Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades

Nicholas J. Matzke[1,2,*]

[1]*Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA; and* [2]*National Institute for Mathematical and Biological Synthesis and Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-3410, USA*
[*]*Correspondence to be sent to: 1122 Volunteer Boulevard, Suite 106, University of Tennessee, Knoxville, TN 37996-3410, USA;*
E-mail: matzke@nimbios.org.

Abstract.—Founder-event speciation, where a rare jump dispersal event founds a new genetically isolated lineage, has long been considered crucial by many historical biogeographers, but its importance is disputed within the vicariance school. Probabilistic modeling of geographic range evolution creates the potential to test different biogeographical models against data using standard statistical model choice procedures, as long as multiple models are available. I re-implement the Dispersal–Extinction–Cladogenesis (DEC) model of LAGRANGE in the R package BioGeoBEARS, and modify it to create a new model, DEC+J, which adds founder-event speciation, the importance of which is governed by a new free parameter, *j*. The identifiability of DEC and DEC+J is tested on data sets simulated under a wide range of macroevolutionary models where geography evolves jointly with lineage birth/death events. The results confirm that DEC and DEC+J are identifiable even though these models ignore the fact that molecular phylogenies are missing many cladogenesis and extinction events. The simulations also indicate that DEC will have substantially increased errors in ancestral range estimation and parameter inference when the true model includes +J. DEC and DEC+J are compared on 13 empirical data sets drawn from studies of island clades. Likelihood-ratio tests indicate that all clades reject DEC, and AICc model weights show large to overwhelming support for DEC+J, for the first time verifying the importance of founder-event speciation in island clades via statistical model choice. Under DEC+J, ancestral nodes are usually estimated to have ranges occupying only one island, rather than the widespread ancestors often favored by DEC. These results indicate that the assumptions of historical biogeography models can have large impacts on inference and require testing and comparison with statistical methods. [BioGeoBEARS; cladogenesis; extinction; founder-event speciation; GeoSSE; historical biogeography; jump dispersal; LAGRANGE.]

Probabilistic modeling of geographic range evolution (Ree et al. 2005; Ree and Smith 2008; Ree and Sanmartín 2009) was a major advance in historical biogeography, making biogeographical problems accessible to model-based maximum-likelihood (ML) and Bayesian inference methodologies (reviewed by Ronquist and Sanmartín 2011; Sanmartín 2012). The most popular model, used in hundreds of publications, is Dispersal–Extinction–Cladogenesis (DEC), implemented in the software LAGRANGE in Python (Ree 2013) or C++ (Smith and Ree 2010). In DEC, geographic range is allowed to change across a phylogeny through several types of events. Along the branches of a phylogenetic tree (anagenetic evolution), the events allowed are "dispersal" (range expansion by adding an area) and "extinction" (range reduction through extirpation in an area), and these are treated as continuous-time Markov processes.

DEC's Cladogenesis Model

Under DEC, geographic range is also allowed to change at cladogenesis events. Several types of scenarios are allowed, conditional on the ancestral range. The first is sympatric speciation in an ancestor with a range size of one area. Here, the ancestral range (e.g., A) is simply copied to both daughter species (left: A, right: A). The second sympatric speciation scenario occurs when an ancestor is widespread, inhabiting multiple areas. Under DEC, inheritance of the widespread range by both daughters is not allowed. However,

sympatric speciation may still take place, where one daughter inherits the complete ancestral range, and the other daughter begins with a range of a single area, somewhere within the ancestral range. Here, scenario #2 is termed "sympatric-subset" speciation. The third cladogenesis scenario allowed in DEC is vicariance, where a widespread ancestral range is divided up between the two daughter species. In the LAGRANGE DEC model, the assumption that one of the daughters must have a range of size 1 is kept for the vicariance process. Thus, an ancestor with range ABCD may divide into daughters with unequal range sizes (e.g., A, BCD; C, ABD), but the daughter ranges are not allowed to split evenly (e.g., AB, CD), (AD, BC), etc. DEC's cladogenesis model descends with modification from the older, parsimony-based biogeography program DIVA (Dispersal–Vicariance Analysis; Ronquist 1997). DIVA allows all vicariance events, but does not include sympatric-subset speciation (Ronquist and Sanmartín 2011).

Most researchers who use LAGRANGE DEC have focused on the inference of ranges at ancestral nodes, and on the estimation of the parameters governing the rates of anagenetic dispersal and extinction. The cladogenesis model receives less attention, probably because it is a fixed, hard-coded model, and is therefore automatically part of any LAGRANGE analysis. Having a fixed model prohibits model comparison, giving users no capability to detect when the cladogenesis model is appropriate or inappropriate for the clade under study. DEC might well be the correct model, or at

least a wrong but useful model (Box and Draper 1987), but researchers are unable to check these assumptions unless additional plausible models are constructed and statistically compared with DEC.

FOUNDER-EVENT SPECIATION

One well-known form of speciation that is left out of standard DEC analyses is "founder-event speciation" (Paulay and Meyer 2002; Templeton 2008), sometimes termed speciation through long-distance dispersal (Heads 2012) or allopatric mode II speciation (Wiley 1981; Maguire and Stigall 2008; Lomolino et al. 2010). In founder-event speciation, a small number of individuals, sometimes even a single individual, take part in a rare, long-distance colonization event which founds a population which is instantly genetically isolated from the ancestral population. Founder-event speciation has received extensive theoretical attention (Mayr 1954; Gould and Eldredge 1972; Carson and Templeton 1984), and it remains a controversial question whether or not the population genetic founder effect and resulting genetic bottlenecks significantly contribute to shifting of adaptive peaks (Carson and Templeton 1984; Coyne and Orr 2004) or morphological evolution. However, to many biogeographers (but not all; see Heads 2012) it seems undeniable that founder-event speciation is an important mode of lineage splitting, and of moving taxa around the planet. Founder-event speciation is particularly likely to be important in oceanic island systems (Carlquist 1974; Paulay and Meyer 2002; de Queiroz 2005; Cowie and Holland 2006; Gillespie et al. 2012). If founder-event speciation is important for explaining certain biogeographical patterns, this would remain true whether or not founder-event speciation plays any major role in production of non-biogeographical macroevolutionary patterns.

Given the recognized importance of founder-event speciation in the literature, it is peculiar that the most popular inference methods in historical biogeography, DIVA, and LAGRANGE, fail to take it into account. It may be that some think that founder-event speciation is covered by the "dispersal" process of DEC and DIVA. However, as noted above, these concepts of dispersal are anagenetic range-expansion events (Ronquist and Sanmartín 2011), and are not identical with founder events simultaneous with lineage splitting.

Figure 1 shows examples of the types of cladogenesis events allowed (gray shading) and disallowed (white) in the DEC model. Some commentators (Ree and Sanmartín 2009; Kodandaramaiah 2010; Goldberg et al. 2011) have noted the absence of founder-event speciation in DIVA and DEC, and suggested that it should be taken into account. Here, I take their advice. I re-implement the DEC model in the R package BioGeoBEARS, and then add founder-event jump dispersal to create a "DEC+J" model. The two models are compared using standard likelihood and Bayesian model-choice methods, on both simulated and empirical data.
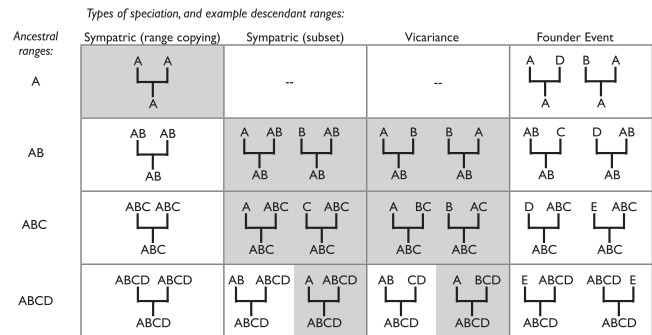

FIGURE 1. Example range-inheritance events at cladogenesis. The events allowed by LAGRANGE's DEC model are highlighted in gray. Each allowed event is fixed to have equal probability in the DEC model. DEC requires that, at cladogenesis, at least one daughter species must have a range size of 1, which excludes some events. Founder-event speciation (column 4) is also excluded from DEC, but included in DEC+J.

METHODS

Implementation of the DEC and DEC + J Models

DEC has two free parameters, the rate of range expansion ("dispersal", parameter d) and range contraction ("extinction", parameter e). These are combined into a rate matrix (Ree and Smith 2008) which is exponentiated to calculate the probabilities of each state at the bottom of a branch, as a function of the branch length and the probabilities of the states at the top of the branch. The rate matrix has m rows and m columns, where m represents the number of states in the state space, and each state is a possible geographic range. The number of possible states/geographic ranges rises at $m = 2^n$, where n equals the number of discrete areas in the analysis, assuming no constraints on the maximum size of ranges. Most phylogenetic methods use continuous-time models, where states evolve under the same model as they do on the branches. Under this condition, when two branches meet at an ancestral node, the probabilities at the two branch bottoms are multiplied to produce data likelihoods at the ancestral node. This is the basis of the Felsenstein pruning algorithm (Felsenstein 1981; 2004), and it represents the assumption that the ancestral state before cladogenesis (e.g., nucleotide G) is copied exactly to both descendants just after cladogenesis (left: G, right: G). In biogeography, this would correspond to assuming complete sympatric speciation, copying the exact ancestral range to both descendants, regardless of range size. However, in models where states evolve differently at cladogenesis events, the downpass probabilities must be combined via a cladogenesis model, which can be visualized as a rectangular transition matrix, where the $m-1$ rows represent the possible ancestral states (the null state, a range of 0 areas, is excluded here), and the $(m-1)^2$ columns represent every possible pair of (left, right) descendant states just after cladogenesis.

In the DEC cladogenesis model, conditional on a particular ancestral geographic range, a number of

cladogenesis events are possible (Fig. 1). DEC assigns a weight of 1 to each allowed event, and then divides by the sum of the weight of all events allowed conditional on a particular ancestral range, thus assigning equal probability to each allowed event. This procedure has the advantage of allowing rapid calculation, but the disadvantage of assuming a fixed cladogenesis model. Examples of cladogenesis models for a three-area analysis are shown in Table 1, for DEC and DEC+J, respectively. The DEC+J model is implemented by assigning a parameter, j, to specify the weight of each jump dispersal event in the cladogenesis matrix. To calculate the probabilities of a particular range inheritance event at cladogenesis, conditional on a particular ancestral range, the weight of each of the allowed range inheritance events is divided by the sum of the weights. A more detailed depiction of the DEC and DEC+J cladogenesis models for analyses with 2, 3, and 4 areas is available in an Excel file in Supplementary Data S1. Readers may vary the value of j and see the resulting effect on conditional probabilities.

In the BioGeoBEARS model parameters table, users may set parameters to be fixed, free, or deterministic functions of each other. The weights of individual sympatry, sympatric-subset, and vicariance events are controlled by y, s, and v. In this study's DEC analysis, $y = s = v = 1$, reproducing the LAGRANGE model. For DEC+J, $y = s = v = (3-j)/3$, to facilitate comparison with DEC. The limits on j are 0–3, and when $j = 0$, DEC+J reduces to DEC. Note that parameter identifiability must be considered when parameters are set. For example, the model $y = s = v = j = 1.0$ would be identical with the model $y = s = v = j = 0.5$. Due to the way the numbers and types of cladogenesis events are conditional on the ancestral range (e.g., Table 1), it is not feasible to parameterize the model as a simplex where y, s, v, and j are total probabilities of each class of events. However, a simplex (where y, s, v, and j sum to 1) could be used to specify the relative probabilities of individual events of each type, and this would be useful, for example, to allow application of a Dirichlet prior in a Bayesian analysis where all four parameters are free (Ronquist F., personal communication). This could be implemented through the BioGeoBEARS parameters table.

