PNAS RESEARCH ARTICLE | EVOLUTION

Phylogenetic inference of where species spread or split across barriers

Michael J. Landis[a,1], Ignacio Quintero[b], Martha M. Muñoz[c], Felipe Zapata[d], and Michael J. Donoghue[c]

Edited by David Hillis, The University of Texas at Austin, Austin, TX; received September 17, 2021; accepted February 14, 2022

Regional features of geography, such as size or distance, are expected to shape how lineages disperse, go extinct, and speciate. Yet this fundamental link between geographical context and evolutionary consequence has not been fully incorporated into phylogenetic models of biogeography. We designed a model that allows variation in regional features (size, distance, insularity, and oceanic separation) to inform rates of biogeographic change. Our approach uses a Bayesian hierarchical modeling framework to transform regional values of quantitative and categorical features into evolutionary rates. We also make use of a parametric range split score to quantify range cohesion for widespread species, thereby allowing geographical barriers to initiate "range-splitting" speciation events. Applying our approach to *Anolis* lizards, a species-rich neotropical radiation, we found that distance between regions, especially over water, decreases dispersal rates and increases between-region speciation rates. For distances less than ∼470 km over land, anoles tended to disperse faster than they speciate between regions. Over oceans, the equivalent maximum range cohesion distance fell to ∼160 km. Our results suggest that the historical biogeography of founder event speciation may be productively studied when the same barriers that inhibit dispersal also promote speciation between regions.

phylogenetics | biogeography | speciation | dispersal | statistical inference

Significance

Geography molds how species evolve in space. Strong geographical barriers to movement, for instance, both inhibit dispersal between regions and allow isolated populations to diverge as new species. Weak barriers, by contrast, permit species range expansion and persistence. These factors present a conundrum: How strong must a barrier be before between-region speciation outpaces dispersal? We designed a phylogenetic model of dispersal, extinction, and speciation that allows regional features to influence rates of biogeographic change and applied it to the neotropical radiation of *Anolis* lizards. Separation by water induces a threefold steeper barrier to movement than equivalent distances over land. Our model will help biologists detect relationships between evolutionary processes and the spatial contexts in which they operate.

How species diversify, disperse, and go extinct differs in tempo and mode depending on where they live (1). For instance, broader access to suitable habitat types in larger regions can both drive the local formation of new species and buffer against extinction (2, 3). Long distances and strong barriers between regions can both limit dispersal (4) and instigate divergence in widespread species, possibly resulting in allopatric speciation (5). Additionally, terrestrial species that inhabit large contiguous continental regions may face characteristically different ecological and geographical conditions when compared to species inhabiting smaller and more isolated insular regions (like islands), which could also influence how those species evolve (6).

Such intuitive and widely accepted relationships that link regional features to biogeographic rates of evolution are nonetheless difficult to quantify. For example, the effect of distance on the frequency of dispersal (7) should yield specific and predictable evolutionary outcomes (8). Species ranges should tend to remain intact below some critical distance for maintaining range cohesion, but fragment into separate allopatric populations beyond this distance. It is also true that dispersal over a far enough distance to sever gene flow should ultimately yield new daughter species, rather than merely result in range expansion. But how far is far enough? And, for terrestrial species, assuming that water is indeed a strong barrier to organismal movement, we suppose that this distance should be shorter over water than over land. But how much shorter? Biogeography has no lack of described relationships marrying process to pattern (4, 5, 9–11) that, once quantified, can be tested across a wide variety of differing clades and regions (12).

From a phylogenetic perspective, testing hypotheses that link regional differences to evolutionary rates inherently connects to the biogeographic histories of lineages. Few clades have complete fossil records, so historical biogeography is often studied using phylogenetic event-based models to reconstruct the geographical past for living species (13). The two most widely used models for biogeographic inference are the Dispersal–Extinction–Cladogenesis (DEC) (14) and the Geographic State Speciation Extinction (GeoSSE) (15) models. Both approaches use phylogenetic stochastic processes to infer how species evolve among discrete regions through dispersal, regional extinction, and cladogenesis events. One important way in which they differ, however, is that GeoSSE allows species range configurations to influence speciation and extinction rates, whereas DEC does not. Even though biogeographers generally consider GeoSSE to be mechanistically superior to DEC (16–18), the computational and statistical performance of GeoSSE is understood to scale poorly for biogeographical systems with more than two regions. For example, the standard GeoSSE model is extremely

Author contributions: M.J.L., I.Q., M.M.M., F.Z., and M.J.D. designed research; M.J.L., I.Q., M.M.M., and F.Z. performed research; M.J.L. contributed new reagents/analytic tools; M.J.L. and I.Q. analyzed data; and M.J.L., I.Q., M.M.M., F.Z., and M.J.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: michael.landis@wustl.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116948119/-/DCSupplemental.

Published March 25, 2022.

parameter rich, requiring 7 free parameters for two regions, 120 parameters for five regions, and 9,420 parameters for nine regions (*SI Appendix*).

For this study, we developed a feature-informed GeoSSE model (FIG) that allows regional features, such as size or distance, to inform biogeographic rates of dispersal, extinction, and speciation. Rather than treating all possible event rates as freely estimated parameters, FIG, instead, models functional relationships between regional features and biogeographic rates, using 12 parameters regardless of how many regions are used. To illustrate the utility of FIG, we modeled a neotropical radiation of 379 lizard species (*Anolis*), whose conspicuous mainland–island distribution has been the subject of intense study (19–22). We tested whether, and how, four features (regional size, distance, insularity, and oceanic barrier presence) measured across nine biogeographic regions jointly shaped the processes of dispersal, extinction, within-region speciation, and between-region speciation for *Anolis* lizards. We found that, while region size and insularity had no detectable effect on *Anolis* biogeography, long distances, particularly over oceans, did. With these measurements in mind, and in light of an ongoing debate in the biogeographical literature (16, 17), we propose that founder event speciation may be modeled productively by using barriers and distance to align how dispersal and between-region speciation operate in space.

Model

FIG models biogeographic rates in terms of the local regional features that species encounter as they evolve (Fig. 1 and *Materials and Methods*). Species and their ranges (sets of inhabited discrete regions) stochastically change through four distinct processes: dispersal, (local and global) extinction, within-region speciation, and between-region speciation (Fig. 1*A*) (15). Event rates for each process may vary in response to multiple regional features; for example, within-region speciation rates might be highest in regions that are simultaneously large and continental (Fig. 1*B*), or

