UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing

Permalink

https://escholarship.org/uc/item/8227p52c

Author

Matzke, Nicholas Joseph

Publication Date

2013

Peer reviewed|Thesis/dissertation

Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing

by

Nicholas J. Matzke

A dissertation submitted in partial satisfaction of the requirements for the degree of
Doctor of Philosophy

in

Integrative Biology

and the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Huelsenbeck, Chair

Professor David Jablonski

Professor Anthony Barnosky

Professor Roger Byrne

Fall 2013

Abstract

Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing

by

Nicholas J. Matzke

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor John Huelsenbeck, Chair

Historical biogeography has a diversity of methods for inferring ancestral geographic ranges on phylogenies, but many of the methods have conflicting assumptions, and there is no common statistical framework by which to judge which models are preferable. Probabilistic modeling of geographic range evolution, pioneered by Ree and Smith (2008, *Systematic Biology*) in their program LAGRANGE, could provide such a framework, but this potential has not been implemented until now.

I have created an R package, "BioGeoBEARS," described in chapter 1 of the dissertation, that implements in a likelihood framework several commonly used models, such as the LAGRANGE Dispersal-Extinction-Cladogenesis (DEC) model and the Dispersal-Vicariance Analysis (DIVA, Ronquist 1997, *Systematic Biology*) model. Standard DEC is a model with two free parameters specifying the rate of "dispersal" (range expansion) and "extinction" (range contraction). However, while dispersal and extinction rates are free parameters, the cladogenesis model is fixed, such that the geographic range of the ancestral lineage is inherited by the two daughter

lineages through a variety of scenarios fixed to have equal probability. This fixed nature of the cladogenesis model means that it has been indiscriminately applied in all DEC analyses, and has not been subjected to any inference or formal model testing.

The process of founder-event speciation, thought to be crucial especially in island systems, is completely left out of the DEC and DIVA models, but it is implemented as an option in BioGeoBEARS, enabling the creation of models such as DEC+J, DIVA+J, etc. The models in BioGeoBEARS are fully parameterized, so that users can easily create new models of their own devising (e.g., vicariance only, founder-event speciation only, any combination of these, etc.) by setting parameters to 0 or 1. Alternatively, parameters controlling various processes can be set to be free parameters, and estimated from the data. Implementation of all models in a common framework allows use of standard statistical model choice procedures such as the Likelihood Ratio Test (LRT) or Akaike Information Criterion (AIC) to objectively compare models and hypothesis about the biogeographical processes operating in different clades and regions.

BioGeoBEARS also adds a number of features not previously available in most historical biogeography software, such as distance-based dispersal, a model of imperfect detection, and the ability to include fossils either as ancestors or tips on a time-calibrated tree.

In Chapter 2, I validate BioGeoBEARS by showing that it exactly reproduces the log-likelihoods and parameter inferences made by the LAGRANGE DEC model on the LAGRANGE test dataset of the Hawaiian *Psychotria* clade. I further validate the method by taking the *Psychotria* phylogeny and simulating geographic range evolution under the DEC and DEC+J models, and then conducting inference under the two models. Model choice using LRT is highly accurate, with false positive and false negative rates of approximately 5%, indicating that the test has the desired frequentist properties, and also indicating that DEC and DEC+J are easy to distinguish from the data, even on a small phylogeny. The simulation results also indicate that when DEC+J is the true model, DEC+J has 87% accuracy in inferring ancestral states, while DEC has only 57% accuracy.

The DEC and DEC+J models are then applied to 13 island clades, most of them classic Hawaiian study systems (*Drosophila*, silverswords, etc.), under a variety of dispersal constraint scenarios. Standard statistical model comparisons show that DEC+J is vastly superior to standard DEC for all clades, for the first time verifying the importance of founder-event speciation in island clades via statistical model choice, and falsifying vicariance-dominated models of island biogeography. The case of *Psychotria* is typical: the DEC+J model is about 300,000 times more probable than the DEC model in an unconstrained analysis, according to AIC weights. Furthermore, the inferred maximum likelihood (ML) estimates of parameters often differ radically under the DEC+J model, with the "DE" part of the model sometimes playing no role (i.e., the parameters *d* and *e*, controlling anagenetic range expansion and range

contraction, are inferred to be 0). Further more, under DEC+J, ancestral nodes are usually estimated to have ranges occupying only one island, rather than the widespread ancestors often favored by DEC.