LAGRANGE allows users to manually specify a dispersal matrix, or a series of time-stratified dispersal matrices, which constitute multipliers to the d parameter when the anagenetic transition matrix is constructed. This dispersal matrix can consist of ones and zeros, indicating for which regions dispersal (range expansion) is allowed or disallowed; or it can consist of intermediate values, indicating the relative probability of dispersal between regions. When comparing DEC and DEC+J, it would be unfair to the DEC model to impose constraints on the dispersal (range-expansion) process, but not impose them on the founder-event process. Therefore, in the current implementation, dispersal matrices and other constraints are applied identically to d and j as probabilities are calculated—that is, if dispersal (range expansion) between two areas is prohibited, jump dispersal during cladogenesis is also prohibited. (It is not actually certain that the two processes should share the same constraints in all situations, but this is a topic for further research.)

Inference in BioGeoBEARS

BioGeoBEARS requires two accessory R packages: rexpokit (Matzke and Sidje 2013) and cladoRcpp (Matzke 2013b). Rexpokit uses the FORTRAN EXPOKIT library (Sidje 1998) to improve the speed of matrix exponentiation (Moler and Van Loan 2003), and cladoRcpp speeds cladogenesis calculations using C++ code linked to R with Rcpp (Eddelbuettel and Francois 2013). Given the phylogeny, geographic ranges at the tips, and anagenetic and cladogenetic models, and values for the d, e, and j parameters, the likelihood of the observed range data at the phylogeny tips can be calculated. By default, BioGeoBEARS uses the R package optimx (Nash and Varadhan 2011; 2012) to perform ML estimation of the free parameters, using the quasi-Newton method with box constraints (Byrd et al. 1995).

Once the ML estimate of the parameters has been made, BioGeoBEARS estimates the probability of each possible ancestral state at each ancestral node. This is done for all nodes by conducting a downpass and an uppass through the tree (Felsenstein 2004). The resulting ancestral state probabilities represent the ancestral states estimated under the globally optimum model. LAGRANGE, in contrast, does a global ML search, but then estimates ancestral states by iteratively constraining each state to be true at each node, and then performing a separate ML search under each constraint (Ree and Smith 2008). This is ancestral state estimation under local optimization (Mooers and Schluter 1999; Felsenstein 2004; Mooers 2004). Although local optimization is available in BioGeoBEARS, by default, BioGeoBEARS calculates ancestral states under the globally optimum model, both for reasons of speed and because Mooers (2004) found that ancestral states estimated under the global ML model may be slightly more accurate.

Validation

BioGeoBEARS was validated against both Python LAGRANGE (versions 20120508 and 20130526) and C++ LAGRANGE (version 0.20, commit 33). Validation was conducted in two ways. First, for each test data set, analyses were run in both versions of LAGRANGE. The ML estimates of d and e, and the resulting log-likelihood of the data, were recorded. Then, these d and e estimates were input into BioGeoBEARS, along with the same tree and tip range data used in the LAGRANGE runs, and the resulting log-likelihood was checked against the LAGRANGE log-likelihood. Second, BioGeoBEARS was set to let d and e vary as free parameters, and used to obtain the ML estimates. These were compared with the LAGRANGE estimates.

TABLE 1. Comparison of the cladogenesis models for the DEC model (as implemented in LAGRANGE or BioGeoBEARS) and the DEC+J model (as implemented in BioGeoBEARS)

A. Parameterization of LAGRANGE'S DEC cladogenesis model for three areas. DEC assumes equal event weights ($y=v=s$), with each transition probability equal to the weight divided by the sum of the weights.

Ranges of descendant pairs

B. Parameterization of the DEC+J cladogenesis model for three areas. Transition probabilities calculated from the weights as above. Under DEC+J, $y=s=v=(3-j)/3$. In both models, blank cells have 0 probability.

Ranges of descendant pairs

Notes: These transition tables assume a study system of three areas and a maximum range size of 3 areas; analyses with a different number of areas will have different transition matrices (see Supplementary Material and the BioGeoBEARS package). *Abbreviations: y,* sympatric range-copying speciation; *s,* sympatric-subset speciation; *v,* vicariance; *j,* jump dispersal or founder-event speciation. In LAGRANGE DEC, the cladogenesis model is fixed, and all allowed events have weight 1. Each allowed cladogenesis event has a probability of 1/(number of cladogenesis events possible from that ancestor). That is, for an ancestor of range size 3 (ABC), there are 12 possible cladogenesis events under the model, thus, each is assigned a probability of 1/12. In BioGeoBEARS, the per-event weights, *y, s, v,* and *j,* can all be fixed, or made into free parameters, according to user preference. The right side of each table shows counts of each type of event under DEC and DEC+J, and the total number of possible events, conditional on each ancestral range.

Study System: Hawaiian *Psychotria*

The focal clade used as the main test example was Hawaiian *Psychotria*, using the same phylogeny that Ree and Smith (2008) originally simplified from Nepokroeff et al. (2003). Ree and Smith's constraints models were also reproduced: M0 was an unconstrained analysis; M1 was unconstrained, except that the maximum range size was set to two areas; M2 added the constraint that dispersal could only happen from west-to-east. Finally, their "stratified" model (here called M3) only allowed dispersal between islands when the islands are known to have emerged (Clague 1996; Nepokroeff et al. 2003; Ree and Smith 2008). Each of these constrained models was run in all versions of LAGRANGE as well as BioGeoBEARS-DEC for validation, as well as BioGeoBEARS DEC+J.

Bayesian Analysis

DEC and DEC+J were also compared within a Bayesian framework on the *Psychotria* data set using the R package LaplacesDemon (Statisticat 2013b, available at http://www.bayesian-inference.com/software, last accessed August 7, 2014). Uniform priors applied to the d, e, and j parameters, with the bounds set to match the limits used in ML search (0–5 for d and e, and 0–3 for j), and likelihoods calculated as during ML analysis. First, random-walk Metropolis MCMC (Link and Barker 2009; Statisticat 2013a) was used to compare the models, with DEC and DEC+J models run separately, each for 10,000 generations. Inspection of trace and autocorrelation plots indicated that this run length was more than sufficient to accomplish burnin (which occurred within 200 generations), establish stationarity, and achieve sufficient sampling to characterize the posterior. The marginal log-likelihood of each model was estimated using the nonparametric self-normalized importance sampling algorithm (LaplacesDemon LML function, NSIS option) of Escoto (2011). The Bayes factor was calculated from the ratio of the approximate marginal likelihoods. Second, reversible-jump MCMC (Green 1995) was used to sample both DEC and DEC+J according to their posterior probabilities. Achieving adequate sampling of both models required a prior probability of 0.000001 on DEC+J to balance out its log-likelihood advantage (~14 log-likelihood units). The analysis was run for 50,000 generations and assessed as described above. Here, the Bayes factor was calculated from the posterior sampling frequencies of the two models as the ratio of posterior to prior model probabilities.

Simulations for *Psychotria* M0 Focal Example

Biogeographical histories were simulated on the *Psychotria* phylogeny, using the ML-estimated parameters from the DEC and DEC+J runs on the M0 model. 1000 simulations from each model were conducted, and the simulated histories were saved. Then, on each of the 2000 simulations, ML inference was conducted under DEC and DEC+J, using the same default settings described above for the original inference. The difference in log-likelihoods produced by each model was calculated, and a likelihood-ratio test (LRT) with one degree of freedom (Burnham and Anderson 2002) was conducted on each pair of inferences to test the null hypothesis that the two models confer the same likelihood on the simulated data. In the 1000 simulations where the DEC model is the true one, the null hypothesis should be rejected (producing a false positive) no more frequently than would be expected by chance with a *P*-value cutoff of 0.05. The false-negative rate (failure to reject the null hypothesis, when the null hypothesis is false) was also measured by repeating the inference/LRT procedure on the 1000 data sets simulated under the DEC+J model. The accuracy of ancestral state estimates was measured for each of the 4000 simulation/inference pairs by calculating the fraction of the nodes for which the most probable state under ancestral range estimation matched the true simulated range.

Simulations to Test Effects of Varying Parameters, Missing Speciation and Extinction Events, and State-Dependence in Speciation and Extinction

The DEC+J model and the simulation tests above implicitly assume that the observed phylogeny is the true phylogeny. This is unlikely to be the case in empirical data sets, where many species will be extinct or unsampled, and thus many cladogenesis events will be missing from the observed tree. Missing data can cause difficulties for all forms of ancestral state inference, but biogeographical models might be particularly vulnerable, considering their emphasis on range-changing processes at cladogenesis. This potential criticism applies to DEC and DIVA as well as DEC+J. The GeoSSE (Goldberg et al. 2011) and ClaSSE (Goldberg and Igić 2012) models naturally take into account extinction as well as the potential dependence of speciation and extinction rates on geographic range. However, the number of free parameters in such models expands rapidly with the number of possible geographic ranges, which may cause difficulties in applying these methods in typical biogeography analyses (single clade, <200 species, >4 areas).

The potential impact of these factors was tested with simulations where geographic range evolved simultaneously with the phylogeny under a range of parameter combinations. The ML estimates of the speciation rate (λ) and extinction rate (μ) for the *Psychotria* tree are 0.33 and 0 events per million years (estimated with the "birthdeath" function of the R package geiger; Harmon et al. 2008), so, for rough comparability, the base speciation rate was set to $\lambda_{base} = 0.3$ in all simulations, and μ_{base} was set to 0, 0.1, or 0.3. Combinations of these parameters produce Yule

(pure-birth) trees, or birth–death (BD) trees, and these models assume that λ and μ are independent of geographic range. A plausible biogeographic model for state-dependence of λ and μ on geographic range is that species living in multiple areas have increased rates of speciation and decreased rates of extinction. Therefore, for each of the three parameter combinations above, corresponding state-dependence in speciation and extinction (SSE) simulations were constructed where $\lambda_r = \lambda_{\text{base}} \cdot r^\alpha$ and $\mu_r = \mu_{\text{base}} \cdot r^\varpi$, where r equals range size (number of areas occupied) and λ_r and μ_r represent the speciation and extinction rates given r. The effect of range size on speciation and extinction is controlled by the exponents α and ϖ. In the SSE simulations performed here, $\alpha = 1$, and $\varpi = -1$. This introduces dramatic range size-dependent heterogeneity in rate, for example, when $\lambda_{\text{base}} = \mu_{\text{base}} = 0.3$, a lineage inhabiting $r = 4$ areas will have $\lambda_r = 1.2$ and $\mu_r = 0.075$, a 16-fold difference. Setting $\alpha = \varpi = 0$ reduces the SSE models to BD models.

Note that an alternative way of conceptualizing the joint process of speciation/extinction and range evolution would be to have a single Markov rate for dispersal of all types, and then assign a probability for each dispersal event being a cladogenetic jump dispersal event, rather than a range-expansion event. This would be equivalent to the present simulation model, given appropriate parameterization, as can be seen by the fact both conceptualizations are simplifications of a generic model where a unique Markov rate parameter is assigned to each possible pair of cladogenetic and range-change events.

