the genome that show evidence of divergence among populations in the context of gene flow and selection (Feder et al. 2012, 2013; Seehausen et al. 2014; Wolf and Ellegren 2016; Foote 2018). We see broad-scale comparative methods and taxonomically focused genomic and population genetic studies as complementary. At present, a broad taxonomic overview of inferred speciation modes can be achieved only by using geographic and phylogenetic data, and the outcomes of such analyses can serve to generate hypotheses about speciation mechanisms that might then be testable in particular species groups through the use of genomic data.

Acknowledgments

A.S. is supported by the Australian Government Research Training Program. M.C. acknowledges support from Australian Research Council grant DP160103942 and the Research School of Biology, Australian National University. We thank C. Moritz, R. Dinnage, X. Hua, N. Matzke, and D. Rosauer for feedback that was useful in developing this study.

Literature Cited

- Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences* 164:S165–S184.
- 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 B* 281:20132980. doi:10.1098/rspb.2013.2980.
- Baack, E., M. C. Melo, L. H. Rieseberg, and D. Ortiz-Barrientos. 2015. The origins of reproductive isolation in plants. *New Phytologist* 207:968–984.
- Barluenga, M., K. N. Stölting, W. Salzburger, M. Muschick, and A. Meyer. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439:719–723.
- Barraclough, T. G., and A. P. Vogler. 2000. Detecting the geographic pattern of speciation from species level phylogenies. *American Naturalist* 155:419–434.
- Barraclough, T. G., A. P. Vogler, and P. H. Harvey. 1998. Revealing the factors that promote speciation. *Philosophical Transactions of the Royal Society B* 353:241–249.
- BirdLife. 2016. Bird species distribution maps of the world. Version 6. BirdLife, Cambridge.
- Blum, M. G. B., and O. François. 2005. On statistical tests of phylogenetic tree imbalance: the Sackin and other indices revisited. *Mathematical Biosciences* 195:141–153.
- Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics* 38:459–487.