Chapter 3 expands this analysis to compare the cladogenesis models used by the programs LAGRANGE, DIVA, and BayArea (Landis et al. 2013, *Systematic Biology*). (The BayArea program actually ignores cladogenesis, which identical to assuming that the ancestral range is copied, unmodified, to both daughter lineages at each cladogenesis event.) These models, along with +J versions, are run on a samples of island clades and non-island (continental and oceanic) clades. Almost all analyses, including continental clades, strongly favored the "+J" models over the models without founder-event speciation. However, founder-event speciation was measurably less frequent in non-island analyses, being 2-4 times weaker than in analyses of island clades. Only one clade was found ("*Taygetis* clade" butterflies from the Neotropics) which favored the DEC model over all others.

Chapter 4 addresses the problem of including fossils in the inference of geographic range evolution on phylogenies. This is done by taking into account the fact that detection of presence and absence in regions will often be imperfect for fossil taxa. A hierarchical model is use to link a probabilistic model of imperfect detection with the traditional likelihood calculations of geographic range evolution. The NEOMAP database is used to provide occurrence data through time for two example clades with good fossil records, namely, North American Canidae and Equinae. The

database is also used to provide counts of occurrences of taphonomic control groups that are used to measure relative sampling effort in each region and time bin. The two clades are found to prefer different models for cladogenesis: equids favor DEC, but canids favor BAYAREA+J. This result is found both with and without usage of the imperfect detection model. Ironically, in test data chosen because of their high-quality fossil record, the record was so good that the model for imperfect detection had little impact. However, modeling imperfect detection is likely to be extremely useful in situations with poorer data, or with subsampled data.

Several important conclusions may be drawn from this research. First, formal model selection procedures can be applied in phylogenetic inferences of historical biogeography, and the relative importance of different processes can be measured. These techniques have great potential for strengthening quantitative inference in historical biogeography. No longer are biogeographers forced to simply assume, consciously or not, that some processes (such as vicariance or dispersal) are important and others are not; instead, this can be inferred from the data. Second, founder-event speciation appears to be a crucial explanatory process in most clades, the only exception being some intracontinental taxa showing a large degree of sympatry across widespread ranges. This is not the same thing as claiming that founder-event speciation is the *only* important process; founder event speciation as the only important process is inferred in only one case (*Microlophus* lava lizards from the Galapagos). The importance of founder-event speciation will not be surprising to most island biogeographers. However, the results are important

nonetheless, as there are still some vocal advocates of vicariance-dominated approaches to biogeography, such as Heads (2012, *Molecular Panbiogeography of the Tropics*), who allows vicariance and range-expansion to play a role in his historical inferences, but explicitly excludes founder-event speciation *a priori*. The commonly-used LAGRANGE DEC and DIVA programs actually make assumptions very similar to those of Heads, even though many users of these programs likely consider themselves dispersalists or pluralists. Finally, the inclusion of fossils and imperfect detection within the same likelihood and model-choice framework clears the path for integrating paleobiogeography and neontological biogeography, strengthening inference in both.

Model choice is now standard practice in phylogenetic analysis of DNA sequences: a program such as ModelTest is used to compare models such as Jukes-Cantor, HKY, GTR+I+G, and to select the best model before inferring phylogenies or ancestral states. It is clear that the same should now happen in phylogenetic biogeography. BioGeoBEARS enables this procedure. Perhaps more importantly, however, is the potential for users to create and test new models. Probabilistic modeling of geographic range evolution on phylogenies is still in its infancy, and undoubtedly there are better models out there, waiting to be discovered. It is also undoubtedly true that different clades and different regions will favor different processes, and that further improvements will be had by linking the evolution of organismal traits (e.g., loss of flight) with the evolution of geographic range, within a common inference framework. In a world of rapid climate change and habitat loss,

biogeographical methods must maximize both flexibility and statistical rigor if they are to play a role. This research takes several steps in that direction.