For each of the six macroevolutionary models described above, geographic range simultaneously evolved for a four-area, unconstrained (M0) system under several values of d (0, 0.03, 0.15), e (0, 0.03, 0.15), and j (0, 0.02, 0.1, 0.3), with all but the highest values based on common estimates made in empirical data sets. All combinations of these parameters were used, except combinations where $e > d$, producing a total of 138 combinations of d, e, j, λ, μ, α, and ϖ. When range contraction occurred on a lineage occupying a single area, this was treated as another extinction event, meaning that in these simulations, range contraction was an additional source of extinction events, independent of the values of μ and ϖ. 100 simulations were run under each parameter combination, and a simulation was stopped once a tree had 50 living tips. If a tree died out or failed to reach 50 tips within 100 myr, the simulation was attempted again up to 100 times. If 100 tries consistently failed, the parameter combination was abandoned, which occurred in only four cases, BD simulations where d, e, λ, and μ all had their maximum values (a peculiarity of DEC-type models is that the null range is an absorbing state, so that the steady-state equilibrium is a range of zero areas; these BD trees have long branches, and the high value of e means that the geographic evolution process drives all lineages extinct). Simulations were conducted assuming the root state was a single area, and repeated assuming the root state was all areas, for a total of 26,800 simulations. The exact history of events and ancestral states was saved for each simulation, and true versus observed tree ages and event counts were plotted for each parameter combination to confirm expected behavior. Each simulated tree was pruned of all fossil tips and subjected to inference under DEC and DEC+J.

Multiclade Analysis

In order to assess the importance of the founder-event process in the historical biogeography of island clades, ML inference with the DEC and DEC+J models was conducted on a sample of 13 clades exclusively or mostly found on islands. Data sets were gathered from recent published analyses in the historical biogeography literature. The criteria for inclusion were as follows: the clade had to be primarily island-based, the published study had to include the presence/absence information for each tip in the tree, and the published phylogeny had to be either dated or contain molecular branch lengths and sufficient dating information to allow estimation of an ultrametric tree using r8s (Sanderson 2003). Tree files were obtained where possible, or digitized from the published figure with TreeRogue (Matzke 2013c) using methods described in Supplementary Text S1.

Island clades were gathered opportunistically by searching the literature, especially studies that examined Hawaiian clades or that used LAGRANGE. Some older studies were included, despite weaker dating and biogeography methods, because they covered classic island radiations (e.g., the Hawaiian silversword alliance). The full accounting of data sets and their sources is given in Table 2 and Supplementary Text.

To achieve maximum comparability and broadness of results, if the source study in question made use of a dispersal matrix or other constraints, these were converted into the necessary input files and settings for BioGeoBEARS. Wherever a constrained analysis was run, an unconstrained run was also implemented for comparison. For Hawaiian groups, depending on their geography and the quality of the dating, it was usually possible to implement the M0, M1, M2, and M3 (time-stratified) constraints models of Ree and Smith (2008). In addition to these, a refinement of time-stratified dispersal was also implemented, wherein a time-stratified areas-allowed matrix was added to the analysis. For clades older than Kauai, an additional time-stratified analysis was constructed, wherein an ancient area, "Z", existed before the emergence of Kauai. For clades thought to have immigrated to Hawaii from a continent, "Z" remains extant throughout the analysis. This procedure then allows two possible constraints models—models with dispersal from Z being equally probable as any other dispersal, or the more realistic model of extremely low-probability dispersal from a continent. Alternatively, for clades thought to descend from an older Hawaiian high island, Z represents that older island, which disappears soon after the emergence of Kauai, as there was a slowdown in island formation

TABLE 2. Study clades, data sources, and constraints settings (maximum range size, dispersal matrices, time-stratification, areas-allowed matrices, etc.)

Region	Clade	Data set source	Modifications	Number of taxa	Tree age	Abbreviation	Number of areas	Max number of areas	Constraints model
Hawaii	*Psychotria*	Ree and Smith (2008) phylo & geog	None	19	5.2	M0	4	4	Unconstrained
				19	5.2	M1	4	2	2 areas max
				19	5.2	M2	4	2	2 areas + eastwards dispersal only
				19	5.2	M2b	4	4	4 areas max, eastwards-only dispersal
				19	5.2	M3	4	4	Time-stratified dispersal matrix
				19	5.2	M3b	4	4	Time-stratified areas allowed and dispersal matrix
Hawaii	Hawaiian *Drosophila*	Kambysellis et al. (1995) phylo & geog	None; digitized from paper; time-scaled with r8s, LF option	42	20	M0	5	5	Unconstrained
				42	20	M3 areas	5	5	Stratified dispersal and areas
				42	20	M3 areas	5	5	Stratified dispersal and areas, w. Z
Hawaii	Hawaiian *Scaptomyza*	Lapoint et al. (2013) phylo & geog	None; digitized by NJM	49	12.4	M0	4	4	Unconstrained
				49	12.4	M1	4	2	2 areas max
				49	12.4	M1a	4	2	2 areas max; adjacent areas only
				49	12.4	M2	4	2	2 areas max; eastwards dispersal only
				49	12.4	M3b_stratified_w_areas	4	4	Stratified dispersal & areas
				49	12.4	M0	5	5	Unconstrained, w. Z
				49	12.4	M1	5	2	2 areas max, w. Z
				49	12.4	M1a	5	2	2 areas max; adjacent areas only, w. Z
				49	12.4	M2	5	2	2 areas max; eastwards dispersal only, w. Z
				49	12.4	M3b_stratified_w_areas	5	5	Stratified dispersal & areas, w. Z
				49	12.4	M3c_stratified_w_areas	5	5	Stratified dispersal & areas, w. Z, low. prob.
Hawaii	Hawaiian Honeycreepers	Lerner et al. (2011) phylo & geog	None; digitized by RR	19	5.78	M0	4	4	Unconstrained
				19	5.78	M3_strat	4	4	Stratified as with *Psychotria* (Ree and Smith 2008)
				19	5.78	M3a_strat	4	4	Stratified dispersal + areas
Hawaii	*Megalagrion*	Jordan et al. (2003) phylo & geog	None; digitized by NJM	34	9.6	M0	4	4	Unconstrained
				34	9.6	M1_2areas	4	2	2 areas max
				34	9.6	M1a_2areas	4	2	2 areas max; adjacent areas only
				34	9.6	M2_2areasEastOnly	4	2	2 areas max; eastwards dispersal only
				34	9.6	M3b_stratified_w_areas	4	4	Stratified dispersal & areas
				34	9.6	M0	5	5	Unconstrained
				34	9.6	M1_2areas	5	2	2 areas max

(continued)

TABLE 2. Continued

Region	Clade	Data set source	Modifications	Number of taxa	Tree age	Abbreviation	Number of areas	Max number of areas	Constraints model
				34	9.6	M1a_2areas	5	2	2 areas max; adjacent areas only
				34	9.6	M2_2areasEastOnly	5	2	2 areas max; eastwards dispersal only
				34	9.6	M3b_stratified_w_areas	5	5	Stratified dispersal & areas
				34	9.6	M3c_stratified_w_areas	5	5	Stratified dispersal & areas; low prob Z
Hawaii	Nesophrosyne	Bennett and O'Grady (2013) phylo & geog	None; digitized by NJM; left out area #7, *Orosius* (from Australia)	198	3.2	M1 (UM-2)	6	6	Unconstrained
				198	3.2	M1 (UM-2)	6	2	2 areas max
				198	3.2	M3_6max (TS-Ua)	6	6	Unconstrained, but time-stratified dispersal (nonzero) and areas constraints
				198	3.2	M3_2max (TS-2)	6	2	2 areas max; time-stratified dispersal (nonzero) and areas constraints
				198	3.2	M3_6max (TS-U)	6	6	Unconstrained, but time-stratified dispersal
				198	3.2	M3_2max (TS-2)	6	2	2 areas max; time-stratified dispersal and areas constraints
				198	3.2	M3_6max (TS-U)	6	6	Unconstrained, but time-stratified dispersal and areas constraints
Hawaii	*Orsonvelles*	Hormiga et al. (2003) phylo & geog	None; digitized by NJM	12	4.35	M0	5	5	Unconstrained
Hawaii	*Plantago*	Dunbar-Co (2008) phylo & geog	None; digitized by SS	16	5.1	M0	5	5	Unconstrained
				16	5.1	M3_strat	5	5	Stratified as with *Psychotria* (Ree and Smith 2008)
Hawaii	Silversword alliance	Baldwin and Sanderson (1998) phylo, Gillespie and Baldwin (2010) geog	Digitized by NJM; some polytomies arbitrarily resolved	16	5.1	M3_strat	5	5	Stratified dispersal and areas
				29	6.84	M0	5	5	Unconstrained
				29	6.84	M3_strat	5	5	Stratified as with *Psychotria* (Ree and Smith 2008), with areas
Caribbean	*Anolis*	Nicholson et al. (2012) phylo	Geography data and date priors from MCC tree regenerated from BEAST analysis	190	98.9	M0	12	3	Unconstrained
Pacific	*Cyrtandra*	Clark et al. (2008) phylo & geog	None; digitized by NJM	59	41.5	M0	7	7	Unconstrained
Galapagos	*Microlophus*	Benavides et al. (2009) phylo & geog	None; digitized by NJM	59	41.5	M0	7	4	4 areas max
				9	3.75	M1	10	2	2 areas max; otherwise unconstrained
SE Asia	*Vireya*	Webb and Ree (2012) phylo & geog	None; used 11 Ma root	65	11	M1	20	2	2 areas, no other constraints
				65	11	M1+cc	20	2	2 areas + constrained connectivity

Notes: In total, 13 clades and 53 constraints models were run.

before Kauai (Clague 1996). Note that even though no extant tips in these phylogenies inhabit area Z, BioGeoBEARS will still infer it as the ancestral area when all of the extant high islands are submerged. The constraints models are summarized in Table 2, and the details are found in the input files, which are available in Dryad. DEC and DEC+J analyses were run on each data set, and on each available dispersal model for data sets that included these. The DEC and DEC+J models were compared with the likelihood-ratio test, AIC, and AICc (Burnham and Anderson 2002) using custom functions implemented in BioGeoBEARS (Matzke 2013a).

RESULTS

Validation

BioGeoBEARS's DEC model performed well in validation tests against LAGRANGE on the example *Psychotria* data set for constraints M0–M3 (Table 3). In almost all cases, BioGeoBEARS replicated the log-likelihoods and parameter estimates returned by LAGRANGE for the same input data. This is strong evidence that the fundamental logic of LAGRANGE has been understood and incorporated into BioGeoBEARS. The exception was time-stratified analyses, where BioGeoBEARS agrees with 2013 Python LAGRANGE but not C++ LAGRANGE, indicating a difference in the stratified calculations or optimization algorithm of the latter.

Hawaiian Psychotria: DEC versus DEC +J

Comparison of the performance of DEC and DEC+J models on Hawaiian *Psychotria* showed dramatic results on all constraints models (Tables 4 and 5), with the data likelihood jumping by 11–15 log-likelihood units. Approximately 2 log-likelihood units constitute a statistically significant increase when nested models differ by one parameter, and likelihood-ratio tests show that DEC+J is strongly favored, with the null model rejected at the $P < 0.0001$ cut-off level in all cases. AIC weights give a sense of the relative support that the data lend to DEC and DEC+J. The ratio of the weights in favor of DEC+J ranges from 1800 (for the highly constrained, time-stratified M3 model) to 10^6 (for the M2 model where dispersal is eastwards-only). As the data set is small (only 19 OTUs in the *Psychotria* tree), AICc was also calculated, and the results were substantively the same. In sum, from a frequentist perspective, the *Psychotria* data set decisively rejects the equivalence of DEC and DEC+J. From a model selection perspective, DEC+J earns virtually all of the model weight.