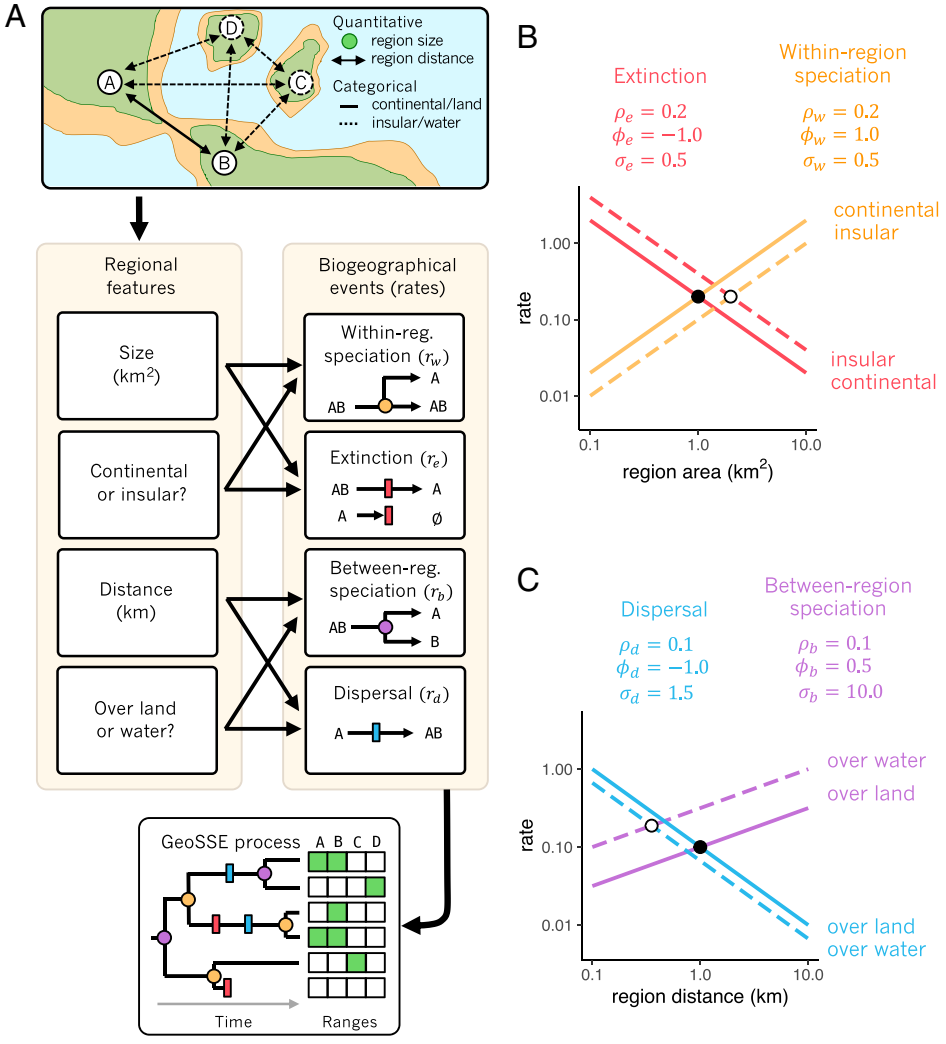

Fig. 1. Regional features, biogeographic events and rates, and example predictions for the FIG model. (*A*) Species may inhabit regions with suitable habitat (green) that are separated by barriers, such as arid zones (brown) or ocean (blue) (*Top*). Inhabitable regions have measurable quantitative and categorical features (*Middle Left*) that are transformed into evolutionary rates (*Middle Right*). Dispersal rates and between-region speciation rates are potentially informed by distance and the separation of land/water between regions. Extinction and within-region speciation rates are potentially informed by region size and type (continental vs. insular). FIG generates biogeographic range data (region presence/absence) conditional on its feature-informed rates and a time-calibrated phylogeny (*Bottom*). (*B* and *C*) Example FIG parameters (*Materials and Methods*) induce opposite responses in biogeographic rates to the same regional features. (*B*) Region size and region type (solid or dashed line) determine the within-region speciation rate (yellow) and extinction rate (red) for each region. (*C*) Distance and barrier type (solid or dashed line) determine the dispersal rate (blue) for any pair of regions and the between-region speciation rate (purple) for any range undergoing a "split." Note that the same regional feature may have different and asymmetrical effects on different biogeographic rates, depending on the model parameter values. Refer to *Materials and Methods* and *SI Appendix*, Table S1 for definitions and interpretations of FIG model parameters.

dispersal rates might be lowest between regions separated by long distances and major bodies of water (Fig. 1*C*). Even though FIG is primarily designed to infer whether such relationships are supported by the data—and, if so, what the signs and strengths of those relationships are—it can also generate ancestral range estimates and stochastic mappings (*SI Appendix*).

As a model, FIG defines four sets of relative rate factors that may influence region-specific rates of dispersal (m_d), extinction (m_e), within-region speciation (m_w), and between-region speciation (m_b). For this study, we assume that extinction and within-region speciation rate factors are informed by region sizes (square kilometers) and region types (insular [islands and/or archipelagoes] or continental), while dispersal and between-region speciation are informed by distances (kilometers) and barrier types (over water or land), and that all regional features held constant over time (*Discussion*). Exact rate factor values are determined by four categorical scaler parameters (σ_d, σ_e, σ_w, σ_b) that rescale the relative sizes of insular regions or over water distances, and four feature exponent parameters (ϕ_d, ϕ_e, ϕ_w, ϕ_b) that control the sign and scale for how the rescaled regional features relate to different evolutionary rates (*Materials and Methods*).

In practice, absolute event rates are computed as the product of a scaling base rate (ρ_d, ρ_e, ρ_w, ρ_b) and the relative rate factor that is directly (m_d, m_e, m_w) or indirectly (m_b through the range split score function, f_b) informed by the designated regional features (*Materials and Methods*). Following the event types of GeoSSE (15), species disperse from region i into j at rate $r_d(i,j) = \rho_d \times m_d(i,j)$. Local extinction of a species in region i proceeds at rate $r_e(i) = \rho_e \times m_e(i)$; if a species goes extinct in its final region, that lineage goes globally extinct and no longer diversifies or disperses. Within-region speciation produces a new species in the ancestral region i at rate $r_w(i) = \rho_w \times m_w(i)$. Lastly, between-region speciation "splits" (bipartitions) a widespread species range into two daughter lineages with ranges s and t at rate $r_b(s,t) = \rho_b \times f_b(s,t;m_b)$.

This formulation allows FIG to detect and quantify correlative relationships between regional features and evolutionary rates when provided a fixed time-calibrated phylogeny, species range data, and measurements of regional features. For example, if FIG estimates $\sigma_d = 3$ and $\phi_d = -1$, dispersal rates are assumed to have an inverse log-linear relationship to distance, with overwater distances effectively treated as 3 times as far as equivalent distances over land (*Materials and Methods*). FIG also permits regional features to have no effect on evolutionary rates. For instance, distance would have no effect on dispersal if $\phi_d = 0$, and distances over land and water would have equivalent effects on dispersal rates if $\sigma_d = 1$; FIG defines analogous relationships between other regional features and evolutionary rates. To detect which features have effects on evolutionary rates, we used Bayesian reversible-jump Markov chain Monte Carlo (RJMCMC) (23) to turn the effects for the eight feature exponent and categorical scaler parameters "on" ($\phi_p \neq 0$ or $\sigma_p \neq 1$) and "off" ($\phi_p = 0$ or $\sigma_p = 1$), where P_θ represents the marginal posterior probability that the parameter, θ, is "on." We consider $P_\theta > 0.95$ as support for a direct or indirect relationship between the relevant regional feature and biogeographic process.

Results

Power Analysis. We measured how reliably FIG detected the effect of regional features on evolutionary rates using simulated data. To do so, we simulated 200 datasets with random effect strengths for the feature exponent (ϕ_d, ϕ_e, ϕ_w, ϕ_b) and categorical scaler (σ_d, σ_e, σ_w, σ_b) parameters, and random tree sizes (from 20 to 250 tips), and then fitted FIG to each of those datasets using RJMCMC (*Materials and Methods*). The effect of regional features on evolutionary rates was detected most reliably (i.e., when the reversible-jump model probability is >0.95; *Materials and Methods*) when tree sizes were large and/or when parameter effect strengths were strong (Fig. 2). FIG reliably detected the effect of quantitative features (region size and distance) on evolutionary rates through the feature exponent parameters (ϕ_p) under a wider range of conditions than it detected effect of categorical features (insularity and barrier presence) on rates through the categorical scaler parameter (σ_p). In part, this is because the effect of a categorical scaler (σ_p) is throttled when its matching feature exponent (ϕ_p) has a weak effect (*Materials and Methods*). Nonetheless, for trees with ≥100 tips, the effect of distance on dispersal and the effect of region size on speciation could be detected when their corresponding feature exponents were only moderate in strength ($\phi_d < -1$ and $\phi_w > 1$, respectively). Trees with 175 or more tips and/or higher effect strengths were generally needed to reliably detect relationships between distance and between-region speciation ($\phi_b > 2$) or between region size and extinction rates ($\phi_e < -2$), possibly because these two evolutionary event types expose lineages to an increased risk of global extinction (i.e., data erasure) through range size reduction. In general, large trees with more than 175 tips and strong effects were necessary to infer the effect of categorical features on evolutionary rates. Barriers could be detected when they effectively doubled distances between regions for dispersal ($\sigma_d > 2$) and, less frequently, for between-region speciation ($\sigma_b > 2$). Only rarely could FIG detect the effect of insularity on within-region speciation ($\sigma_w < 1/3$) or extinction ($\sigma_e < 1/3$) rates. Last of all, FIG almost never detected effects between regional features and evolutionary rates when the parameter effect strength was low (i.e., when $\phi_p \approx 0$ or $\sigma_p \approx 1$), indicating that FIG has a low false discovery rate.