- Burhenne, S., D. Jacob, and G. P. Henze. 2011. Sampling based on Sobol' sequences for Monte Carlo techniques applied to building simulations. Pages 1816–1823 in *Proceedings of Building Simulation 2011: 12th IBPSA Building Simulation Conference*. International Building Performance Simulation Association, Sydney.
- Bush, G. L. 1994. Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology and Evolution* 9:285–288.
- Cabral, J. S., L. Valente, and F. Hartig. 2017. Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects. *Ecography* 40:267–280.
- Cardillo, M. 2015. Geographic range shifts do not erase the historic signal of speciation in mammals. *American Naturalist* 185:343–353.
- Cardillo, M., and D. Warren. 2016. Analysing patterns of spatial and niche overlap among species at multiple resolutions. *Global Ecology and Biogeography* 25:951–963.
- Chesser, R. T., and R. M. Zink. 1994. Modes of speciation in birds: a test of Lynch's method. *Evolution* 48:490–497.
- Chevan, A., and M. Sutherland. 2017. Hierarchical partitioning. *American Statistician* 45:90–96.
- Cowie, R. H., and B. S. Holland. 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *Journal of Biogeography* 33:193–198.
- Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sinauer, Sunderland, MA.
- Csilléry, K., M. G. B. Blum, O. E. Gaggiotti, and O. François. 2010. Approximate Bayesian computation (ABC) in practice. *Trends in Ecology and Evolution* 25:410–418.
- Csilléry, K., O. François, and M. G. B. Blum. 2012. ABC: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* 3:475–479.
- Egan, S. P., G. J. Ragland, L. Assour, T. H. Q. Powell, G. R. Hood, S. Emrich, P. Nosil, and J. L. Feder. 2015. Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecology Letters* 18:817–825.
- Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* 28:342–350.
- Feder, J. L., S. M. Flaxman, S. P. Egan, A. A. Comeault, and P. Nosil. 2013. Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics* 44:73–97.
- Feder, J. L., X. Xie, J. Rull, S. Velez, A. Forbes, B. Leung, H. Dambroski, K. E. Filchak, and M. Aluja. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the USA* 102(suppl. 1):6573–6580.
- Fitzpatrick, B. M., J. A. Fordyce, and S. Gavrillets. 2009. Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology* 22:2342–2347.
- Fitzpatrick, B. M., and M. Turelli. 2006. The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution* 60:601–615.
- Footo, A. D. 2018. Sympatric speciation in the genomic era. *Trends in Ecology and Evolution* 33:85–95.
- Gaston, K. J. 1998. Species-range size distributions: products of speciation, extinction and transformation. *Philosophical Transactions of the Royal Society B* 353:219–230.
- . 2003. *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.
- Gavrillets, S., and A. Hastings. 1996. Founder effect speciation: a theoretical reassessment. *American Naturalist* 147:466–491.
- Gavrillets, S., H. Li, and M. D. Vose. 2000. Patterns of parapatric speciation. *Evolution* 54:1126–1134.
- Getz, W. M., and C. C. Wilmers. 2004. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography* 27:489–505.
- 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.
- Grossenbacher, D. L., S. D. Veloz, and J. P. Sexton. 2014. Niche and range size patterns suggest that speciation begins in small, ecologically diverged populations in North American monkeyflowers (*Mimulus* spp.). *Evolution* 68:1270–1280.
- Hartigan, J. A., and M. A. Wong. 1979. A K-means clustering algorithm. *Applied Statistics* 28:100–108.
- Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913.
- Holt, R. D. 2003. On the evolutionary ecology of species ranges. *Evolutionary Ecology Research* 5:159–178.
- Ingram, T. 2011. Speciation along a depth gradient in a marine adaptive radiation. *Proceedings of the Royal Society B* 278:613–618.
- Jackson, S. T., and J. T. Overpeck. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology* 26:194–220.
- Jónsson, H., M. Schubert, A. Seguin-Orlando, A. Ginolhac, L. Petersen, M. Fumagalli, A. Albrechtsen, et al. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the USA* 111:18655–18660.
- Kuhn, M. 2016. *Caret: classification and regression training*. R package version 6.0-73. <https://cran.r-project.org/web/packages/caret/>.
- Levin, D. A. 2006. Ancient dispersals, propagule pressure, and species selection in flowering plants. *Systematic Botany* 31:443–448.
- Li, K., W. Hong, H. Jiao, G. Wang, K. A. Rodriguez, R. Buffenstein, Y. Zhao, E. Nevo, and H. Zhao. 2015. Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. *Proceedings of the National Academy of Sciences of the USA* 112:11905–11910.
- Losos, J. B., and R. E. Glor. 2003. Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology and Evolution* 18:220–227.
- Luebert, F., P. Jacobs, H. H. Hilger, and L. A. H. Muller. 2013. Evidence for nonallopatric speciation among closely related sympatric *Heliotropium* species in the Atacama Desert. *Ecology and Evolution* 4:266–275.
- Lynch, J. D. 1989. The gauge of speciation: on the frequencies of modes of speciation. Pages 527–553 in D. Otte and J. Endler, eds. *Speciation and its consequences*. Sinauer, Sunderland, MA.
- Mahler, D. L., L. J. Revell, R. E. Glor, and J. B. Losos. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of the Greater Antillean anoles. *Evolution* 64:2731–2745.
- Matzke, N. J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* 63:951–970.
- Mayr, E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press, Cambridge, MA.
- . 1963. *Animal species and evolution*. Harvard University Press, Cambridge, MA.
- Nosil, P. 2008. Speciation with gene flow could be common. *Molecular Ecology* 17:2006–2008.
- Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* 37:637–671.