Consideration of parameter inference under DEC and DEC+J (Table 4) is also revealing. For *Psychotria*, under DEC+J, the *j* parameter is always positive, and the *d* and *e* parameters are inferred to be effectively zero (hitting the lower bound of the optimx search bounds). This is an indication that the "D" and "E" processes of the DEC

model are completely unnecessary for explaining the biogeography of Hawaiian *Psychotria*. Instead, the data are explained with much higher probability by a series of founder events. The only partial exception to this is the M3b model. In this model, which has a time-stratified areas-allowed matrix in addition to a time-stratified dispersal constraints matrix, *d* and *e* are positive even under DEC+J, and the likelihoods are closer.

Figure 2 depicts the ancestral geographic ranges estimated under each model. It can be seen that the DEC model reconstructs several ancestral nodes with widespread ranges (a well-known bias of the model; Clark et al. 2008; Ree and Smith 2008; Buerki et al. 2011), which then break up through vicariance to produce daughter lineages with narrow ranges. Under DEC+J, this phenomenon largely disappears. Ancestral states are usually reconstructed with less uncertainty under the DEC+J model than the DEC model (Fig. 2, right). This should not be surprising, as the fact that the probabilities are less spread out over the possible ancestral ranges is directly related to the higher likelihood of the data. Note that differences in uncertainty can be masked in plots that show only the single most-probable state, so plots or tables showing uncertainty should always be consulted. Note also that plots of the most-probable ancestral state (and of the states at the "corners," which are used to represent the geographic range immediately after a cladogenesis event) are not the same thing as the single best joint history (Felsenstein 2004), so the highest probability ancestral ranges at each individual node cannot always be read naively as a joint history.

The results of Bayesian inference are given in Supplementary Text. Bayes factors calculated via several methods are dominated by the likelihood difference between the models and thus are congruent with the ML analysis. Trace plots and posterior distributions of parameters under the two models are shown in Supplementary Figure 1.1–1.6.

Simulations on Psychotria M0 Focal Example

Supplementary Figure S2 shows the log-likelihood advantage of the DEC+J model over the DEC model, when the true model is DEC or DEC+J. When the true model is DEC, this advantage is very slight. The LRT rejected DEC in favor of DEC+J only 43/1000 times, a rate approximately equal to the P-value cut-off ($P<0.05$). This indicates that the LRT has the desired frequentist properties when comparing DEC to DEC+J. However, in the reverse situation, the log-likelihood advantage of DEC+J is massive, indicating that model selection easily detects DEC+J when DEC+J generated the data.

For each of the 4000 simulation/inference pairs, the fraction of nodes for which the true ancestral range matched the inferred most-probable ancestral range is shown in Supplementary Figure S3. The accuracies were ordered as follows: DEC+J simulations, DEC inference: 0.57; DEC simulations, DEC inference: 0.78; DEC simulations, DEC+J inference: 0.83; DEC+J simulations,

TABLE 3. Validation of the BioGeoBEARS implementation of the two-parameter DEC model against the various versions of LAGRANGE

Program	Taxon	Model	Description	Ln L	d	e	Data set
Python LAGRANGE (2012)	*Psychotria*	M0	Unconstrained	−34.54	0.03505	0.02831	Ree and Smith (2008) phylo & geog
Python LAGRANGE (2013)	*Psychotria*	M0		−34.54	0.03505	0.02831	
C++ LAGRANGE	*Psychotria*	M0		−34.542	0.035017	0.0282904	
BioGeoBEARS DEC model	*Psychotria*	M0		−34.5	0.0351	0.0285	
Python LAGRANGE (2012)	*Psychotria*	M0 strat	Unconstrained, with null stratification	−35.36	0.03818	0.04079	Ree and Smith (2008) phylo & geog
Python LAGRANGE (2013)	*Psychotria*	M0 strat		−34.54	0.03505	0.02831	
C++ LAGRANGE	*Psychotria*	M0 strat		−34.542	0.035017	0.0282904	
BioGeoBEARS DEC model	*Psychotria*	M0 strat		−34.5	0.03506479	0.02833348	
Python LAGRANGE (2012)	*Psychotria*	M1	2 areas max	−35.21	0.04271	0.04305	Ree and Smith (2008) phylo & geog
Python LAGRANGE (2013)	*Psychotria*	M1		−35.21	0.04271	0.04305	
C++ LAGRANGE	*Psychotria*	M1		−35.2111	0.0427272	0.0429474	
BioGeoBEARS DEC model	*Psychotria*	M1		−35.2	0.04270405	0.0427981	
Python LAGRANGE (2012)	*Psychotria*	M2	2 areas + eastwards dispersal only	−31.92	0.1228	0.00000055	Ree and Smith (2008) phylo & geog
Python LAGRANGE (2013)	*Psychotria*	M2		−31.92	0.1228	0.00000055	
C++ LAGRANGE	*Psychotria*	M2		−31.92	0.122973	4.7416E−07	
BioGeoBEARS DEC model	*Psychotria*	M2		−31.9	0.1227461	1.00E−15	
Python LAGRANGE (2012)	*Psychotria*	M3	Time-stratified dispersal matrix	−39.8	0.03687	7.233E−08	Ree and Smith (2008) phylo & geog
Python LAGRANGE (2013)	*Psychotria*	M3		−39.83	0.03645	4.495E−08	
C++ LAGRANGE	*Psychotria*	M3		−36.6318	0.0342543	1.1488E−07	
BioGeoBEARS DEC model	*Psychotria*	M3		−39.8	0.03643885	1.00E−15	
Python LAGRANGE (2012)	*Cyrtandra*	M0	Unconstrained	−46.43	0.001251	4.285E−09	Clark et al. (2008, 2009) phylo & geog
Python LAGRANGE (2013)	*Cyrtandra*	M0		−46.43	0.001251	4.285E−09	
C++ LAGRANGE	*Cyrtandra*	M0		−46.4624	0.00110037	2.9021E−05	
BioGeoBEARS DEC model	*Cyrtandra*	M0		−46.42506	0.00125161	1.00E−15	
C++ LAGRANGE	*Cyrtandra*	M1	4 areas max	−47.2735	0.00144745	3.0655E−05	Clark et al. (2008, 2009)
BioGeoBEARS DEC model	*Cyrtandra*	M1		−47.3	0.00153	1.00E−15	
Python LAGRANGE (2012)	*Lonicera*	M0	Unconstrained	−125	0.005812	0.001049	Smith and Donoghue (2010) phylo & geog
Python LAGRANGE (2013)	*Lonicera*	M0		−125	0.005812	0.001049	
C++ LAGRANGE	*Lonicera*	M0		−124.977	0.0058186	0.00111119	
BioGeoBEARS DEC model	*Lonicera*	M0		−125	0.0058159	0.00104781	

Notes: The full precision output by each program is included to allow detailed comparison.

TABLE 4.   Comparison of DEC and DEC+J on Hawaiian *Psychotria*, for various models

Model	Constraints	Max number of areas	Ln L	Num	d	e	j	alt Ln L	null Ln L	D	P-value
DEC	M0	4	−34.5	2	0.035	0.029	0				
DEC+J	M0	4	−20.9	3	1E−15	1E−15	0.114	−20.95	−34.54	27.2	2E−07
DEC	M1	2	−35.2	2	0.043	0.043	0				
DEC+J	M1	2	−21.0	3	1E−15	1E−15	0.114	−20.96	−35.21	28.5	9E−08
DEC	M2	2	−31.9	2	0.123	1E−15	0				
DEC+J	M2	2	−16.7	3	1E−15	1E−15	0.285	−16.67	−31.92	30.5	3E−08
DEC	M2b	4	−31.4	2	0.115	0.009	0				
DEC+J	M2b	4	−16.7	3	1E−15	1E−15	0.285	−16.67	−31.41	29.5	6E−08
DEC	M3	4	−39.8	2	0.036	1E−15	0				
DEC+J	M3	4	−28.3	3	1E−15	3E−09	0.187	−28.32	−39.83	23.0	2E−06
DEC	M3b	4	−41.7	2	0.075	0.075	0				
DEC+J	M3b	4	−32.8	3	0.024	0.041	0.194	−32.78	−41.71	17.9	2E−05

TABLE 5.   AIC and AICc results for Hawaiian *Psychotria*, as well as the resulting model weights, and the ratio of the weights (DEC+J probability:DEC probability)

Model	Constraints	Max number of areas	Ln L	AIC1	AIC2	wt1	wt2	Ratio	AICc1	AICc2	wt1	wt2	ratio
DEC	M0	4	−34.5										
DEC+J	M0	4	−20.9	47.9	73.1	1	3.4E−06	294,895	49.5	73.8	1	5.2E−06	191,951
DEC	M1	2	−35.2										
DEC+J	M1	2	−21.0	47.9	74.4	1	1.8E−06	566,941	49.5	75.2	1	2.7E−06	371,386
DEC	M2	2	−31.9										
DEC+J	M2	2	−16.7	39.4	67.8	1	6.5E−07	1,537,750	40.9	68.6	1	9.9E−07	1,009,534
DEC	M2b	4	−31.4										
DEC+J	M2b	4	−16.7	39.4	66.8	1	1.1E−06	924,075	40.9	67.6	1	1.6E−06	606,220
DEC	M3	4	−39.8										
DEC+J	M3	4	−28.3	62.6	83.7	1	2.7E−05	36,882	64.2	84.4	1	4.2E−05	23979.6
DEC	M3b	4	−41.7										
DEC+J	M3b	4	−32.8	71.6	87.4	1	0.0004	2771	73.2	88.2	1	0.00055	1816.11

Notes: Models M0–M3 as described in text. M2b is M2 (eastwards-only dispersal) but with maximum range size set to 4 rather than 2. M3b is the M3 model (time-stratified with a changing dispersal matrix), but with the addition of a time-stratified "areas allowed" matrix, restricting not only dispersal to but also occupancy of submerged islands.

DEC+J inference: 0.87. The means were all significantly different from each other, using the paired-sample *t*-test for inferences made on the same simulations, and the unpaired Welch's *t*-test for comparing means between different simulation runs (two-tailed tests, all *P*-values <2e−15). Ancestral state accuracy for the *Psychotria* M0 example under DEC and DEC+J as a function of node age is shown in Supplementary Figure S4, and accuracy of parameter inference under the two models is shown in Supplementary Figure S5. These results are discussed in Supplementary Text as they are essentially special cases of the more comprehensive simulations discussed below.

Simulations Under Yule, BD, and SSE Processes

Representative results of DEC and DEC+J inference on simulations under different macroevolutionary models are shown in Figures 3 and 4. The results of all simulations are shown in Supplementary Text. The Yule simulations (Fig. 3) match the assumption that the observed tree is the true tree, and here DEC+J parameter inference performs well, except when d and e have values (0.15) that are very high relative to λ (0.3). DEC inference, however, suffers significantly when DEC+J is the true model, with d estimates suffering increased upward bias as j increases. The accuracy of ancestral range inference under DEC similarly suffers when DEC+J is the true model, again as a function of j. Accuracy of ancestral range inference decreases for both models when d is high. Taking ΔAICc as an indicator of the ability of the data to choose between models, DEC and DEC+J are very easily distinguished when d and e are zero or moderate. When d and e are high, model distinguishability decreases, although it is still satisfactory for moderate and high values of j.