***Anolis* Biogeography.** We applied FIG to model the biogeography of *Anolis* lizards across nine neotropical regions (five continental, four insular; Fig. 3*A*, *Materials and Methods*, and *SI Appendix*). Roughly half (49%) of extant anoles inhabit one continental region, 43% inhabit one insular region, and the remaining 8% of species in ref. 21 are widespread among continental regions (Fig. 3*B*; see *SI Appendix* for caveats). Notably, four continental regions (B, C, D, E) and two insular regions (F, H) of vastly different sizes (10^5 km^2 to 10^7 km^2) each possess similar numbers of species (35 to 50 spp.).

Biogeographic rates for multiregion processes (m_d and m_b), but not single-region processes (m_e and m_w), varied substantially among regions. Dispersal (m_d; Fig. 3*C*, edges) favored movement between adjacent continental regions. Similar in pattern, but opposite in manner, rates of between-region speciation were elevated for widespread ranges that involved insular regions and/or the most-distant continental regions (m_b; Fig. 3*D*, edges). Extinction rates (m_e; Fig. 3*C*, nodes) were nearly identical, regardless of region size or island status, except that the posterior mean rate for Amazonia (region E) was higher than average. Within-region speciation rates (m_w; Fig. 3*D*, nodes) were also relatively constant across regions.

Between-region speciation rates were much higher than dispersal rates for all pairs of regions that were separated by the Caribbean (Fig. 3*E*). Among continental regions, however, dispersal often met or even exceeded the rate of between-region

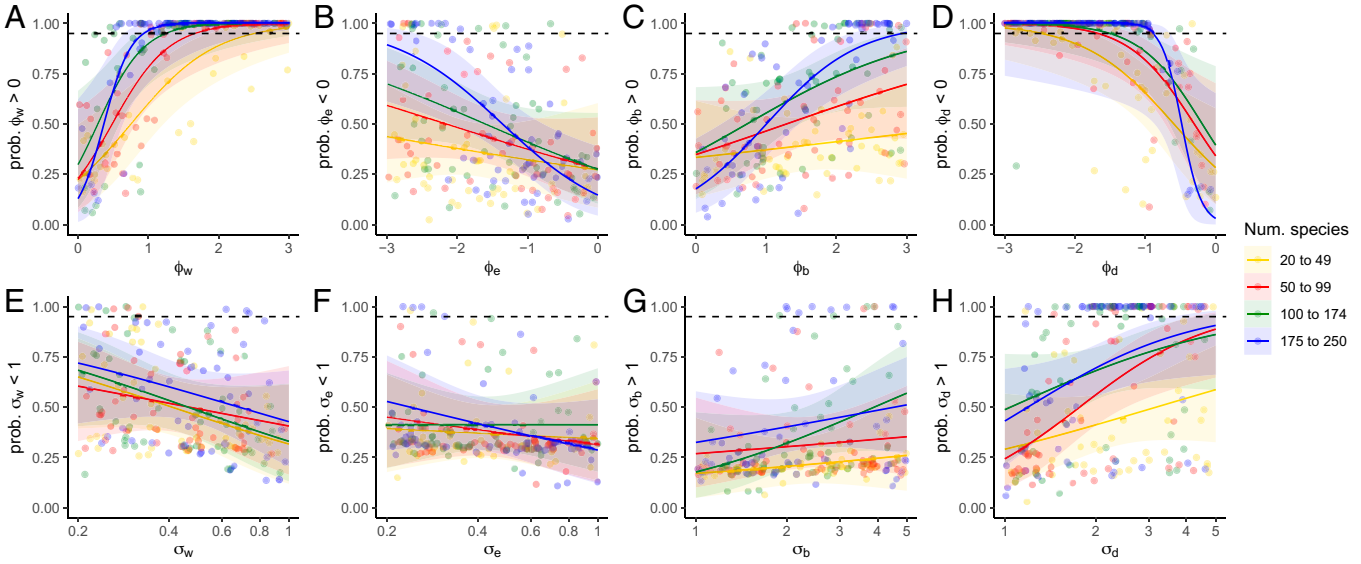

Fig. 2. Hypothesis testing power for FIG. Panels show the posterior model probabilities that RJMCMC correctly detects the effect of region size increasing within-region speciation rates ($\phi_w > 0$; *A*) or decreasing extinction rates ($\phi_e < 0$; *B*), the effect of distance increasing between-region speciation rates ($\phi_b > 0$; *C*) or decreasing dispersal rates ($\phi_d < 0$; *D*), the effect of insularity increasing the impact of region size on within-region speciation rates ($\sigma_w < 1$; *E*) or extinction rates ($\sigma_e < 1$; *F*), and the effect of water increasing the impact of distance on between-region speciation rates ($\sigma_b > 1$; *G*) or dispersal rates ($\sigma_d > 1$; *H*). Each estimate (point) gives the posterior probability (*y* axis) of correctly detecting an effect and its sign relative to the effect strength (*x* axis) and tree size (20 to 49 species, yellow; 50 to 99 species, red; 100 to 174 species, green; 175 to 250 species, blue). Estimates above a black dashed line correctly detected the simulated effect, with a probability of >0.95. Logistic regression curves summarize the distribution of effect strengths and detection probabilities for each parameter and tree size category. Shaded regions indicate regression confidence intervals.

speciation. Geographical distance was strongly associated with decreased dispersal rates (posterior median and 95% highest posterior density interval [HPD], $\phi_d = -1.03$ [-1.27, -0.80]) and increased between-region speciation rates ($\phi_b = 1.77$ [0.85, 2.83]; Table 1). For regions separated by ocean, those distances were effectively tripled for between-region speciation rates ($\sigma_b = 3.34$ [1.45, 5.76]) and dispersal ($\sigma_d = 3.45$ [1.70, 5.69]; Table 1) when compared to equivalent distances over land. In other words, water presents a formidable barrier to movement in anole lizards, both in terms of range expansion and in the maintenance of range cohesion.

Linear models predicted that dispersal outpaced speciation when distances over land were shorter than 279 [156, 472] km and shorter than 97 [46, 164] km over water (Fig. 3*F* and *Materials and Methods*). Validating this maximum range cohesion distance (MRCD) estimate against other aspects of our study, we found that 6 out of 20 possible ordered pairs of continental regions (BA, BC, CD, CE, DC, DE; Fig. 3*F*) had distances below the upper limit of 472 km, which defines a set of adjacent regions occupied by extant widespread continental anoles (Fig. 3*B*). No insular region pairs fell below the analogous over-water MRCD threshold of 164 km.

Within regions, speciation occurred more often than extinction (Fig. 3*G*). An RJMCMC analysis could not determine a positive or negative relationship between region size and extinction rates ($P_{\phi_e} = 0.47$) or speciation rates ($P_{\phi_w} = 0.20$; Table 1). Because of this, linear models do not predict a narrow window of region sizes in which *Anolis* extinction rates exceed speciation rates within regions (Fig. 3*H*).