- Peakall, R., D. Ebert, J. Poldy, R. A. Barrow, W. Francke, C. C. Bower, and F. P. Schiestl. 2010. Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually deceptive *Chiloglottis* orchids: implications for pollinator-driven speciation. *New Phytologist* 188:437–450.
- Phillimore, A. B., C. D. L. Orme, G. H. Thomas, T. M. Blackburn, P. M. Bennett, K. J. Gaston, and I. P. F. Owens. 2008. Sympatric speciation in birds is rare: insights from range data and simulations. *American Naturalist* 171:646–657.
- Pigot, A. L., W. Jetz, C. Sheard, and J. A. Tobias. 2018. The macroecological dynamics of species coexistence in birds. *Nature Ecology and Evolution* 2:1112–1119.
- Pigot, A. L., and J. A. Tobias. 2013. Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters* 16:330–338.
- . 2014. Dispersal and the transition to sympatry in vertebrates. *Proceedings of the Royal Society B* 282:20141929. doi:10.1098/rspb.2014.1929.
- Pinho, C., and J. Hey. 2010. Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics* 41:215–230.
- Poelstra, J. W., E. J. Richards, and C. H. Martin. 2018. Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Molecular Ecology* 27:4270–4288. doi:10.1111/mec.14784.
- Pudlo, P., J.-M. Marin, A. Estoup, J.-M. Cornuet, M. Gauthier, and C. P. Robert. 2016. Reliable ABC model choice via random forests. *Bioinformatics* 32:859–866.
- Qiao, H., E. E. Saupe, J. Soberón, A. T. Peterson, and C. E. Myers. 2016. Impacts of niche breadth and dispersal ability on macroevolutionary patterns. *American Naturalist* 188:149–162.
- Rangel, T. F., J. A. F. Diniz-Filho, and R. K. Colwell. 2007. Species richness and evolutionary niche dynamics: a spatial pattern-oriented simulation experiment. *American Naturalist* 170:602–616.
- Rangel, T. F., N. R. Edwards, P. B. Holden, J. A. F. Diniz-Filho, W. D. Gosling, M. T. P. Coelho, F. A. S. Cassemiro, C. Rahbek, and R. K. Colwell. 2018. Modeling the ecology and evolution of biodiversity: biogeographical cradles, museums, and graves. *Science* 361:eaar5452. doi:10.1126/science.aar5452.
- Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57:4–14.
- Rieseberg, L. H., and J. H. Willis. 2007. Plant speciation. *Science* 317:910–915.
- Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* 46:195–203.
- Rosenzweig, M. L. 1978. Geographical speciation: on range size and the probability of isolate formation. Pages 172–194 in D. Wollkind, ed. *Proceedings of the Washington State University Conference on Biomathematics and Biostatistics*. Washington State University, Pullman.
- Rosser, N., K. M. Kozak, A. B. Phillimore, and J. Mallet. 2015. Extensive range overlap between heliconiine sister species: evidence for sympatric speciation in butterflies? *BMC Evolutionary Biology* 15:125. doi:10.1186/s12862-015-0420-3.
- Rundle, H., and P. Nosil. 2005. Ecological speciation. *Ecology Letters* 8:336–352.
- Savolainen, V., M. Anstett, C. Lexer, I. Hutton, J. J. Clarkson, M. V. Norup, M. P. Powell, D. Springate, N. Salamin, and W. J. Baker. 2006. Sympatric speciation in palms on an oceanic island. *Nature* 441:9–12.
- Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, et al. 2014. Genomics and the origin of species. *Nature Reviews Genetics* 15:176–192.
- Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. J. Mrosso, R. Miyagi, I. van der Sluijs, et al. 2008. Speciation through sensory drive in cichlid fish. *Nature* 455:620–626.
- Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics* 40:415–436.
- Skeels, A., and M. Cardillo. 2019. Data from: Reconstructing the geography of speciation from contemporary biodiversity data. *American Naturalist*, Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.d9j09bm>.
- Sukumaran, J., E. P. Economo, and L. Lacey Knowles. 2016. Machine learning biogeographic processes from biotic patterns: a new trait-dependent dispersal and diversification model with model choice by simulation-trained discriminant analysis. *Systematic Biology* 65:525–545.
- Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. *Trends in Ecology and Evolution* 16:330–343.
- Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution* 16:381–390.
- Walsh, C., and R. MacNally. 2013. hier.part: hierarchical partitioning. R package version 1.0-4. <https://CRAN.R-project.org/package=hier.part>.
- Weber, M. G., C. E. Wagner, R. J. Best, L. J. Harmon, and B. Matthews. 2017. Evolution in a community context: on integrating ecological interactions and macroevolution. *Trends in Ecology and Evolution* 32:291–304.
- Wiens, J. J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* 58:193–197.
- Wisz, M. S., J. Pottier, W. D. Kissling, L. Pellissier, J. Lenoir, C. F. Damgaard, C. F. Dormann, et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* 88:15–30.
- Wolf, J. B. W., and H. Ellegren. 2016. Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics* 18:87–100.

References Cited Only in the Online Appendixes

- Aldous, D. 1996. Probability distributions on cladograms. Pages 1–18 in D. Aldous and R. Pemantle, eds. *Random discrete structures*. Springer, New York.
- Bivand, R., E. J. Pebesma, and V. Gómez-Rubio. 2013. *Applied spatial data analysis with R*. 2nd ed. Springer, New York.
- Bivand, R., and C. Rundel. 2016. rgeos: interface to geometry engine—open source (GEOS). R package version 0.3-20. <https://CRAN.R-project.org/package=rgeos>.
- Blum, M. G. B., and O. François. 2006. Which random processes describe the tree of life? a large-scale study of phylogenetic tree imbalance. *Systematic Biology* 55:685–691.
- . 2010. Non-linear regression models for approximate Bayesian computation. *Statistics and Computing* 20:63–73.
- Blum, M. G. B., M. A. Nunes, D. Prangle, and S. A. Sisson. 2017. A comparative review of dimension reduction methods in approximate Bayesian computation. *Statistical Science* 28:189–208.
- Brennan, I. G., and P. M. Oliver. 2017. Mass turnover and recovery dynamics of a diverse Australian continental radiation. *Evolution* 71:1352–1356.