Comparing the Yule simulations to the most extreme SSE simulations (Fig. 4), the most prominent finding is that the overall picture remains highly similar. Parameter inference and ancestral range estimation both suffer somewhat on SSE simulations in comparison to the Yule simulations, but the effects are moderate. DEC and DEC+J remain distinguishable under realistic parameter values, and inference of d and j do not greatly interfere with each other except at high d.
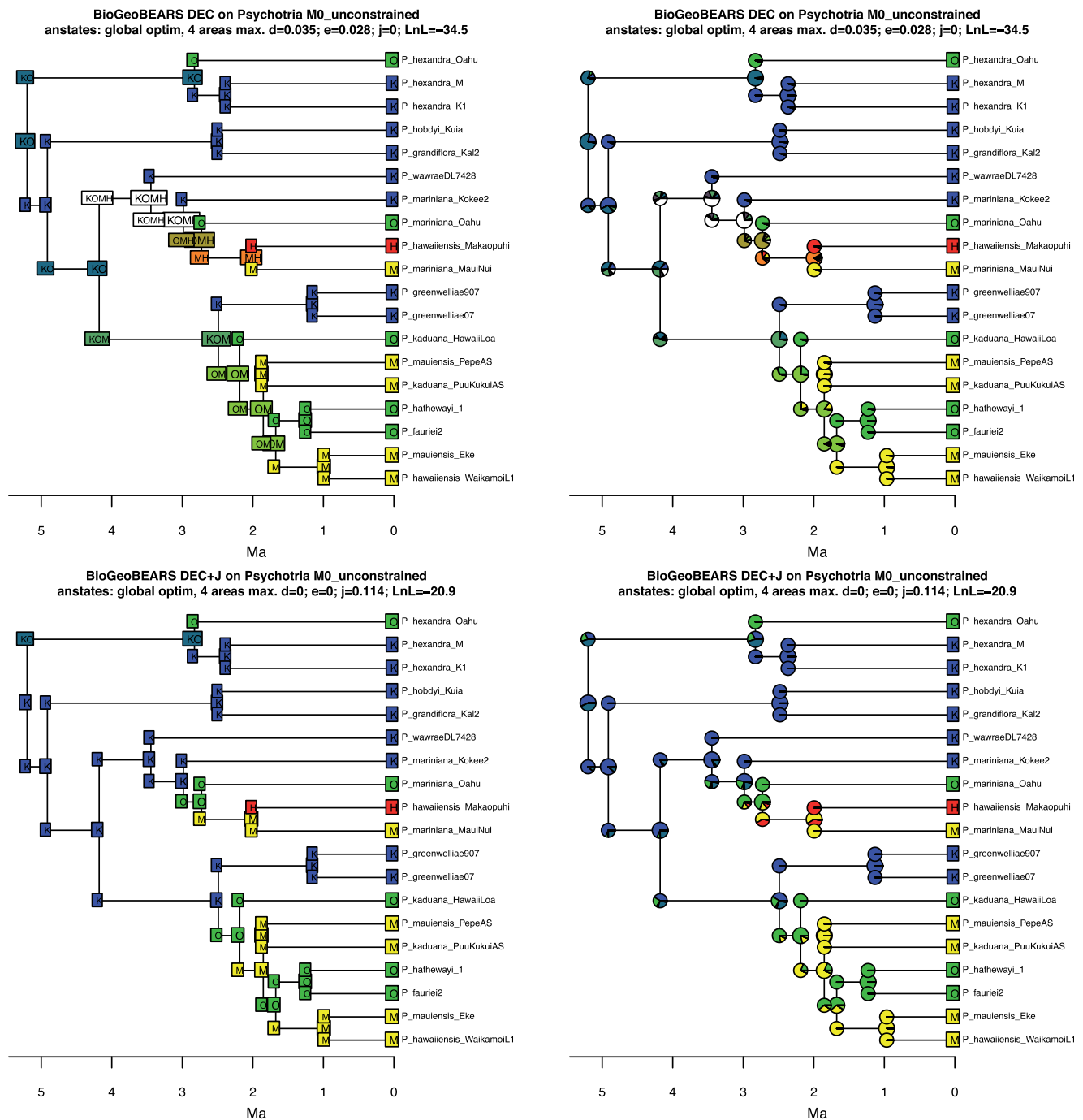
FIGURE 2.   Global ancestral state estimates under M0 (unconstrained) for the DEC and DEC+J models run on Hawaiian *Psychotria*. Left: the single most-probable ancestral range. Right: pie charts show the relative probability of all 16 states/possible geographic ranges. The corner positions are used to represent the geographic range immediately after a cladogenesis event. Key: Kauai, K, blue; O, Oahu, yellow; M, Maui-Nui, green; H, Hawaii Big Island, red. Ranges that are combinations of these four areas have a color that is a mix of the component area colors. A range of all areas (KOMH) is white.

Very similar results were produced by all of the other simulations (Supplementary Figs. S6 and S7), including the BD model where $\lambda = \mu = 0.3$, which is the model that produced the longest branches and the most extinct lineages (shown in Supplementary Fig. S8, along with summary statistics for all parameter combinations). Across the macroevolutionary simulations considered

here, it appears that inappropriately leaving founder-event speciation out of the inference model damages inference much more than leaving out extinction or SSE processes.

Inference of *e* is known to be highly problematic in DEC (Ree and Smith 2008), a result also observed here with both DEC and DEC+J. Interestingly, while DEC
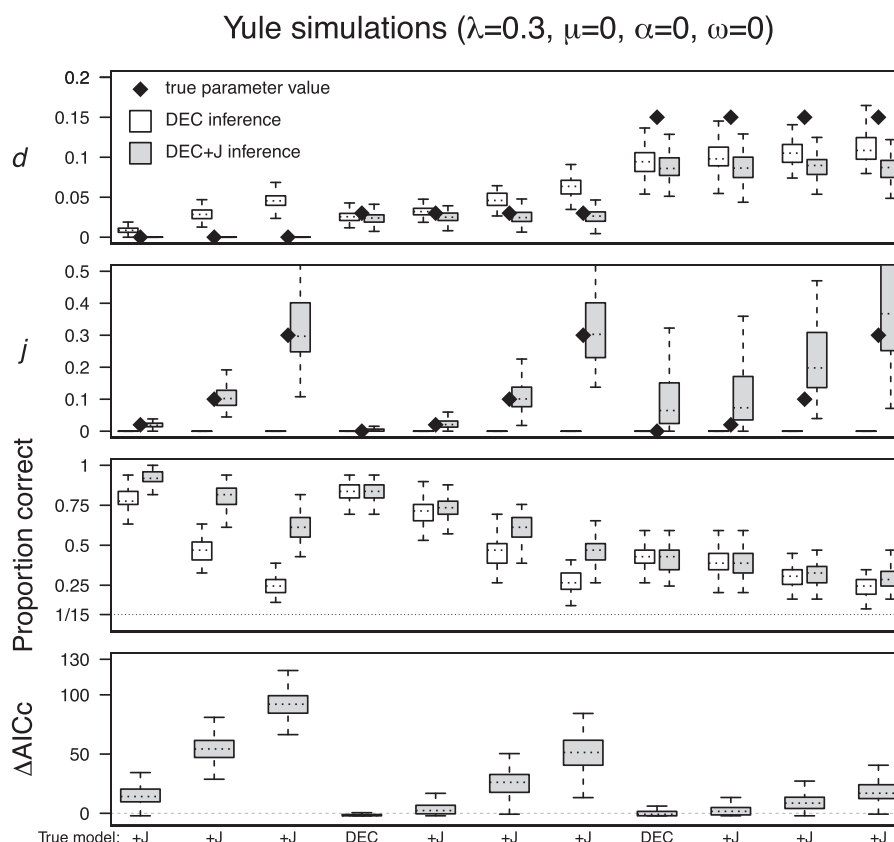
FIGURE 3.   Results of DEC and DEC+J inference on trees/ranges jointly simulated under a Yule process with 11 different combinations of *d*, *e*, and *j* values (black diamonds). Each parameter combination and resulting inferences are shown in a column, with white boxes representing DEC inference, and gray representing DEC+J inference. The first two rows show the accuracy of parameter inference for *d* and *j*. The third row gives the proportion of nodes for which each model inferred the correct ancestral range. The fourth row shows the ΔAICc advantage of DEC+J over DEC. There is no leftmost "DEC" column as *d*, *e*, and *j* would all be 0, and thus from a starting range of one area no change would occur. Only a subset of the simulation parameter combinations is shown, for full results, as well as inference of *e*, see Supplementary Figures S6 and S7.

almost always infers $e = 0$ when the generating model is DEC with positive e, DEC often infers positive e when the true model is DEC+J, even when the true value of e is 0 (Supplementary Figs. S5–S7).

DEC versus DEC + J on Island Clades

All phylogenies, input files, R scripts, and PDF graphics for DEC and DEC+J inferences on the study clades are available on Dryad (http://dx.doi.org/10.5061/dryad.10.2mc1t). Figure 5 shows the relative probabilities of DEC and DEC+J as calculated from AICc weights for each of the 13 study clades and 53 constraints scenarios. Supplementary Table S1 gives LRT *P*-values and AICc weights for all runs. In general, the results overwhelmingly support the inclusion of founder-event speciation, with DEC+J being anywhere from 4 to 10^{52} times more probable than DEC. The only exceptions to this pattern are the time-stratified analyses of *Drosophila* and *Scaptomyza*. In both cases, the issue seems to be highly unrealistic versions of the stratified models, which can occur when clades much older than Kauai are forced into constraints models originally set

up for the younger *Psychotria* clade. In both cases, a more realistic implementation of the constraints model, taken from the source studies, improved the relative performance of DEC+J. In the case of *Drosophila*, this was a constraints model adding a Z area representing an ancient island that later submerged (Kambysellis et al. 1995). In the case of *Scaptomyza*, this meant adding a Z area with a permanently ultralow probability of dispersal (Lapoint et al. 2013), representing a permanent but very distant continent (Supplementary Table S1). The *P*-values of the likelihood-ratio test closely follow the results from AICc model weights, with 51/53 analyses resulting in rejection of the null model (equivalence of DEC and DEC+J) with $P<0.05$. The most common situation is extremely confident rejection of the null ($P<0.00001$), an impressive result given the small size of the data sets (a maximum of 198 OTUs).