BioGeoBEARS is open-source and is freely available at the Comprehensive R Archive Network (http://cran.r-project.org/web/packages/BioGeoBEARS/index.html). A step-by-step tutorial, using the *Psychotria* dataset, is available at PhyloWiki (http://phylo.wikidot.com/biogeobears).

Dedication

This work is dedicated to my parents, Gordon and Mary Ann Matzke, who are

thanked for their endless support and patience as I found my own path.

Acknowledgements

Chapter 1: BioGeoBEARS: an R package for model testing and ancestral state estimation in historical biogeography

Summary

1. Historical biogeography has a diversity of methods for inferring ancestral geographic ranges on phylogenies, but many of the methods have conflicting assumptions, and there is no common statistical framework by which to judge which models are preferable.

2. The R package BioGeoBEARS implements several common models (e.g., LAGRANGE, DIVA) as well as new models including processes such as founder-event speciation. As the models are fully parameterized, users can now statistically compare models using standard model choice metrics such as AIC.

3. BioGeoBEARS adds a number of features, such as distance-based dispersal, a model of imperfect detection, and inclusion of fossils, not previously available in most historical biogeography software.

4. An example compares the default LAGRANGE model (DEC) to a model adding founder-event speciation (DEC+J), on the LAGRANGE test dataset of the Hawaiian shrub *Psychotria*. DEC+J dramatically outperforms DEC, and results in substantially different inferences about ancestral ranges.

Introduction

Historical biogeography is the field devoted to inferring the history by which living species came to have their present geographic ranges on the globe, and to studying the processes that have produced these distributions (Crisci 2001). The field dates back to Wallace (1855) and Darwin (1859), but much of its history is contentious, with rival schools of thought conducting decades-long arguments primarily over starting assumptions. A classic example in the phylogenetic branch of historical biogeography is the "vicariance versus dispersal" debate. Under the twin influences of the cladistics revolution and the plate tectonics revolution, vicariance proponents enjoyed several decades of dominance (Nelson & Rosen 1981; Cracraft & Shipp 1988), but with the advent of molecular phylogenetics and molecular dating in the 1990s, many reached the conclusion that many clades were too young to be explained by vicariance events caused by continental breakups in the Mesozoic (Zink, Blackwell-Rago & Ronquist 2000; Donoghue & Moore 2003; de Queiroz 2005; Cowie & Holland 2006). This left long-distance dispersal as the major explanatory process, and this seems to be the most common assumption in historical biogeography analyses today, although there remain vocal proponents of vicariance biogeography (Ebach & Tangney 2007; Parenti & Ebach 2009; Heads 2012) and modern analyses using the most advanced dating methods sometimes do find groups old enough and with the proper distributions to be explained by Gondwanaland breakup (Wood *et al.* 2013).

Historical biogeography in recent years has come to be dominated by phylogenetic

biogeography (Maguire & Stigall 2008), wherein the evolution of geographic range

is traced on an explicitly hypothesized cladogram, or more recently, a time-

calibrated phylogeny. Phylogenetic biogeographers use several algorithmic, event-

based (Sanmartin 2007) inference methods, most prominently the computer

programs DIVA (Ronquist 1996; Ronquist 1997) and LAGRANGE (Ree *et al.* 2005;

Ree & Smith 2008; Ree & Sanmartín 2009; Smith 2010) and approaches derived

from them (Nylander *et al.* 2008; Yu, Harris & He 2010; Wood *et al.* 2013; Yu, Harris