- Cardillo, M., and R. Pratt. 2013. Evolution of a hotspot genus: geographic variation in speciation and extinction rates in *Banksia* (Proteaceae). *BMC Evolutionary Biology* 13:155. doi:10.1186/1471-2148-13-155.
- Cardillo, M., P. H. Weston, Z. K. M. Reynolds, P. M. Olde, A. R. Mast, E. M. Lemmon, A. R. Lemmon, and L. Bromham. 2017. The phylogeny and biogeography of *Hakea* (Proteaceae) reveals the role of biome shifts in a continental plant radiation. *Evolution* 71:1928–1943.
- Colwell, R. K., and T. F. Rangel. 2010. A stochastic, evolutionary model for range shifts and richness on tropical elevational gradients under Quaternary glacial cycles. *Philosophical Transactions of the Royal Society B* 365:3695–3707.
- Esquerré, D., I. G. Brennan, R. A. Catullo, F. Torres-Pérez, and J. S. Keogh. Forthcoming. How mountains shape biodiversity: the role of the Andes in biogeography, diversification and reproductive biology in South America's most species rich lizard radiation (Squamata: Liolaemidae). *Evolution*.
- Fritz, S. A. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters* 12:538–549.
- Hennig, C. 2015. fpc: flexible procedures for clustering. R package version 2.1-10. <https://CRAN.R-project.org/package=fpc>.
- Hijmans, R. J. 2016. Raster: geographic data analysis and modeling. R package version 2.5-8. <https://CRAN.R-project.org/package=raster>.
- Ingram, T., and Y. Kai. 2014. The geography of morphological convergence in the radiations of Pacific *Sebastes* rockfishes. *American Naturalist* 184:E115–E131.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. *Nature* 491:444–448.
- Litsios, G., C. A. Sims, R. O. Wüest, P. B. Pearman, N. E. Zimmermann, and N. Salamin. 2012. Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology* 12:212. doi:10.1186/1471-2148-12-212.
- MacNally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between—and reconciliation of—“predictive” and “explanatory” models. *Biodiversity and Conservation* 9:655–671.
- Metallinou, M., E. N. Arnold, P.-A. Crochet, P. Geniez, J. C. Brito, P. Lymberakis, S. Baha El Din, R. Sindaco, M. Robinson, and S. Carranza. 2012. Conquering the Sahara and Arabian deserts: systematics and biogeography of *Stenodactylus* geckos (Reptilia: Gekkonidae). *BMC Evolutionary Biology* 12:258. doi:10.1186/1471-2148-12-258.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pebesma, E. J. 2004. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences* 30:683–691.
- Pigot, A. L., A. B. Phillimore, I. P. F. Owens, and C. D. L. Orme. 2010. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. *Systematic Biology* 59:660–673.
- Pitteloud, C., N. Arrigo, T. Suchan, A. Mastretta-Yanes, R. Vila, V. Dincă, J. Hernández-Roldán, et al. 2017. Climatic niche evolution is faster in sympatric than allopatric lineages of the butterfly genus *Pyrgus*. *Proceedings of the Royal Society B* 284:20170208. doi:10.1098/rspb.2017.0208.
- Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B* 267:2267–2272.
- Pyron, A. R., and J. J. Wiens. 2013. Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proceedings of the Royal Society B* 280:20131622. doi:10.1098/rspb.2013.1622.
- Rosauer, D. A. N., S. W. Laffan, M. D. Crisp, and S. C. Donnellan. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology* 18:4061–4072.
- Sabath, N., E. E. Goldberg, L. Glick, M. Einhorn, T.-L. Ashman, R. Ming, S. P. Otto, J. C. Vamosi, and I. Mayrose. 2016. Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms. *New Phytologist* 209:1290–1300.
- Skeels, A., and M. Cardillo. 2017. Environmental niche conservatism explains the accumulation of species richness in Mediterranean-hotspot plant genera. *Evolution* 71:582–594.
- Valente, L. M., G. Reeves, J. Schnitzler, I. P. Mason, M. F. Fay, T. G. Rebelo, M. W. Chase, and T. G. Barraclough. 2009. Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution* 64:745–760.
- Vidal-García, M., P. G. Byrne, J. D. Roberts, and J. S. Keogh. 2014. The role of phylogeny and ecology in shaping morphology in 21 genera and 127 species of Australo-Papuan myobatrachid frogs. *Journal of Evolutionary Biology* 27:181–192.
- Weih, C., U. Ligges, K. Luebke, and N. Raabe. 2005. klaR analyzing German business cycles. Pages 335–343 in D. Baier, R. Decker, and L. Schmidt-Thieme, eds. *Data analysis and decision support*. Springer, Berlin.
- Wiens, J. J., T. N. Engstrom, and P. T. Chippindale. 2006. Rapid diversification, incomplete isolation, and the “speciation clock” in North American salamanders (genus *Plethodon*): testing the hybrid swarm hypothesis of rapid radiation. *Evolution* 60:2585–2603.

Associate Editor: Roger K. Butlin
Editor: Daniel I. Bolnick