Figure 5 and Supplementary Table S1 also give some evidence that the degree of support for DEC+J over DEC is often related to the size of the tree—large analyses can give extremely high AICc ratios and extremely low *P*-values, whereas in small clades (*Plantago*, *Orsonwelles*, silverswords), the support is relatively weaker (although still strong, and statistically significant). However, this
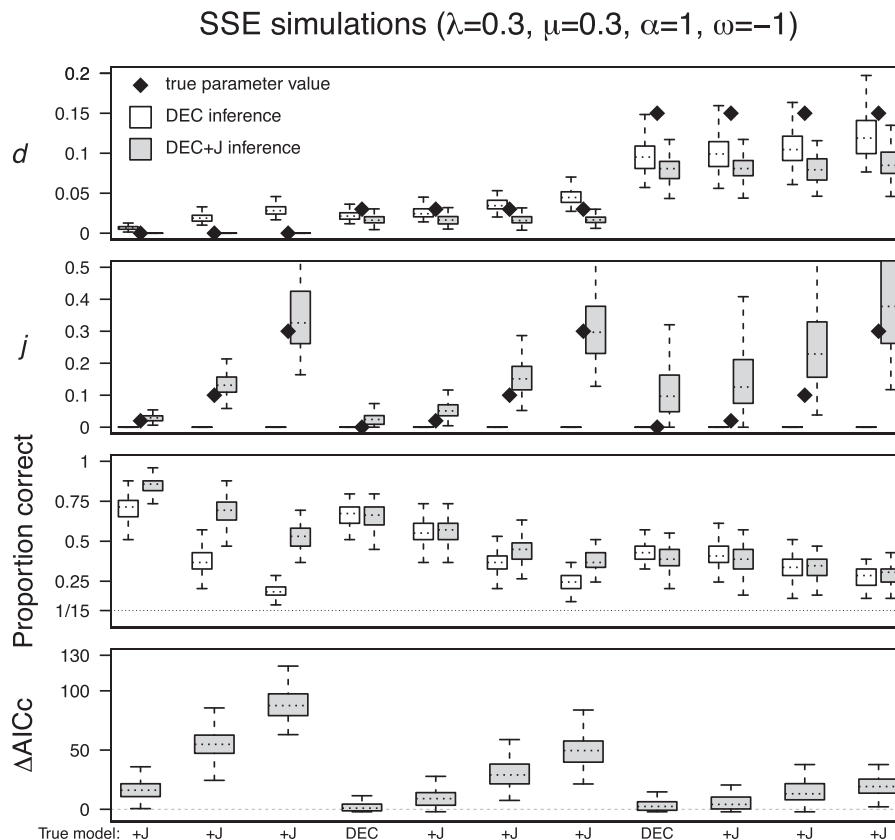
FIGURE 4.   Results of DEC and DEC+J inference on trees/ranges jointly simulated with 11 different combinations of d, e, and j values (black diamonds), and using a SSE (state-dependent speciation/extinction) macroevolutionary process. Interpretation is the same as Figure 3. Only a subset of the simulations is shown, for full results, as well as inference of e, see Supplementary Figures S6 and S7.

pattern is complex, partially because it is confounded with the constraints models, and more complex constraints tend to result in somewhat weaker support for DEC+J. In addition, clades undoubtedly differ in terms of the importance of founder-event speciation in the clade's biogeographic history.

Supplementary Figure S9 shows the ML point estimates of d, e, and j for each clade/constraint combination, under DEC and DEC+J. When j is moderate or large, d and e often drop to zero. As noted above for *Psychotria*, this is evidence that the "DE" part of DEC is not necessary in these clades. However, d and e do not always drop to 0; sometimes all three processes are supported. There also appears to be some correlation of d inference between the DEC and DEC+J models. One data configuration in which d will usually be positive, of course, is when some tips have ranges of more than one area. The founder-event speciation process, as implemented here, only buds off offspring with a range size of one area. As widespread ranges may not be inherited by both descendents under either the DEC or the DEC+J model, most phylogenies with more than one tip with widespread ranges will require a positive d to explain the data. Inference of e is extremely problematic, as discussed previously, so no great confidence should be put on the e estimates. The j parameter is always

positive, but ranges from small to the upper limit ($j =$ 3). Small values of j indicate that only a few nodes of the phylogeny have configurations of tip geography that are conducive to founder-event speciation explanations, that is, subtrees with area arrangements such as (A(A(A,B))). However, when $j = 3$, this means that the values of y, s, and v are 0. In this situation (observed in *Microlophus*), jump dispersal is the *only* process that needs to be invoked to explain distributions, if d and e are 0. Ancestral state inferences for all inference runs are available in PDFs found in Supplementary Data.

DISCUSSION

The picture that emerges across all island clades examined is that the patterns found in *Psychotria* are general: island clades prefer the DEC+J model, and under this model, ancestors are usually inferred to have narrower ranges, ancestral ranges are estimated with higher confidence, and the estimated history is a relatively simple series of within-island and between-island speciation events. This conclusion will of course not surprise many island biogeographers, who have accepted the dominant role of founder-event speciation for decades. However, it should be surprising that a
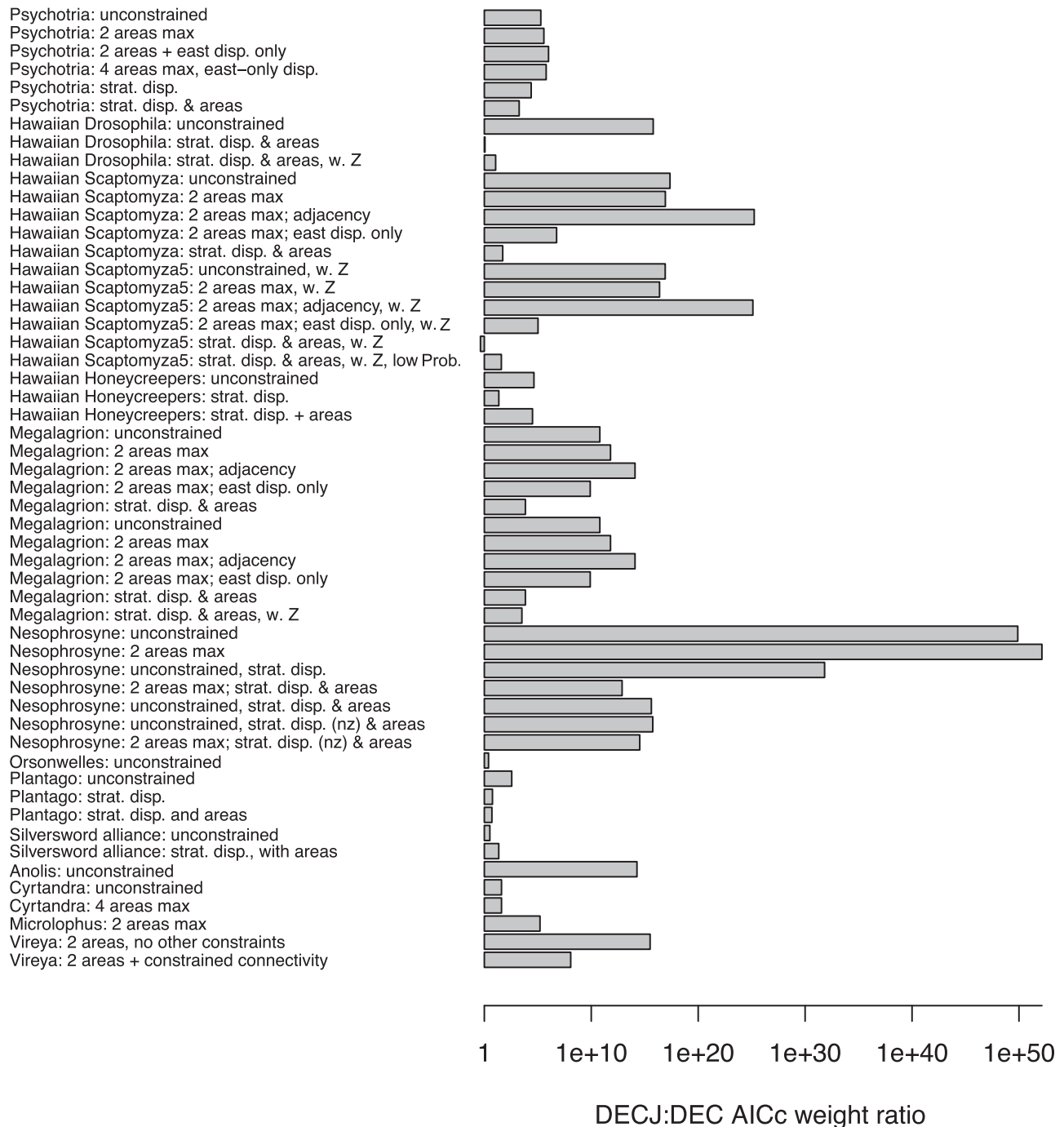
FIGURE 5. The ratio of probabilities of DEC+J to DEC, calculated from AICc weights. The probability ratio is presented for ease of interpretation; ΔAICc or ΔAIC plots look very similar. Note that the probability ratio is plotted on a log scale. Under the likelihood-ratio test, all comparisons of DEC and DEC+J reject DEC at the $P<0.05$ significance level, except for "Hawaiian *Drosophila*: strat. disp. & areas" and "Hawaiian *Scaptomyza*5: strat. disp. & areas". See text for details. Abbreviations: strat., time-stratified; disp., manual dispersal probability multiplier matrix; w. Z, analysis was run with "Z", an old, ancestral area outside of the extant Hawaiian high islands (*Scaptomyza*5 refers to adding this fifth area; "low prob." refers to an extremely low, but nonzero, dispersal probability; see Supplementary Files); nz, zeros in the manual dispersal probability multiplier matrix have been replaced with a small nonzero value.

model that is apparently inappropriate in island systems, namely DEC, has been so widely applied to them.

The DEC model was revolutionary. It was the advent of parametric biogeography (Ree and Sanmartín 2009), wherein rather than simply relying on parsimony reconstructions, events are assigned probabilities, and optimal estimates of history are made by ML or Bayesian approaches. However, probably because it was the first major effort, LAGRANGE did not fully explore all of the potential inherent in the parametric approach.

A major advantage of an explicit, probabilistic modeling framework is access to standard tools for model comparison and model choice. All that is required to use these tools is the implementation of new models within the same framework. This has been done here by the creation of the DEC+J model. Simulation experiments showed that the models are identifiable even with few data, and also showed acceptable estimation of d and j, but not e. Finally, the simulation experiments showed that there is an intimate relationship between assumptions about biogeographic process, and the accuracy of estimations of biogeographic history. The hazards of inferring history under the incorrect model are dramatic—in the *Psychotria* M0 example, when DEC+J was the true process, the accuracy of ancestral range estimation under DEC was 57%, but under DEC+J it was 87%.

Comparisons of DEC and DEC+J on real data sets show that across a broad sample of island clades, there is massive support for DEC+J over DEC, over a wide range of constraints models. The differences are of approximately the same order of magnitude as those found when comparing DEC inference and DEC+J inference on simulated DEC+J data. In other words, real-world island data sets appear as if they were simulated under the DEC+J model, rather than the DEC model. Several side-benefits accrued from using a better model, namely, simpler estimated histories, more confident estimates at ancestral nodes, and elimination of the bias toward widespread ancestors sometimes found in DEC.

What is the mathematical cause of the dramatic improvement in likelihoods under the DEC+J model, on island data sets consisting primarily of single-island endemics? A hypothesis is presented here. In order to distribute lineages across the study region, the DEC model is forced to use range-expansion events, which are anagenetic events controlled by a continuous-time rate matrix, and these events must be followed by the correct cladogenesis events. This means that in order for vicariance or subset speciation to happen at a node on the tree, some time before that node, the correct dispersal event must have happened, and then, by chance, no other (cumulative) changes, either dispersal or extinction, may happen before the speciation event. Then, in order for the tip to be a single-island endemic, no further cumulative changes to range may occur. Thus, a positive d is required to spread lineages out, but having a positive d means that branch histories that produce the correct ranges at the correct times will be somewhat improbable, compared with all of the other possible histories. These waiting times are what lowers likelihoods in the DEC model. By contrast, on the DEC+J model, d and e are often low or zero, meaning the probability of no change anagenetically along a branch is high. Cladogenesis and dispersal are perfectly correlated in founder-event speciation events, and if the species are single-island endemics, a simple history of founder events can explain the distribution. Simulation results make the crucial point that it is perfectly possible to observe geographic distributions that do *not* favor

DEC+J over DEC; however, it appears that taxa with such distributions are rare in island systems.