& He 2013). In these methods, biogeography is simplified into a limited number of

discrete regions. Each OTU (operational taxonomic unit, typically a species or

monophyletic population) may be present or absent in each region. The goal is then

to find the best estimate of the history of lineages occupying and moving between

these discrete regions, according to some optimality criterion. DIVA is a parsimony

inference program that allows species to occupy one or more states, and searches

for biogeographic histories with minimum cost, where dispersal and extinction

events have a positive cost, while zero cost is assigned to vicariance events for

widespread ancestors, and to sympatric speciation events for ancestors occupying a

single area. LAGRANGE, probably the most popular method today, uses a DEC

(dispersal-extinction-cladogenesis) model that assigns probabilities to various

range-changing events. These event probabilities are used to calculate the

likelihood of the observed geographic range data at the tips of the tree. Maximum

likelihood (ML) optimization is then performed, and the values of the parameters

and ancestral states that confer maximum probability on the data constitute the ML

estimate. Additional advantages of LAGRANGE include the ability to specify

dispersal matrices that change in different time-strata, allowing researchers to

implement hypotheses of differing connectivity and distances between regions over

time. LAGRANGE was the origin of parametric, model-based biogeography (Ree &

Sanmartín 2009). Although DIVA and LAGRANGE are conceptually similar in many

ways, they make significantly different assumptions in the cladogenesis events

allowed (Ronquist & Sanmartin 2011; Webb & Ree 2012).

A variety of other methods are also in use, including the still-common practice of

simply treating biogeography as a standard multistate morphological character and

then employing parsimony, ML, or Bayesian approaches (Clark *et al.* 2008;

Sanmartin, Van der Mark & Ronquist 2008; Drummond *et al.* 2012) as well as

entirely new approaches that allow widespread species and take distances into

account using Approximate-Bayesian Computation (ABC) -like (Webb & Ree 2012)

or fully Bayesian (Landis *et al.* 2013) approaches. Finally, on top of the diversity of

computational approaches, decades-old, nonalgorithmic techniques such as

panbiogeography remain in use (Morrone 2006; Grehan & Schwartz 2009; Heads

2012), although not without substantial protest (Waters *et al.* 2013).

Historical biogeography is thus burdened with a variety of methods based on

conflicting assumptions, and no consensus on which are the best methods to use, or

even on how to determine what methods are the best. The main reason the issue

remains unresolved is the lack of any common statistical framework by which to

judge the competing models and assumptions that go into historical biogeography analyses. Some studies do compare different methods on the same datasets, and measure the amount of agreement and disagreement in ancestral range estimates made by different methods (e.g., Clark *et al.* 2008; Buerki *et al.* 2011; Diaz Gomez 2011). However, it is hard to reach any general conclusions about which methods and assumptions are the best, both because the method outputs are different (e.g. parsimony vs. likelihood estimates of ancestral range) and because any single clade being studied might, or might not, have been the product of the processes assumed in the inference method. When single clades are examined, and the different inference methods use different optimality criteria, there is no easy way to generalize about which biogeographic study systems are likely to follow or disobey the assumptions of the various inference methods.

BioGeoBEARS

In order to begin to address this problem, I created the package BioGeoBEARS (Matzke 2013b) in the R programming language (R Core Team 2013). "BioGeoBEARS" stands for <u>BioGeo</u>graphy with <u>B</u>ayesian (and likelihood) <u>E</u>volutionary <u>A</u>nalysis in <u>R</u> Scripts. The package allows probabilistic inference of both historical biogeography (ancestral geographic ranges on a phylogeny) and comparison of different models of range evolution. It reproduces the model available in LAGRANGE (Ree *et al.* 2005; Ree & Smith 2008), conducting ML optimization with optimx (Nash & Varadhan 2011; Nash & Varadhan 2012), but

makes available numerous additional models. For example, LAGRANGE as typically

run has two free parameters, d (dispersal rate, i.e. the rate of range addition along a

phylogenetic branch) and e (extinction rate, really the rate of local range loss along a

phylogenetic branch), and a fixed cladogenesis model that gives equal probability to

a number of allowed range inheritance events. Specifically, the allowed

cladogenesis events are (1) vicariance, (2) a new species starts in a subset of the

ancestral range (sympatric-subset speciation), and (3) the ancestral range is copied

to both species. In all cases, at least one descendant species must have a starting

range of size 1. LAGRANGE assigns equal probability to each of these events, and

zero probability to any other events. BioGeoBEARS allows the user to turn on an

additional cladogenic event, namely founder-event speciation, wherein the new

species jumps to a range outside of the ancestral range.