Phylogenetic uncertainty had no major effect on our parameter estimates (*SI Appendix*, Fig. S3). Combined posterior estimates for 15 sampled phylogenies from the original Poe et al. (21) output deviated little from our primary estimates that we obtained using the maximum clade credibility tree (Table 1 and *SI Appendix*, Fig. S3). Combined reversible jump probabilities across posterior tree samples also favored relationships for distance and water with dispersal ($P_{\phi_d} \approx 1$, $P_{\sigma_d} \approx 1$) and between-region speciation ($P_{\phi_b} = 0.98$, $P_{\sigma_b} = 0.96$), and not for regional features with extinction ($P_{\phi_e} = 0.53$, $P_{\sigma_e} = 0.50$) or within-region speciation ($P_{\phi_w} = 0.19$, $P_{\sigma_w} = 0.50$).

Discussion

Geographical distances were simultaneously related to decreased dispersal rates and increased between-region speciation rates in *Anolis* (Table 1). When we assume that dispersal rates represent how easily species move between regions, our estimates suggest that *Anolis* ranges lose cohesion and tend to split beyond ~500 km for continental species (Table 1). Only continental pairs of regions fell below this threshold, and these are precisely the regions that are, today, inhabited by widespread anoles (Fig. 3 *B* and *F*). Water is a more formidable barrier to movement in *Anolis*, as insular regions were effectively 3 times farther apart than continental regions when relating distance to dispersal and between-region speciation (Table 1). The shortest distance among insular regions (an average of 266 km from the Bahamas [I] into Cuba [G]) was still greater than the oceanic MRCD of about 160 km. Numerous phylogeographic studies in *Anolis* similarly found that intraspecific genetic divergence often met or surpassed typical levels of interspecific divergence ($F_{ST} > 0.3$) (24) when anole populations were separated by hundreds of kilometers over land and, in some cases, just tens of kilometers over water (Dataset S1).

The brown anole (*Anolis sagrei*), which is naturally found in both Cuba and the Bahamas, provides plausible support for a single species spanning multiple insular regions. Recent work, however, points to surprisingly deep divergences (2.8 Ma to 4.1 Ma) between the *A. sagrei ordinatus* subspecies in the Bahamas and its closest relatives in Cuba (25). Even within-island divergence in Cuba is steep for *A. sagrei* (25), and for other island anoles, like *Anolis distichus* from Hispaniola, which has suggested the need for taxonomic revision (e.g., ref. 26). Separate exploratory analyses that recoded the Poe et al. (21) range for *A. sagrei* to include the occurrence of the *A. sagrei ordinatus*

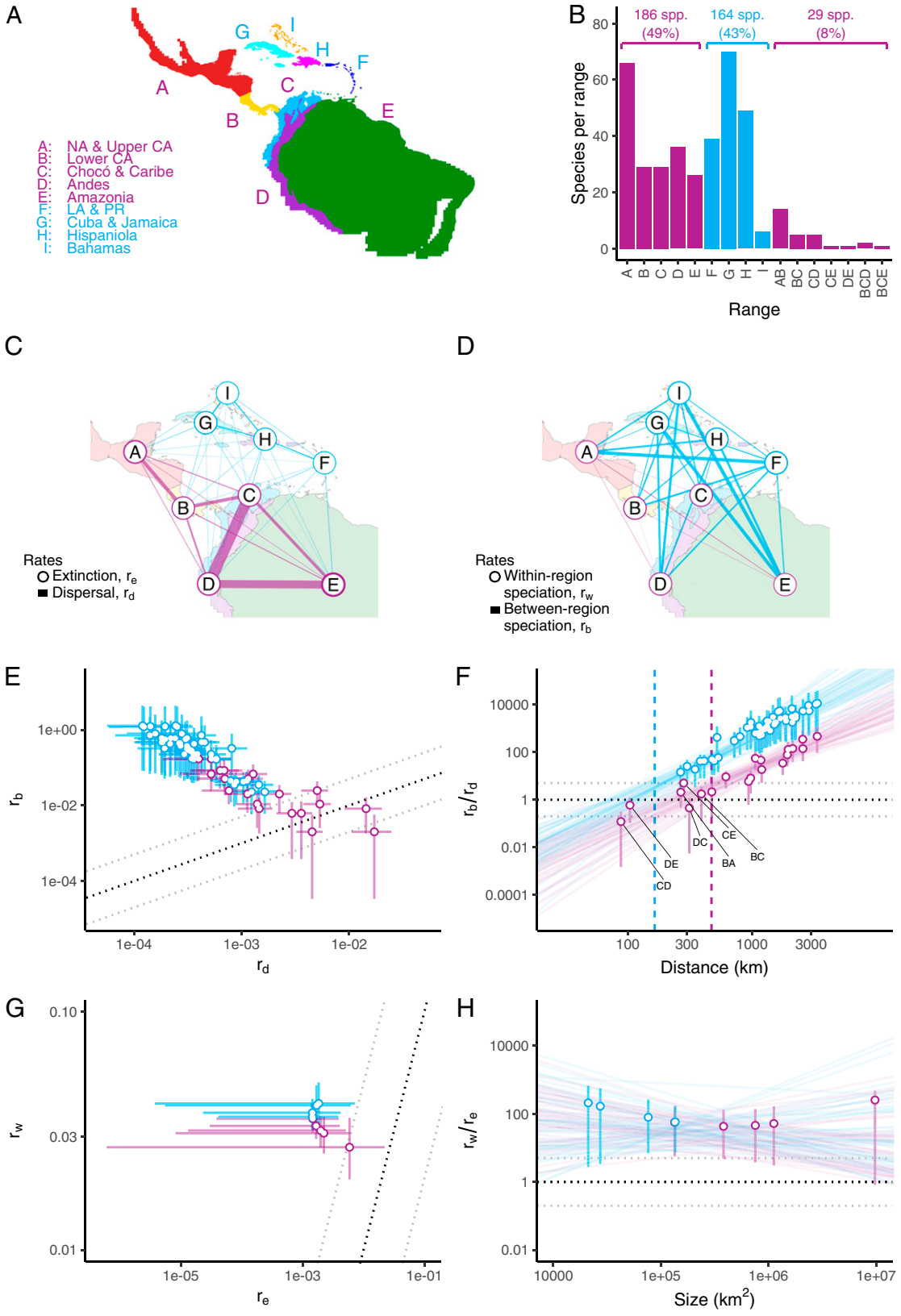

Fig. 3. Inferred relationships between regional features and biogeographic rates in *Anolis*. (*A*) Five continental (purple) and four insular (blue) regions drawn in proportion. (*B*) Counts of geographical ranges (sets of inhabited regions) among extant species. Counts and percentages for continental endemics (left), insular endemics (center), and widespread continental (right) species are listed above the histogram. (*C* and *D*) Relative regional rate estimates are displayed as networks. Thicker widths for node borders correspond to larger rates of extinction (*C*) or within-region speciation (*D*). Thicker widths for edges correspond to larger rates of dispersal (*C*) or between-region speciation (*D*); dispersal features are averaged across each source–destination pair. (*E*) Rates of dispersal (*x* axis) versus between-region speciation (*y* axis) for all pairs of two continental regions (purple) and all pairs that involve any insular regions (blue). (*F*) Ratios of between-region speciation over dispersal rates (*y* axis) plotted against distance (*x* axis). The vertical dashed lines approximate the maximum distance where $r_d \geq r_b$ over land (purple) and over water (blue). (*G*) Rates of extinction (*x* axis) versus within-region speciation (*y* axis) for all continental regions (purple) and insular regions (blue). (*H*) Ratios of within-region speciation over extinction rates (*y* axis) plotted against size (*x* axis). For *E*–*H*, rates are plotted as posterior means (points) and their 95% highest posterior densities (segments). Dotted lines indicate rate ratios of 1:1 (black) and 5:1 or 1:5 (gray).