Comparison of DEC + J to Other Methods

A variety of models and methods already exist in historical biogeography. DEC+J is not identical with any of them, but exploration of the features it does share with each method helps to shed light on the assumptions and limitations of each method.

Fitch parsimony.—Because DEC+J often estimates histories with a small number of between-region "jumps", it is tempting to compare biogeographical inference under DEC+J to inference under Fitch parsimony (FP), a method that long predates probabilistic methods in historical biogeography (Bremer 1992). Under FP, geographic range is treated as an unordered character, and lineages may not inhabit more than one area at a time. The biogeographic histories estimated by DEC+J often resemble those of FP. In situations where DEC+J infers $d = e = 0$ (common but not universal in this study's data sets), the length of branches becomes irrelevant, increasing the resemblance. However, DEC+J is not FP. Most importantly, the optimality criteria are different: just like DEC, DEC+J is a probabilistic model that calculates the probability of the data under all histories possible under the model, and thus confers a likelihood on the tip data. This allows use of standard ML and Bayesian inference machinery. DEC+J contains DEC as a nested sub-model, and thus all of the advantages of DEC over FP accrue to DEC+J, including allowing species to live in multiple areas, allowing range expansion, vicariance, and subset sympatry to be inferred when indicated by the tip data, and formal inclusion of time in the analysis, such that users can specify changes in geography in different time strata. The likelihood produced by probabilistic model can be statistically compared with the likelihoods produced by other probabilistic models that use the same data, for example, comparing DEC and DEC+J. Comparison of two parsimony methods, in contrast, is difficult. There is no formal statistical method biogeographers can use to decide whether FP or DIVA is the better method to use on their data.

Probabilistic island models.—FP is a stepping-stone "island model", but DEC+J might also be compared with parametric "island models" reviewed by Ronquist and Sanmartín (2011). Such models are widespread, for example in "discrete phylogeography" analyses in BEAST. As with FP, these models assume that lineages must live in one area at a time, however, like DNA evolution models, the probability of change is controlled by a rate matrix operating in continuous time. This has the peculiar effect of modeling change in distribution as an instantaneous "range-switching" process, where a lineage jumps from region A to B without leaving any members of the population behind in A. The

crucial difference between this process and founder-event speciation is that the latter is associated with cladogenesis events, where only one of the two daughter lineages jumps to a new region. This improvement in realism is a reason to favor DEC+J, but the models could be statistically compared, and it is possible that in certain situations (e.g., strong correlation between the number of speciation events and branch length, perhaps due to high extinction) "range-switching" could be a reasonable approximation.

Another point of dissimilarity concerns the "base frequencies" aspect of parametric island models, which model residence times of lineages in each island. Because range contraction to a null range is irreversible, DEC-type models with positive e have a long-term stationary distribution of "extinct in all areas." When likelihoods are calculated under DEC and DEC+J, base frequencies are ignored at the root, which in effect assumes equal frequencies of each state. These are strange features and ideally should be eliminated either by correcting for exclusion of the null range from the state space (Landis et al. 2013) and/or explicit inclusion of the macroevolutionary birth/death process as in SSE models (e.g., Goldberg et al. 2011; FitzJohn 2012).

The fact that DEC+J infers "island-hopping" histories for island data, but infers the DEC model when DEC is the generating model, means that DEC+J has taken a significant step toward unifying Ronquist and Sanmartín's categories of "island models" and "reticulate models." Further elaboration of this approach with the BioGeoBEARS supermodel may allow inclusion of additional categories of models, including novel ones not before considered.

DIVA.—Founder-event speciation is not formally included in the processes assumed by DIVA (Kodandaramaiah 2010; Ronquist and Sanmartín 2011;. However, as a parsimony method, DIVA ignores branch lengths, and thus the cladogenesis aspect of founder-event speciation might be "interpreted" into DIVA by the user, by assuming that dispersal (range expansions) happened just before vicariance events at cladogenesis. There is nothing in the DIVA algorithm that says whether the inferred range-expansion events usually occur in the middle of branches or at the end of them, although it appears that it is common to assume the former, for example when range expansion in DIVA and DEC are analogized (Ronquist and Sanmartín 2011). A more practical consideration is that DIVA would only be likely to produce founder-event-like inferences if the maximum range size is set to 2 areas (Kodandaramaiah 2010). Without this constraint, DIVA often infers ancestors that are more and more widespread at deeper nodes, suggesting a history where widespread ranges persist throughout entire branches (Kodandaramaiah 2010). The conflation of cladogenetic and anagenetic range-changing processes in DIVA (and FP) is a significant problem that can be addressed by explicitly modeling each process of interest in BioGeoBEARS.

SSE models.—Finally, DEC+J and DEC merit comparison to "SSE" models such as GeoSSE and ClaSSE. GeoSSE (Goldberg et al. 2011) assumes the same types of cladogenesis events as DEC, but each is treated as a macroevolutionary process that is controlled by a different rate parameter. This results in a large number of free parameters even for a two-area system, and would require many more parameters for systems with more areas, suggesting the need for large data sets for effective inference. The closest approach that GeoSSE could make to founder-event speciation would be to have a very high rate of vicariant speciation when a lineage occupies multiple ranges. This would tend to produce speciation events immediately following range expansion. Arguably, this follows the intuition that immediately after long-distance dispersal, the same species exists in two regions, and it would take some amount of time and evolutionary change before a taxonomist would split the two populations into two species.

However, such a formulation has the disadvantage of conflating two very different processes, namely, classic vicariance, wherein a geological or environmental event breaks up the population independently of the time since a lineage became widespread, and founder-event speciation, where cladogenesis is intimately connected to the original dispersal event. It is possible that both processes have occurred in the history of a particular clade, especially if different lineages have different dispersal ability, and if so, modeling the two processes with a single rate parameter would produce only an "average rate of vicariant speciation" as a compromised estimate giving the mean of what is actually a bimodal process. This consideration argues for modeling classic vicariance and founder-event speciation as distinct processes, at least until more sophisticated models are available that explicitly link traits and distance to maintenance of gene flow and probability of speciation.

In addition, while the intuition that the same species exists in two places after a rare long-distance dispersal event is appealing from the point of view of a morphological species concept, this concept may be a misleading representation of genetic and phylogenetic reality in cases where a rare colonization event establishes a genetically isolated population. Such a population would be completely isolated from its ancestral gene pool and thus be a new lineage on any mechanistic definition, whether or not it would initially be "different enough" to be named a new species. It would take some time after a founder event for genetic isolation to become detectable in genetic data, but the bottleneck effects of population genetic founder events are so strong that genetic isolation is likely to be detectable within a few generations, effectively instantaneous on historical biogeography timescales.

An SSE method which could include both classic vicariance and founder-event speciation as separate processes is ClaSSE (Goldberg and Igić 2012), as it is fully generic, creating parameters for each anagenetic or cladogenetic process conceivable in the state space.

However, to limit the explosion of free parameters, many of the parameters would have to be constrained to be equal, or fixed to zero, as is implicitly done in DEC-type models. This would necessitate careful exploration of which cladogenetic range-changing processes are supported by the data. The SSE models have the advantage of automatically correcting for extinction and state dependence, but even simple SSE models have been shown to require large data sets (Davis et al. 2013). Also, state spaces in biogeographic problems often contain over 500 geographic ranges, and it remains to be seen if the systems of differential equations used to calculate likelihoods under SSE can be effectively scaled to typical historical biogeography problems.

In any event, simulation results show that leaving out extinction and SSE processes may not be as damaging to inference as leaving out founder-event speciation. It appears that as long as geographic change happens rarely enough that geographic range shows phylogenetic signal on the observed tree, then enough data exist for effective model choice and inference of ancestral states.

Conclusion

Development and testing of improved models for DNA substitution have a long and distinguished history in phylogenetics (Posada and Crandall 2001). The most important point of this article is that model choice can and should be applied in historical biogeography. I suggest that model selection be instituted as a regular procedure in historical biogeography analyses, as is already the case in molecular phylogenetics. Obviously this should not be restricted merely to the DEC and DEC+J models. Within the BioGeoBEARS framework, it is easy to convert the DEC model into a DIVA-like model (Ronquist 1997; Ronquist and Sanmartín 2011), or a BayArea-like model (Landis et al. 2013), and from there to create, for example, DIVALIKE+J and BAYAREALIKE+J models. Additional models, as-yet unnamed, can be implemented by fixing y, s, or v to 0 or other values, or allowing them to vary independently to test if the improvement in likelihoods warrants the increase in model complexity.

In this study, a strong signal for founder-event speciation was discovered. How general are these results likely to be? It is difficult to extrapolate from the unusual situation of oceanic islands to the much more complex biogeography within and between continents. However, simulation results showed that there are data configurations for which DEC+J will not outperform DEC. Investigation of these simulated ranges indicates that whereas DEC+J realizations tend to have tip species with ranges of single areas (Supplementary Fig. S8.1–S8.6, "range size" panel), often allopatric, DEC realizations tend to have more species that are widespread, and a fair degree of overlap and sympatry. I have found one empirical data set where DEC is not rejected (data not shown), namely "*Taygetis* clade" butterflies (Matos-Maraví et al. 2013), many

species of which are widespread across the Neotropics. Thus, intracontinental clades with many widespread, sympatric species might be appropriately modeled by DEC. On the other hand, studies that treat continents as areas, with most species endemic to single continents, may be fit better by the DEC+J model. An interesting data point is a new analysis of the global biogeography of Amphibia using DEC and DEC+J (Pyron 2014), which finds either weak or no support for DEC+J over DEC, depending on the subanalysis. This finding is in accord with amphibians' extremely low, but not quite zero (de Queiroz 2014), powers of dispersal over salt water.

Given that the importance of founder-event speciation has long been conventional wisdom among island biogeographers, some might consider the use of elaborate modeling and statistical techniques to confirm conventional wisdom to be unnecessary. However, categorical dismissal of founder-event speciation is still popular among some proponents of vicariance biogeography (Santos 2007; Heads 2012). Vicariance biogeography has been widely criticized (Waters et al. 2013; de Queiroz 2014), but much less attention has been paid to the fact that the most popular inference programs in the field make similar strong assumptions, and these assumptions have a strong impact on inference. Now that direct statistical comparison of biogeographical models has become accessible, rather than simply assuming *a priori* which processes will be considered during analysis, we may let the data tell us which models are to be preferred (Huelsenbeck 1995; Huelsenbeck et al. 2001).

Supplementary Material

Supplementary material, including data files and/or online-only appendices, can be found in the Dryad data repository http://dx.doi.org/10.5061/dryad.10.2mc1t.

Landis, Josh Schraiber, Brian Moore, Michael Donoghue, Hannah Wood, Rosie Gillespie, Chelsea Specht, Kevin Padian, Brent Mishler, David Lindberg, Roger Byrne, David Jablonski, and Tony Barnosky. Thanks especially to Rick Ree and Stephen Smith for putting their programs online and for helpfully answering questions. I am indebted to the editors, and reviewers Rick Ree and Fredrik Ronquist, for helpful comments on this article.

REFERENCES

Baldwin B.G., Sanderson M.J. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Nat. Acad. Sci., 95:9402–9406.