BioGeoBEARS also allows the relative weighting of the different sorts of

cladogenesis events to be made into free parameters, allowing use of optimization

and standard model choice procedures (e.g., AIC or the Likelihood Ratio Test;

Burnham & Anderson 2002) to pick the best model. The relative probability of

different descendent range sizes is also parameterized and thus can also be

specified or estimated, freeing researchers from the LAGRANGE assumption that

every cladogenesis event has at least one daughter species of range size 1. This is

important if, for example, the researcher wishes to implement and test a DIVA-like

model in a likelihood framework; the DIVA model assigns probability 0 to

sympatric-subset speciation, and assigns equal probability to all descendent range

sizes during a vicariance event. Thus, in DIVA, an ancestor with range ABCD may

split into descendant species with ranges AB and CD; but in the LAGRANGE DEC

model, this is disallowed (Ree & Smith 2008; Ronquist & Sanmartin 2011), and the

only vicariance scenarios allowed are those that split off a single area, e.g. (A, BCD).

This limitation of the LAGRANGE model can cause considerable difficulty when

researchers wish to test a classic vicariance hypothesis with modern parametric

methods, as Wood et al. (2013) discovered. However, in BioGeoBEARS, converting

the LAGRANGE DEC model into a DIVA-like model simply involves changing 2

parameters. The parameter *s*, which controls the relative probability of sympatric-

subset events, is changed from 1 to 0. The parameter *mx01v*, which controls the

relative probability of different descendent range sizes, is changed from 0.0001

(which forces the smaller daughter species to always have range size 1 during

vicariance, i.e. the LAGRANGE DEC assumption) to 0.5 (which gives equal

probability to all descendant range sizes during vicariance).

Model parameters

In a similar fashion, BioGeoBEARS enables access to many different models by

parameterizing each process or assumption made in programs such as DIVA or

LAGRANGE. These parameters may then be fixed to values that reproduce specific

models, or the parameter may be set to be free, and the researcher can test whether

or not the additional parameter contributes significant improvement to the data

likelihood. The parameters and brief descriptions of each are given in Table 1, along with the values that reproduce various popular models.

Additional parameters are included that allow new model features not available in any current package. For example, the parameter b exponentiates the branch lengths of the input phylogeny. If b is set to zero, then all the branch lengths are rescaled to have length 1, resulting in inference that will resemble that of a speciational, or parsimony, model (Pagel 1999b; Pagel 1999a). (This operation is only meaningful for non-time-stratified analyses.) Several parameters (Table 1) also allow users to implement a model for the imperfect detection of species presence in each region (Link & Barker 2009). This can occur either as a positive constraint on the possible ranges of a species, or, if count data are available for both detections of a species and detections of a taphonomic control (Bottjer & Jablonski 1988), as a probabilistic statement of the likelihood the count data given each possible range at the tip. These revised data likelihoods (much as with uncertain states in DNA; Felsenstein 2004) are then fed into the inference machinery.

Fossils

The flexibility available in BioGeoBEARS also enables the natural incorporation of fossil geographic range data. There are three ways to include fossils in BioGeoBEARS analyses. (1) If the researchers think they know the ancestral range(s) at a particular node, this can be hard-coded into the analysis via the