Table 1. Posterior estimates for analysis of Anolis lizards

Variable	Mean	Lower CI	Upper CI	RJ prob.	Interpretation
ρ_d	5.12e-4	2.04e-4	8.94e-4	–	–
ρ_e	1.50e-3	3.02e-7	3.93e-3	–	–
ρ_w	3.45e-2	3.04e-2	3.84e-2	–	–
ρ_b	1.37e-1	3.65e-2	2.91e-1	–	–
ϕ_d	−1.03	−1.27	−0.80	1.00	Dispersal limited by distance
ϕ_e	0.13	−0.97	1.00	0.47	Extinction independent of size
ϕ_w	−0.07	−0.14	0.01	0.20	Within-region speciation independent of size
ϕ_b	1.77	0.85	2.83	0.99	Between-region speciation accelerated by distance
σ_d	3.45	1.70	5.69	1.00	Dispersal limited by water
σ_e	1.07	0.18	2.45	0.50	Extinction independent of insularity
σ_w	0.88	0.22	1.91	0.50	Within-region speciation independent of insularity
σ_b	3.34	1.45	5.76	0.99	Between-region speciation accelerated by water
MRCD, over land	279	156	472	–	Measured in kilometers; see main text
MRCD, over water	97	46	164	–	Measured in kilometers; see main text

Credible interval (CI) bounds were computed from 95% highest posterior densities. CIs that did not contain $\phi_p = 0$ or $\sigma_p = 1$ were considered significant; no significance test was applied to ρ_p. Reversible jump probabilities (RJ prob.) greater than 0.95 also indicate support that $\phi_p \neq 0$ or $\sigma_p \neq 1$. Interpretations follow Table 2 (*Materials and Methods*).

subspecies (25) in the Bahamas had no discernible impact on our parameter estimates (discussed in *SI Appendix*). We expect that our findings are robust to other comparably minor adjustments to *Anolis* taxonomy, phylogeny, and/or geographic ranges that we used in our analysis.

Previous work concentrating on Caribbean anoles found that speciation rates increased with island size (20, 27). Our analysis found no conclusive relationships between region size and within-region speciation rates or extinction rates across coarser continental and insular regions (Table 1), nor did we find that insularity influenced these rates (28), which is consistent with recent studies (22).

Lending credence to our estimates, we note that the size of our *Anolis* dataset (379 ingroup taxa, nine regions) furnished our analysis with more statistical power than the largest datasets (250 taxa, five regions) in our simulation study. Even phylogenies with as few as 50 taxa may be sufficient for FIG to identify cases where within-region speciation and dispersal are influenced by regional features. Larger phylogenies with >175 taxa may be necessary to detect analogous relationships for extinction and between-region speciation. Because the effect of a categorical scaler (σ_p) indirectly depends on its corresponding feature exponent (ϕ_p), it is difficult to infer the former as $\sigma_p \neq 1$ when the latter is estimated as $\phi_p \approx 0$ (*Materials and Methods* and *SI Appendix*, Fig. S1). That said, our power analysis suggests that FIG has a low false discovery rate for detecting the relationships between features and rates, which should make the model safe to apply to radiations of any size with little risk of being positively misled.

Furthermore, the *Anolis* parameter estimates that we derived from a maximum clade credibility tree (Table 1 and Fig. 3) were highly congruent with those derived from 15 posterior tree samples (*SI Appendix*, Fig. S3). Biogeographers should suspect that phylogenetic uncertainty may influence FIG estimates in other clades, particularly for those phylogenies with few taxa, weak clade support, and/or highly uncertain node ages. At present, we suggest combining posterior estimates across phylogenetic hypotheses to assess the impact of phylogenetic uncertainty (*SI Appendix*). The ideal solution, however, will be to jointly estimate phylogeny and biogeography under FIG. Joint estimation not only is the correct statistical approach for handling phylogenetic uncertainty (29), it also creates opportunities for biogeography to inform phylogenetic divergence times (30, 31) and taxon placement (e.g., for fossils) (32).

We emphasize that the four regional features we selected (size, distance, insularity, and barrier presence) are likely only indirect drivers of biogeographic diversification (33, 34). Size, distance, and other factors may be strongly correlated with a host of more direct drivers, including ecological complexity, topographical complexity, climatological stability, and region age—drivers that are generally more difficult to measure and model. Such factors are likely key to uncovering what forces govern adaptive radiation. For instance, Amazonia is two orders of magnitude larger than Hispaniola but contains fewer than half as many species. In this case, direct measurements of ecological and geographical complexity may predict differences between insular and continental anoles better than size alone. We also note that our analysis used modern-day features to model biogeographic diversification. Previous research suggests that modeling historical features of paleogeography is not always necessary to accurately estimate biogeographic parameters (35). While we do not expect that modeling Caribbean paleogeography would change our findings that distance and water act as barriers for *Anolis* movement, it is possible that fluctuations in island sizes over time detectably influenced within-region speciation and extinction rates.

More broadly, this line of inquiry reveals one purpose for the model design of FIG: It allows biologists to study a wide range of first-order biogeographic hypotheses concerning how regional features are related to evolutionary dynamics for different clades and regions (12). It also establishes a framework to test second-order hypotheses, such as whether biogeographic tendencies are correlated with the phylogenetic distributions of traits. For example, does the tendency for species ranges to spread or split correlate with dispersal mode (36), physiology (37, 38), reproductive system (39), ploidy level (40), or ecology (41)?

The feature-informed rates of FIG also mitigate concerns raised by biogeographers (13, 15, 16, 18, 42), that GeoSSE-type models are doomed to be overparameterized when applied to systems involving more than two or three regions. When available, additional regional features can readily be incorporated into the construction of the parameterized geography functions (m_b, m_d, m_w, m_e). Independent of the number of regions, each additional feature would require at least one additional model parameter, and its explanatory power could be assessed using RJMCMC for Bayesian model selection.

Range split scores (*Materials and Methods*) are central to the model's efficiency, as they allow a small number of parameters

to modulate the explosive number of between-region speciation rates (*SI Appendix*) through a biologically plausible mechanism (e.g., allopatric speciation via isolation by distance). In particular, regions separated by the weakest barriers prevent a species range from splitting (e.g., gene flow generates range cohesion) and thus generate the lowest split scores. Even trace amounts of gene flow between the most distant regions in a widespread range will marginally decrease the split score. None of this, however, implies that widespread ranges split at slower rates than species with narrower ranges, as larger ranges may split in more distinct ways, and thus have greater total rates of splitting in any way. Although we interpret the mechanism for range splitting in anoles as allopatric speciation between regions, ecological speciation with gene flow should behave similarly between regions when the feature catalyzing divergence is spatially autocorrelated (11).

Several aspects of FIG are admittedly "crude" (2) but not beyond repair. Reformulated range split scores, for instance, could depend directly on metrics of genetic differentiation (e.g., F_{ST}) rather than on its proxies, like distance. It is also possible to link phylogenetic shifts in biogeographic tempo and mode to observed (43) and hidden (18) evolutionary factors. Paleogeographical features, which may influence estimates of ancestral ranges and rates (44), can inform evolutionary rates through parameterized functions of time and space (30). In model development, first steps are necessary before taking such next steps (45, 46).

Still, even simple models can cast light upon complex issues. Few biologists would be surprised that our model inferred that anoles rarely disperse great distances or over water or that, after they do, they tend to imminently split into two species. This finding does, however, highlight a deeper unresolved problem in historical biogeography concerning founder event speciation. Conceptual models of the phenomenon begin with any number of colonists expanding the range of the progenitor species into a new region, with the new colonizing population having sufficiently low levels of gene flow with the progenitor population to allow rapid divergence and the fixation of genetic incompatibilities, resulting in allopatric speciation (4). This has provoked critical discussion regarding what statistical features are necessary, or can even be estimated, when modeling founder event speciation in a phylogenetic context (14–17, 47).