Benavides E., Baum R., Snell H.M., Snell H.L., Sites J.W. 2009. Island biogeography of galápagos lava Lizards (Tropiduridae: *Microlophus*): species diversity and colonization of the Archipelago. Evolution, 63:1606–1626.

Bennett G.M., O'Grady P.M. 2013. Historical biogeography and ecological opportunity in the adaptive radiation of native Hawaiian leafhoppers (Cicadellidae: *Nesophrosyne*). J. Biogeogr., 40:1512–1523.

Box G.E.P., Draper N.R. 1987. Empirical model-building and response surfaces. New York (NY): Wiley.

Bremer K. 1992. Ancestral areas: a cladistic reinterpretation of the center of origin concept. Syst. Biol. 41:436–445.

Buerki S., Forest F., Alvarez N., Nylander J.A.A., Arrigo N., Sanmartín I. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. J. Biogeogr. 38:531–550.

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference : a practical information-theoretic approach. 2nd ed. New York (NY): Springer.

Byrd R.H., Lu P., Nocedal J., Zhu C. 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16:1190–1208.

Carlquist S.J. 1974. Island biology. New York (NY): Columbia University Press.

Carson H.L., Templeton A.R. 1984. Genetic revolutions in relation to speciation phenomena: the founding of new populations. Annu. Rev. Ecol. Syst. 15:97–131.

Clague D.A. 1996. The growth and subsidence of the Hawaiian-Emperor volcanic chain. In: Keast A., Miller S.E., editors. The origin and evolution of Pacific Island Biotas, New Guinea to Eastern Polynesia: patterns and processes. Amsterdam: SPB Academic Publishing. p. 35–50.

Clark J.R., Ree R.H., Alfaro M.E., King M.G., Wagner W.L., Roalson E.H. 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. Syst. Biol. 57:693–707.

Clark JR, Wagner WL, Roalson EH. 2009. Patterns of diversification and ancestral range reconstruction in the southeast Asian–Pacific angiosperm lineage *Cyrtandra* (Gesneriaceae). Mol. Phylogenet. Evol., 53:982–994.

Cowie R.H., Holland B.S. 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. J. Biogeogr. 33:193–198.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Davis M.P., Midford P.E., Maddison W. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evol. Biol. 13:38.

de Queiroz A. 2005. The resurrection of oceanic dispersal in historical biogeography. Trends Ecol. Evol. 20:68–73.

Dunbar-Co S. 2008. Island evolution: phylogeny, adaptive radiation, and biogeography of *Plantago* (Plantaginaceae) in the Hawaiian Islands. Ph.D., University of Hawai'i.

de Queiroz A. 2014. The monkey's voyage: how improbable journeys shaped the history of life. New York (NY): Basic Books.

Eddelbuettel D, Francois R. 2013. Rcpp: Seamless R and C++ Integration, CRAN: The Comprehensive R Archive Network, Vienna, Austria. http://cran.r-project.org/package=Rcpp, (last accessed August 7, 2014).

Escoto B. 2011. Marginal likelihood computation via arrogance sampling. arXiv.org:1-12.

Felsenstein J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution 35:1229–1242.

Felsenstein J. 2004. Inferring phylogenies. Sunderland (MA): Sinauer Associates.

FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.

Gillespie R.G., Baldwin B.G. 2010. Island biogeography of remote archipelagos: Interplay between ecological and evolutionary processes. In: Losos J.B., Ricklefs R.E., MacArthur R.H., editors. The theory of island biogeography revisited. Princeton, Princeton University Press, p. 358–387.

Gillespie R.G., Baldwin B.G., Waters J.M., Fraser C.I., Nikula R., Roderick G.K. 2012. Long-distance dispersal: a framework for hypothesis testing. Trends Ecol. Evol. 27:47–56.

Goldberg E.E., Igić B. 2012. Tempo and mode in plant breeding system evolution. Evolution 66:3701–3709.

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

Gould S.J., Eldredge N. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In: Thomas J. M. Schopf, editor., Models in paleobiology. San Francisco: Freeman & Cooper, p. 82–115.

Green P.J. 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732.

Harmon L.J., Weir J.T., Brock C.D., Glor R.E., Challenger W. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129–131.

Heads M.J. 2012. Molecular panbiogeography of the tropics. Berkeley: University of California Press.

Hormiga G., Arnedo M., Gillespie R.G. 2003. Speciation on a conveyor belt: sequential colonization of the Hawaiian Islands by *Orsonwelles* Spiders (Araneae, Linyphiidae). Syst. Biol., 52:70–88.

Huelsenbeck J.P. 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.

Huelsenbeck J.P., Ronquist F., Nielsen R., Bollback J.P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

Jordan S., Simon C., Polhemus D. 2003. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus megalagrion (Odonata: Coenagrionidae). Syst. Biol., 52:89–109.

Kambysellis M.P., Ho K.-F., Craddock E.M., Piano F., Parisi M., Cohen J. 1995. Pattern of ecological shifts in the diversification of Hawaiian *Drosophila* inferred from a molecular phylogeny. Curr. Biol. 5:1129–1139.

Kodandaramaiah U. 2010. Use of dispersal–vicariance analysis in biogeography—a critique. J. Biogeogr. 37:3–11.

Landis M., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62(6): 789–804.

Lapoint R.T., O'Grady P.M., Whiteman N.K. 2013. Diversification and dispersal of the Hawaiian Drosophilidae: the evolution of Scaptomyza. Mol. Phylogenet. Evol. 69:95–108.

Lerner H.R.L., Meyer M., James H.F., Hofreiter M., Fleischer R.C. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian Honeycreepers. Curr. Biol., 21:1838–1844.

Link W.A., Barker R.J. 2009. Bayesian inference: with ecological applications. Boston: Elsevier/Academic.

Lomolino M.V., Riddle B.R., Whittaker R.J., Brown J.H. 2010. Biogeography. 4th ed. Sunderland (MA): Sinauer Associates.

Maguire K.C., Stigall A.L. 2008. Paleobiogeography of miocene equinae of North America: a phylogenetic biogeographic analysis of the relative roles of climate, vicariance, and dispersal. Palaeogeogr. Palaeoclimatol. Palaeoecol. 267:175–184.

Matos-Maraví P.F., Peña C., Willmott K.R., Freitas A.V.L., Wahlberg N. 2013. Systematics and evolutionary history of butterflies in the "Taygetis clade" (Nymphalidae: Satyrinae: Euptychiina): towards a better understanding of Neotropical biogeography. Mol. Phylogenet. Evol. 66:54–68.

Matzke N.J. 2013a. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts, CRAN: The Comprehensive R Archive Network, Vienna, Austria. http://cran.r-project.org/package=BioGeoBEARS (last accessed August 7, 2014).

Matzke N.J. 2013b. cladoRcpp: C++ implementations of phylogenetic calculations, CRAN: The Comprehensive R Archive Network, Vienna, Austria. http://cran.r-project.org/package=cladoRcpp (last accessed August 7, 2014).

Matzke N.J. 2013c. TreeRogue: R code for digitizing trees. https://stat.ethz.ch/pipermail/r-sig-phylo/2010-October/000816.html (last accessed August 7, 2014).

Matzke N.J., Sidje R.B. 2013. rexpokit: R wrappers for EXPOKIT; other matrix functions, CRAN: The Comprehensive R Archive Network, Vienna, Austria. http://cran.r-project.org/package=rexpokit (last accessed August 7, 2014).

Mayr E. 1954. Change of genetic environment and evolution. In: Huxley J., Hardy A.C., Ford E.B., editors. Evolution as a process. London: Allen & Unwin. p. 157–180.

Moler C., Van Loan C. 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev. 45:3–49.

Mooers A. 2004. Effects of tree shape on the accuracy of maximum likelihood-based ancestor reconstructions. Syst. Biol. 53: 809–814.

Mooers A., Schluter D. 1999. Reconstructing ancestor states with maximum likelihood: support for one-and two-rate models. Syst. Biol. 48:623–633.

Nash J.C., Varadhan R. 2011. Unifying optimization algorithms to aid software system users: optimx for R. J. Stat. Softw. 43:1–14.

Nash J.C., Varadhan R. 2012. optimx: a replacement and extension of the optim() function, CRAN: The Comprehensive R Archive Network, Vienna, Austria. http://cran.r-project.org/package=optimx (last accessed August 7, 2014).

Nepokroeff M., Sytsma K.J., Wagner W.L., Zimmer E.A. 2003. Reconstructing ancestral patterns of colonization and dispersal in the Hawaiian understory tree genus Psychotria (Rubiaceae): a comparison of parsimony and likelihood approaches. Syst. Biol. 52:820–838.

Nicholson K.E., Crother B.I., Guyer C., Savage J.M. 2012. It is time for a new classification of anoles (Squamata: Dactyloidae). Zootaxa, 3477:1–108.

Paulay G., Meyer C. 2002. Diversification in the tropical pacific: comparisons between marine and terrestrial systems and the importance of founder speciation. Integr. Comp. Biol. 42: 922–934.

Posada D., Crandall K.A. 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50:580–601.

Pyron R.A. 2014. Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. Syst. Biol. Available online 19 June 2014.

Ree R.H. 2013. Lagrange configurator version 20130526 (beta). http://www.reelab.net/lagrange/configurator/index (last accessed August 7, 2014).

Ree R.H., Sanmartín I. 2009. Prospects and challenges for parametric models in historical biogeographical inference. J. Biogeogr. 36:1211–1220.

Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.

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

Ronquist F. 1997. Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. Syst. Biol. 46:195–203.

Ronquist F., Sanmartín I. 2011. Phylogenetic methods in biogeography. Annu. Rev. Ecol. Evol. Syst. 42:441–464.

Sanderson M.J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.

Sanmartín I. 2012. Historical biogeography: evolution in time and space. Evo. Edu. Outreach. 5:555–568.

Santos C.M.D. 2007. On basal clades and ancestral areas. J. Biogeogr. 34:1470–1469.

Sidje R.B. 1998. Expokit: a software package for computing matrix exponentials. ACM Trans. Math. Softw. 24:130–156.

Smith S.A., Donoghue, M.J. 2010. Combining historical biogeography with niche modeling in the Caprifolium clade of *Lonicera* (Caprifoliaceae, Dipsacales). Syst. Biol. 59:322–341.

Smith S.A., Ree R.H. 2010. Lagrange: likelihood analysis of geographic range evolution, https://code.google.com/p/lagrange/ (last accessed August 7, 2014).

Statisticat L. 2013a. LaplacesDemon Tutorial. Statisticat, Hot Springs, South Dakota. http://www.bayesian-inference.com/softwaredownload (last accessed August 7, 2014).

Statisticat L. 2013b. LaplacesDemon: complete environment for Bayesian inference. Statisticat, Hot Springs, South Dakota. http://www.bayesian-inference.com/softwaredownload (last accessed August 7, 2014).

Templeton A.R. 2008. The reality and importance of founder speciation in evolution. BioEssays 30:470–479.

Waters J.M., Trewick S.A., Paterson A.M., Spencer H.G., Kennedy M., Craw D., Burridge C.P., Wallis G.P. 2013. Biogeography off the tracks. Syst. Biol. 62:494–498.

Webb C.O., Ree R.H. 2012. Historical biogeography inference in Malesia. In: Gower D., Johnson K., Richardson J., Rosen B., Ruber L., Williams S. editors. Biotic evolution and environmental change in Southeast Asia. Cambridge, Cambridge University Press, p. 191–215.

Wiley E.O. 1981. Phylogenetics: the theory and practice of phylogenetic systematics. New York (NY): Wiley.