"fixnode" and "fixlike" parameters of the BioGeoBEARS run object. (2) The more likely situation is that the ancestral nodes on the tree will have unknown ranges, but the researcher will have fossil species with biogeographic range information. These can be included as fossil tips in the tree, as done by Wood et al. (2013). Unlike LAGRANGE, BioGeoBEARS explicitly implements likelihood calculations for both non-ultrametric trees that include fossil tips, for both time-stratified and non-time-stratified analyses. LAGRANGE accepts non-ultrametric trees and fossil tips appear to function well in non-time-stratified analyses (Wood *et al.* 2013), but including fossil tips in time-stratified analyses, as done in BioGeoBEARS, is somewhat more complex, as it cannot be assumed that all tips start in the time stratum closest to the present. (3) When fossils are anagenetic ancestors (for example, in cases where species-level mammalian phylogenies are available, complete with the stratigraphic time ranges of each species, e.g. Tedford, Wang & Taylor 2009), they may be included as very short side branches; BioGeoBEARS will treat any branches shorter than a user-specified cutoff length as direct ancestors, and will not apply the cladogenesis model at the node that links them to the rest of the phylogeny. This fixes the flaw in early attempts to include anagenetic ancestors in LAGRANGE (Matzke & Maguire 2011). As noted above, the use of taphonomic control groups to estimate sampling effort is implemented, and this may be used to modify the likelihoods at any tips or nodes. Although a different context (estimating lineage diversity patterns in space and time), Valentine et al. (2013) provide a dramatic demonstration of how severe mistakes in inference can be caused by failure to account for biases in fossil sampling.

Additional features

BioGeoBEARS implements a number of additional features that increase analysis

flexibility and researchers' ability to propose and test hypotheses. While

LAGRANGE allows users to manually specify a dispersal matrix representing the

relative probability of dispersal between each region, and allows different dispersal

matrices in different time strata, this approach has some limitations. Setting these

relative dispersal probabilities is often a "seat-of-the-pants" operation on the part of

the researcher – for example, who is to say whether trans-oceanic dispersal is 100

times less likely than intracontinental dispersal, or 1000 times less likely? Ideally,

such probabilities would be controlled by free parameters in an explicit model. The

values of the parameters may then be inferred with maximum likelihood (ML) or

Bayesian approaches. BioGeoBEARS enables several parameterizations, for example

the user can make dispersal probabilities a function of distance to some power, x

(Webb & Ree 2012; Landis *et al.* 2013). If x is negative, then long-distance dispersal

has a lower probability than short-distance dispersal. If x is 0, then distance has no

effect. Users can combine this with a traditional dispersal matrix that is fixed to

specifying connectivity or other geographical and environmental features thought to

influence dispersal probabilities.

Another limitation of the dispersal matrix approach is that it is not identical with

specifying the disappearance of areas in the past. For example, each of the various

Hawaiian Islands emerged at a particular time point, and in LAGRANGE, the traditional approach is to set the dispersal rates to nonexistent islands to 0. However, this still formally allows the possibility of long-term persistence on an island, which can cause unexpected results in the likelihood calculations. Therefore, BioGeoBEARS allows users to include an "areas allowed" matrix for every time stratum. This even allows the inference ancestral species existing in ranges that are not found in the present (as is expected for some older clades in the Hawaiian islands; e.g., Jordan, Simon & Polhemus 2003; Lapoint, O'Grady & Whiteman 2013).

Given the bewildering variety of historical biogeography packages and their features, it is useful to have a systematic summary of the features of each. In Table 2, I compare BioGeoBEARS to the most popular and/or most recent historical biogeography software available.

Number of areas

An extremely important consideration in all discrete historical biogeography analyses is the number of areas, and the maximum number of areas that any species may occupy at a given time. This is because the size, s, of the state space under a maximum range size constraint m, is the sum of number of possible combinations of ranges of size 1, 2,…, m, or

$$s = \sum_{k=1}^{m} \binom{N}{k} \qquad \text{(eqn 1)}$$

(This equation is implemented for users in the cladoRcpp function

numstates_from_numareas.)

This means that for 5 areas, there are 2^5=32 possible states, where each state is a

possible geographic range (a combination of presences and absences in the study

area). But for 10 areas, there are 1024. This means that the rate matrix used to

calculate the transition probabilities along the branches is of size 1024x1024.

Exponentiation of large matrices becomes extremely slow. Cladogenesis calculations

can also become slow (Ronquist 1997) because, formally, the probability of every

possible combination of (ancestral state, left descendant state, right descendant

state) must be assessed. For 10 areas, this is 1024^3=1,073,741,824 combinations.