Here, we take a particular interest in whether founder event speciation should be modeled as one compound event (i.e., dispersal with speciation) or two sequential events (i.e., dispersal, then speciation). We conclude from our analysis of *Anolis* biogeography that two-event models can help refine our understanding of geographical speciation. Dispersal and between-region speciation are both oriented around a common factor, namely, the movement of individuals (and genes) across distances and barriers, and our biogeographic models should somehow recognize their shared spatial basis. Widespread ranges comprising adjacent regions, where gene flow is uninhibited, are expected to be more genetically cohesive and evolutionarily stable when compared to more fragmented ranges with less gene flow that is attenuated across distant regions (48). Rates of allopatric speciation should be lower in the first case of nearby regions. Short distances and the absence of barriers should permit widespread ranges to persist, and, conversely, long distances and the presence of barriers should result in rare dispersal events followed by rapid speciation events.

Put more simply, what limits dispersal should also promote allopatric speciation. Looking ahead, we think it will be important for models to distinguish between interpopulation migration, the basis for gene flow, and dispersal, the basis for range expansion. Migration [*sensu* Wright (49)] is the movement of individuals/genes among established populations, whereas dispersal involves the movement of individuals to establish new populations. It is unclear whether migration rates or dispersal rates should generally be higher given the same set of regional features. For example, divergent ecological selection among demes might limit the entry of migrants and thus limit gene flow. On the other hand, individuals dispersing into new ecosystems might be too few in number to establish new populations, or they might be competitively excluded due to priority effects. Carefully parsing the relationships among migration, dispersal, and speciation will help to unify microevolutionary and macroevolutionary perspectives in phylogenetic biogeography.

Materials and Methods

FIG Model Definition. We developed a hierarchical phylogenetic model (FIG) to relate regional features to evolutionary rates of biogeography. *SI Appendix* diagrams FIG as a probabilistic graphical model (*SI Appendix*, Fig. S1) and provides a compact summary of all model variables (*SI Appendix*, Table S1). FIG is a GeoSSE (15) variant, so it is composed of four biogeographic processes: dispersal (d), extinction (e), within-region speciation (w), and between-region speciation (b). Relationships between regional features and relative evolutionary rates are primarily governed by two sets of parameters. Categorical scaler parameters (σ_d, σ_e, σ_w, σ_b) rescale the effective value of a quantitative feature (e.g., size or distance) depending on the corresponding categorical feature for a region (e.g., insular region or water barrier). Feature exponent parameters (ϕ_d, ϕ_e, ϕ_w, ϕ_b) control the sign and scale of the relationship between a quantitative regional feature and an effect on the matched evolutionary rate. Values for all eight parameters are estimated from the data during inference, where each parameter can induce a different relationship between one regional feature and one evolutionary rate (Table 2) by defining relationships between features and evolutionary effects using the equations below. This version of FIG assumes that regional feature measurements are constant over time; this assumption will be relaxed in future versions of the model.

Measurements for regional features were transformed into relative evolutionary rates in two steps. To simplify notation, we use the subscript p to indicate variables for processes d, e, w, b. First, we constructed four separate sets of categorically scaled features, g_p, for the extinction ($p = e$) and within-region speciation ($p = w$) processes as

$$g_p(i) = \begin{cases} y_q(i) & \text{if } y_c(i) \text{ is continental} \\ y_q(i) \times \sigma_p & \text{if } y_c(i) \text{ is insular} \end{cases},$$

where the vector y_q quantifies a feature (e.g., size) and y_c encodes categorical values (e.g., continental or insular) for region i, and σ_p is the categorical scaler that rescales quantitative features for insular regions when determining relative within-region speciation or extinction rates among regions. For example, if $\sigma_e = 0.5$, then the extinction process will consider an insular region of size 4 and a continental region of size 2 as both having the same effective size.

Similar functions for dispersal ($p = d$) and between-region speciation ($p = b$) were defined as

$$g_p\left(i,j\right) = \begin{cases} z_q(i,j) & \text{if } z_c(i,j) \text{ is over land} \\ z_q(i,j) \times \sigma_p & \text{if } z_c(i,j) \text{ is over water} \end{cases},$$

where, for each ordered pair of regions, the matrix $z_q(i,j)$ quantifies a feature (e.g., distance) and $z_c(i,j)$ indicates a categorical value (e.g., presence/absence of seaway barrier), and σ_p is a categorical scaler that rescales distances over water relative to distances over land for dispersal and between-region speciation. Dispersal would consider distances over water to be effectively 3 times greater than the same distances over land if $\sigma_d = 3$, for instance.

In practice, we normalized y_q and z_q by their respective means to simplify how we interpret the scale of each σ_p relative to the quantitative features.

Each set of scaled features (g_p) was then transformed into a set of exponentiated features, m_p, for extinction ($p = e$) and within-region speciation ($p = w$) as

$$m_p(i) = \left[\frac{g_p(i)}{gm(g_p)}\right]^{\phi_p}$$

Table 2. FIG hypothesis testing

Parameter	Interpretation
Feature exponent $-\infty < \phi_p < \infty$	Size has no effect on extinction ($\phi_e = 0$) or within-region speciation rate ($\phi_w = 0$)
	Distance has no effect on dispersal ($\phi_d = 0$) or between-region speciation rate ($\phi_b = 0$)
	Size decreases extinction ($\phi_e < 0$) or within-region speciation rate ($\phi_w < 0$)
	Distance decreases dispersal ($\phi_d < 0$) or between-region speciation rate ($\phi_b < 0$)
	Size increases extinction ($\phi_e > 0$) or within-region speciation rate ($\phi_w > 0$)
	Distance increases dispersal ($\phi_d > 0$) or between-region speciation rate ($\phi_b > 0$)
Categorical scaler $0 < \sigma_p < \infty$	Insular region sizes unchanged for extinction ($\sigma_e = 1$) or within-region speciation ($\sigma_w = 1$)
	Over water distances unchanged for dispersal ($\sigma_d = 1$) or between-region speciation ($\sigma_b = 1$)
	Insular region sizes effectively decreased for extinction ($\sigma_e < 1$) or within-region speciation ($\sigma_w < 1$)
	Over water distances effectively decreased for dispersal ($\sigma_d < 1$) or between-region speciation ($\sigma_b < 1$)
	Insular region sizes effectively increased for extinction ($\sigma_e > 1$) or within-region speciation ($\sigma_w > 1$)
	Over water distances effectively increased for dispersal ($\sigma_d > 1$) or between-region speciation ($\sigma_b > 1$)

Different values for the four feature exponent (ϕ_d, ϕ_e, ϕ_w, ϕ_b) and four categorical scaler (σ_d, σ_e, σ_w, σ_b) parameters determine if and how regional features (size, distance, region type, barrier type) influence biogeographic rates. Larger quantitative features (size, distance) increase the relevant regional rates when $\phi_p > 0$, decrease rates when $\phi_p < 0$, and leave rates unchanged when $\phi_p = 0$. Categorical features (insularity, ocean barrier presence) effectively increase the corresponding quantitative feature (size, distance) value when $\sigma_p > 1$, effectively decrease the value when $\sigma_p < 1$, and leave values unchanged when $\sigma_p = 1$. Additional detail is provided in *Materials and Methods*.

and, similarly, for dispersal ($p = d$) and between-region processes ($p = b$) as

$$m_p(i,j) = \left[\frac{g_p(i,j)}{gm(g_p)}\right]^{\phi_p},$$

where $gm(g_p)$ is the geometric mean for the scaled features, g_p. For example, suppose that the scaled feature for extinction in region A was $g_e(\mathrm{A}) = 4$, and the corresponding geometric mean across all scaled features for extinction was $\mathrm{gm}(g_e) = 2$. If the feature exponent $\phi_e = -1$, then the exponentiated feature $m_e(\mathrm{A}) = [4/2]^{-1} = 1/2$; that is, the extinction rate in region A would be half that for any region whose size equaled the geometric mean of all region sizes.

Absolute state-dependent rates of evolution (r_d, r_e, r_w, r_b) are then computed as the product of a base rate parameter (ρ_d, ρ_e, ρ_w, ρ_b) and a term involving the respective exponentiated features (m_d, m_e, m_w, m_b) for each event type: $r_w(i) = \rho_w \times m_w(i)$ is the rate of within-region speciation in region i; $r_e(i) = \rho_e \times m_e(i)$ is the rate of extinction in region i; $r_d(i,j) = \rho_d \times m_d(i,j)$ is the rate of dispersal from region i into j; and $r_b(s,t) = \rho_b \times f_b(s,t;m_b)$ is the rate at which the ancestral range, $s \bigcup t$, splits into the nonempty and complementary daughter ranges s and t; the range split score function, f_b, is defined below. All rates are then assigned to a GeoSSE process for simulation and inference.

Range Split Scores. The range split score computes the relative rate of between-region speciation, where higher scores accelerate speciation. Range split scores treat each range as a fully connected graph, with nodes corresponding to inhabited regions and edge weights measured by the parametrized geography matrix, m_b. The cutset, $E(s,t)$, defines the minimal set of edges that must be removed (cut) from the ancestral range graph, a, to fully separate the daughter ranges graphs, s and t, into distinct components. We then define the range split score function as

$$f_b(s,t;m_b) = \frac{1}{\mathrm{gm}(f_b)} \times \left[\sum_{(u,v)\in E(s,t)} \bar{m}_b(u,v)^{-1}\right]^{-1},$$

which is the inverse sum of inverse weights between all pairs of regions in the cutset, with edge weights between nodes u and v equaling $\bar{m}_b(u,v) = [m_b(u,v) + m_b(v,u)]/2$, normalized by the geometric mean of all range split scores. Each additional edge in the cutset for a particular split decreases its corresponding range split score (e.g., increases range cohesion). Small range split scores are dominated by those cut edges with the smallest weights (e.g., nearby region pairs), while larger-valued weights further decrease the split score only marginally.

Power Analysis for Hypothesis Testing. We analyzed 200 simulated datasets using FIG to assess how reliably the model correctly detected the effect of regional features on evolutionary rates. All simulations used the same five-region system that contained variation in region size, distance, region type (continental vs. insular), and barrier type (over land vs. over water; *SI Appendix*, Fig. S2). All simulations ran for ten units of evolutionary time under the same base rates ($\rho_d = 0.1$, $\rho_w = 0.1$, $\rho_e = 0.1$, $\rho_b = 0.2$), but each simulation randomly sampled its own effect strengths for feature exponent (ϕ_d, ϕ_e, ϕ_w, ϕ_b) and categorical scaler (σ_d, σ_e, σ_w, σ_b) parameters. Our simulations (but not inferences) assumed that larger region sizes decreased extinction rate, $\phi_e \sim \mathrm{Unif}(-3,0)$, and increased within-region speciation rate, $\phi_w \sim \mathrm{Unif}(0,3)$, and insular region size was effectively reduced relative to equivalent continental region sizes, σ_e, $\sigma_w \sim \mathrm{Logunif}(0.2,1.0)$. Longer distances between regions decreased dispersal rates, $\phi_d \sim \mathrm{Unif}(-3,0)$, and increased between-region speciation rates, $\phi_b \sim \mathrm{Unif}(0,3)$, and distances over water were effectively increased relative to equivalent distances over land, σ_d, $\sigma_b \sim \mathrm{Logunif}(1,5)$. We obtained 50 simulated datasets for each of four tree size categories: 20 to 49 species, 50 to 99 species, 100 to 174 species, and 175 to 250 species. We then estimated the marginal posterior model probabilities that the correct signed nonzero estimate was obtained for each ϕ_p and σ_p parameter under FIG using RJMCMC. Model probabilities greater than 0.95 were used to classify when FIG correctly detected the effect of a regional feature on an evolutionary rate. Simulated datasets were analyzed using RevBayes (50).

Maximum Range Cohesion Distance. We predicted the maximum distance at which dispersal rates exceeded between-region speciation rates ($r_b/r_d \leq 1$) using the y-intercept estimate of the log-linear model, $\log_{10}(z_q) = \log_{10}(r_b/r_d) + \varepsilon$. Separate posterior means and HPDs for range cohesion distances were computed for continental and insular regions. We preferred the HPD upper bound as a conservative estimator for MRCD.

***Anolis* Biogeography.** We reanalyzed the Poe et al. (21) *Anolis* lizard dataset, which described 379 anole species ranges across 14 neotropical regions. We collapsed their regions into a system with five continental regions (Mexico and Upper Central America [A]; Lower Central America [B]; the Chocó and Interandean Valleys [C]; the Andes [D]; Amazonia [E]) and four insular regions (the Lesser Antilles and Puerto Rico [F]; Cuba, Jamaica, and the Caymans [H]; Hispaniola [G]; and the Bahamas [I]). Region sizes and spatially averaged distances were computed using modern day features of geography and anole ranges (*SI Appendix*). Maximum range sizes were limited to four regions (255 possible range states). All species ranges from Poe et al. (21) were translated directly into our collapsed system of regions, except for four that required minor manual adjustments. *SI Appendix* describes the dataset treatment and the phylogenetic uncertainty analysis in detail and expands on our interpretation of our *Anolis* results and ancestral range estimates (*SI Appendix*, Fig. S4).

Bayesian Model Design and Inference. Model design and inference was performed in RevBayes (commit 724fa6a204) (50). SSE model likelihoods were computed using the TensorPhylo plugin for RevBayes (commit 5cfd2bfc93) (51). All simulated and empirical analyses used the same priors: $\rho_p \sim \mathrm{Exp}(1)$, $\phi_p \sim \mathrm{Norm}(0,1)$, and $\phi_p \sim \mathrm{Lognorm}(-0.172, 0.587)$. Standard MCMC was used to generate the main empirical results and parameter estimates. RJMCMC analyses were conducted to estimate the model probabilities for $\phi_p \neq 0$

and $\sigma_p \neq 1$ for the simulation study and for the empirical analysis. MCMC mixing and convergence for the *Anolis* parameter estimates and reversible-jump probabilities (Table 1) were validated by visual inspection and through small potential scale reduction factors (≤ 1.05) (52).

Data Availability. RevBayes analysis scripts, R plotting scripts, 200 simulated datasets, two versions of the *Anolis* dataset, and the MCMC results for the simulation study and empirical study are hosted through the public GitHub repository (https://github.com/mlandis/fig_model) (53).

ACKNOWLEDGMENTS. We are grateful to Tracy Heath, Jonathan Losos, the editor, and two anonymous reviewers for their feedback on this work, and to Steven Poe and Julián Velasco for sharing data and advice relating to their monumental 2017 study. This research was supported by the NSF (Grant DEB-2040347 to M.J.L., Grant DEB-2040081 to F.Z.), start-up funds from Washington University in St. Louis (M.J.L.), and the Yale Institute for Biospheric Studies (M.J.L.). This project also received funding from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant Agreement 897225 (I.Q.).