C++ LAGRANGE and BioGeoBEARS make use of the FORTRAN library EXPOKIT

(Sidje 1998) to speed up the exponentiation of large matrices. While EXPOKIT

enables much more rapid calculations than available in Python LAGRANGE, the

exponential increase in matrix size with number of areas means that C++

LAGRANGE and BioGeoBEARS only have modestly higher limits on the number of

areas. Both C++ LAGRANGE and BioGeoBEARS also implement the optional use of

the sparse matrix exponentiation routines from EXPOKIT. In theory, sparse matrix

exponentiation should allow the rapid exponentiation of larger matrices and thus

use of more areas. However, the accuracy of the results of sparse matrix

exponentiation is lower, and when analyses are run with and without sparse matrix exponentiation in BioGeoBEARS, the exponentiation results are correlated but not identical, apparently resulting in different ML inferences (NJM, unpublished data). Thus, caution is recommended in using the sparse matrix exponentiation routines.

Accessory packages

BioGeoBEARS accesses EXPOKIT via an accessory R package, rexpokit (Matzke & Sidje 2013) which contains the FORTRAN and C++ source code, linked to R with Rcpp (Eddelbuettel & Francois 2013; Eddelbuettel, Francois & Bates 2013). Compilation from source can be a complex affair, depending on one's operating system and installed compilers, so most users should simply install the binary of rexpokit, available from CRAN.

The cladogenesis calculations that BioGeoBEARS requires would be extremely slow in an R-only implementation, as a great many for-loops are required to check whether various combinations of ancestral and descendant states are allowed events during cladogenesis. Therefore, these were implemented in C++ using Rcpp, in the accessory R package cladoRcpp (Matzke 2013c). cladoRcpp includes a number of tricks to minimize the combinatorial explosion of cladogenesis events, as a great many theoretically possible events can be excluded as having probability zero. The easiest way to install cladoRcpp is also via the binary at CRAN.

Table 2 lists approximate computation times, in order to give users a sense of what is possible, given their research timeline, computational resources, and biogeographical study system. These should be treated as rough estimates of the amount of computational time required, not formal time trials. They are based on several years' experience working with these programs, and in the case of SHIBA, the published description of the runtime (Webb & Ree 2012). As is commonly observed, the great flexibility and interpretability of R, which enabled the implementation of the diverse feature suite of BioGeoBEARS, also comes at the cost of slow run times. Thus, for analyses with small numbers of areas, BioGeoBEARS is as slow or slower than Python LAGRANGE. However, as the number of areas increases, the computation time comes to be dominated by the time required for matrix exponentiation. Thus, for large numbers of areas, BioGeoBEARS approaches C++ LAGRANGE in terms of computation speed.

Computation times in BioGeoBEARS can be improved by use of parallel processing on multiple cores. If the user specifies more than one core, then the matrix exponentiation calculations are parallelized via R's parallel package, mclapply function. Note that the parallel functions are not available in R.app, so users with Macintoshes should run large analyses with R started from the Terminal command line, rather than R.app.

Bayesian analyses

This introduction to BioGeoBEARS has focused on the models available, and likelihood implementations of them. Bayesian analysis has been implemented through use of the LaplacesDemon package (Statisticat 2013b), however, that package is now maintained off of CRAN, so its usage is not formally included in BioGeoBEARS at the current time. Essentially all that is required is for the user to specify priors on parameters of interest for LaplacesDemon, and then use BioGeoBEARS to provide the data likelihoods. LaplacesDemon makes available dozens of flavors of Bayesian MCMC sampling once these inputs are available. Example functions making use of LaplacesDemon may be found in the BioGeoBEARS extdata/a_scripts directory.

Utility and plotting functions

BioGeoBEARS also contains a large number of utility functions for manipulating time-scaled phylogenies (e.g., prt prints an R phylo3 object to an interpretable table format), geographic range data, and the outputs from other biogeography inference programs. Biogeographers may have fond memories of puzzling over the text output of programs like DIVA and LAGRANGE, which produce their primary output in the form of a list of nodes and ancestral states. Frustratingly, DIVA, Python LAGRANGE, C++ LAGRANGE, and R's ape package (Paradis 2012) each use a *different* numbering scheme for internal nodes. BioGeoBEARS has functions for reading these outputs and plotting them on phylogenies; see e.g. get_lagrange_nodenums and related functions.

cladoRcpp contains the function numstates_from_numareas, which implements

Equation 1, above, and which is key for planning runs with large numbers of

areas/states.