Author affiliations: [a]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130; [b]Institut de Biologie, École Normale Supérieure, 75005 Paris, France; [c]Department of Ecology & Evolutionary Biology, Yale University, New Haven, CT 06520; and [d]Department of Ecology & Evolutionary Biology, University of California, Los Angeles, CA 90095

1. R. E. Ricklefs, Community diversity: Relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
2. R. H. MacArthur, E. O. Wilson, *The Theory of Island Biogeography* (Princeton University Press, Princeton, NJ, 1967).
3. R. J. Whittaker, K. A. Triantis, R. J. Ladle, A general dynamic theory of oceanic island biogeography. *J. Biogeogr.* **35**, 977–994 (2008).
4. S. Carlquist, The biota of long-distance dispersal. I. Principles of dispersal and evolution. *Q. Rev. Biol.* **41**, 247–270 (1966).
5. E. Mayr, *Systematics and the Origin of Species* (Columbia University Press, New York, NY, 1942).
6. B. H. Warren *et al.*, Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
7. R. Nathan, Long-distance dispersal of plants. *Science* **313**, 786–788 (2006).
8. L. R. V. de Alencar, T. B. Quental, Linking population-level and microevolutionary processes to understand speciation dynamics at the macroevolutionary scale. *Ecol. Evol.* **11**, 5828–5843 (2021).
9. H. F. Osborn, The law of adaptive radiation. *Am. Nat.* **36**, 353–363 (1902).
10. G. Nelson, N. I. Platnick, A vicariance approach to historical biogeography. *Bioscience* **30**, 339–343 (1980).
11. H. D. Rundle, P. Nosil, Ecological speciation. *Ecol. Lett.* **8**, 336–352 (2005).
12. M. D. Crisp, S. A. Trewick, L. G. Cook, Hypothesis testing in biogeography. *Trends Ecol. Evol.* **26**, 66–72 (2011).
13. R. H. Ree, I. Sanmartín, Prospects and challenges for parametric models in historical biogeographical inference. *J. Biogeogr.* **36**, 1211–1220 (2009).
14. R. H. Ree, B. R. Moore, C. O. Webb, M. J. Donoghue, A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* **59**, 2299–2311 (2005).
15. E. E. Goldberg, L. T. Lancaster, R. H. Ree, Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.* **60**, 451–465 (2011).
16. N. J. Matzke, Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. *Syst. Biol.* **63**, 951–970 (2014).
17. R. H. Ree, I. Sanmartín, Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.* **45**, 741–749 (2018).
18. D. S. Caetano, B. C. O'Meara, J. M. Beaulieu, Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* **72**, 2308–2324 (2018).
19. J. B. Losos, T. R. Jackman, A. Larson, K. Queiroz, L. Rodríguez-Schettino, Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–2118 (1998).
20. D. L. Rabosky, R. E. Glor, Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22178–22183 (2010).
21. S. Poe *et al.*, A phylogenetic, biogeographic, and taxonomic study of all extant species of *Anolis* (Squamata; Iguanidae). *Syst. Biol.* **66**, 663–697 (2017).
22. E. D. Burress, M. M. Muñoz, Ecological opportunity from innovation, not islands, drove the anole lizard adaptive radiation. *Syst. Biol.* **71**, 93–104 (2021).
23. P. J. Green, "Trans-dimensional Markov chain Monte Carlo" in *Highly Structured Stochastic Systems*, P. J. Green, N. L. Hjort, S. Richardson, Eds. (Oxford University Press, 2003), pp. 179–198.
24. C. Roux *et al.*, Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* **14**, e2000234 (2016).
25. R. G. Reynolds *et al.*, Phylogeographic and phenotypic outcomes of brown anole colonization across the Caribbean provide insight into the beginning stages of an adaptive radiation. *J. Evol. Biol.* **33**, 468–494 (2020).
26. D. J. MacGuigan, A. J. Geneva, R. E. Glor, A genomic assessment of species boundaries and hybridization in a group of highly polymorphic anoles (*distichus* species complex). *Ecol. Evol.* **7**, 3657–3671 (2017).
27. J. B. Losos, D. Schluter, Analysis of an evolutionary species-area relationship. *Nature* **408**, 847–850 (2000).
28. M. Conway, B. J. Olsen, Contrasting drivers of diversification rates on islands and continents across three passerine families. *Proc. Biol. Sci.* **286**, 20191757 (2019).
29. J. P. Huelsenbeck, B. Rannala, Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* **57**, 1237–1247 (2003).
30. M. J. Landis, Biogeographic dating of speciation times using paleogeographically informed processes. *Syst. Biol.* **66**, 128–144 (2017).
31. M. J. Landis, W. A. Freyman, B. G. Baldwin, Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution* **72**, 2343–2359 (2018).
32. M. J. Landis *et al.*, Joint phylogenetic estimation of geographic movements and biome shifts during the global diversification of *Viburnum*. *Syst. Biol.* **70**, 67–85 (2021).
33. W. P. Maddison, P. E. Midford, S. P. Otto, Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* **56**, 701–710 (2007).
34. D. L. Rabosky, E. E. Goldberg, Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* **64**, 340–355 (2015).
35. M. Landis, E. J. Edwards, M. J. Donoghue, Modeling phylogenetic biome shifts on a planet with a past. *Syst. Biol.* **70**, 86–107 (2021).
36. B. R. Moore, M. J. Donoghue, Correlates of diversification in the plant clade Dipsacales: Geographic movement and evolutionary innovations. *Am. Nat.* **170**, S28–S55 (2007).
37. K. G. Ashton, M. C. Tracy, A. Queiroz, Is Bergmann's rule valid for mammals? *Am. Nat.* **156**, 390–415 (2000).
38. M. M. Muñoz *et al.*, Evolutionary stasis and lability in thermal physiology in a group of tropical lizards. *Proc. Biol. Sci.* **281**, 20132433 (2014).
39. D. L. Grossenbacher *et al.*, Self-compatibility is over-represented on islands. *New Phytol.* **215**, 469–478 (2017).
40. K. T. David, K. M. Halanych, Spatial proximity between polyploids across South American frog genera. *J. Biogeogr.* **48**, 991–1000 (2021).
41. I. Quintero, M. J. Landis, Interdependent phenotypic and biogeographic evolution driven by biotic interactions. *Syst. Biol.* **69**, 739–755 (2020).
42. J. Hackel, I. Sanmartín, Modelling the tempo and mode of lineage dispersal. *Trends Ecol. Evol.* **36**, 1102–1112 (2021).
43. J. Sukumaran, E. P. Economo, L. Lacey Knowles, Machine learning biogeographic processes from biotic patterns: A new trait-dependent dispersal and diversification model with model choice by simulation-trained discriminant analysis. *Syst. Biol.* **65**, 525–545 (2016).
44. S. Louca, M. W. Pennell, Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
45. W. C. Wimsatt, "False models as means to truer theories" in *Neutral Models in Biology*, M. Nitecki, A. Hoffman, Eds. (Oxford University Press, New York, NY, 1987), pp. 23–55.
46. M. R. Servedio *et al.*, Not just a theory–The utility of mathematical models in evolutionary biology. *PLoS Biol.* **12**, e1002017 (2014).
47. N. J. Matzke, Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. OSFPreprints [Preprint] (2021). https://osf.io/vqm7r (Accessed 20 September 2021).
48. R. Lande, Genetic variation and phenotypic evolution during allopatric speciation. *Am. Nat.* **116**, 463–479 (1980).
49. S. Wright, Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
50. S. Höhna *et al.*, RevBayes: Bayesian phylogenetic inference using graphical models and interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).
51. M. R. May, X. Meyer, TensorPhylo RevBayes plugin. https://bitbucket.org/mrmay/tensorphylo. Accessed 7 January 2022.
52. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).
53. M. J. Landis, I. Quintero, M. M. Muñoz, F. Zapata, M. J. Donoghue, FIG model analysis: Scripts and data. GitHub. https://github.com/mlandis/fig_model. Deposited 26 January 2022.