Example and Conclusion

An example script that conducts a BioGeoBEARS analysis "out of the box" has been

provided in supplementary material. This example runs the default LAGRANGE-

style DEC analysis on the Hawaiian *Psychotria* dataset used as the initial example for

LAGRANGE (Ree & Smith 2008) and available at the LAGRANGE Configurator (Ree

2013). The ML parameter inferences and log-likelihood are identical to those found

by LAGRANGE. Ancestral state probabilities may differ, due to the different defaults

for the type of ancestral state inference run in LAGRANGE versus BioGeoBEARS

(Table 2; BioGeoBEARS by default estimates ancestral states under the global ML

model, which is faster than local optimization, and may be more accurate; Mooers

2004). The script also runs a DEC+J model and compares the results with the

likelihood ratio test and AIC. The likelihood ratio test rejects the hypothesis that

DEC and DEC+J confer equal likelihoods on the data (p=1.80E-07), and relative AIC

model weights indicate that the DEC+J model is 294895 more probable than DEC.

A comparison of the ancestral state inferences under the two models is shown in

Figure 1. Under the DEC+J model, the parameter estimates for *d* and *e* drop to 0. In

16

essence, this means that founder-event speciation/jump dispersal process is doing all of the work in moving *Psychotria* lineages around the Hawaiian Islands. Figure 1 also shows that the DEC+J model results in less uncertainty in the estimation of ancestral nodes, and gives a much simpler picture of the events that led to modern distributions, essentially a small number of founder-events between islands, coupled with common within-island speciation.

In conclusion, BioGeoBEARS shows great potential for enhancing the accessibility of parametric, probabilistic model-based methods in historical biogeography (Ree & Sanmartín 2009), and for testing these same methods and models. Undoubtedly, users will discover additional useful models with novel combinations of free and fixed parameters. In this sense, the situation may resemble that found in the evolution of the simple Jukes-Cantor model of DNA substitution into the complex GTR+I+G model (Posada & Crandall 1998). Hopefully, the model-testing framework implemented in BioGeoBEARS will help to resolve some of the long-standing debates over models and starting assumptions in historical biogeography.

Conflicts of Interest

None.

Figure 1. Ancestral range inferences under the maximum likelihood estimate on the Hawaiian shrub clade *Psychotria* (Ree & Smith 2008) using a BioGeoBEARS implementation of the LAGRANGE DEC (dispersal-extinction-cladogenesis) model (top) and BioGeoBEARS DEC+J model. The DEC model has 2 parameters (*d*, *e*), and the DEC+J model adds a third free parameter (*j*) controlling the relative probability of founder-event speciation events at cladogenesis, versus the traditional LAGRANGE DEC cladogenesis events. The most probable states are shown at left; note that these are the most probable states at each node and corner (corners represent the states just after speciation); they do not necessarily correspond exactly to what a joint reconstruction of the single best joint history would show. The probability of each possible ancestral range at each node, under the ML-estimated parameters for each model, is shown at right. The DEC+J model shows much less uncertainty in ancestral states, favors ancestors with single-island ranges, and confirms the progression of taxa from the oldest island (Kauai) to the younger islands. 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.

BioGeoBEARS DEC on Psychotria M0
anstates: global optim, 4 areas max. d=0.035; e=0.028; j=0; LnL=-34.5

BioGeoBEARS DEC on Psychotria M0
anstates: global optim, 4 areas max. d=0.035; e=0.028; j=0; LnL=-34.5

BioGeoBEARS DEC+J on Psychotria M0
anstates: global optim, 4 areas max. d=0; e=0; j=0.114; LnL=-20.9

BioGeoBEARS DEC+J on Psychotria M0
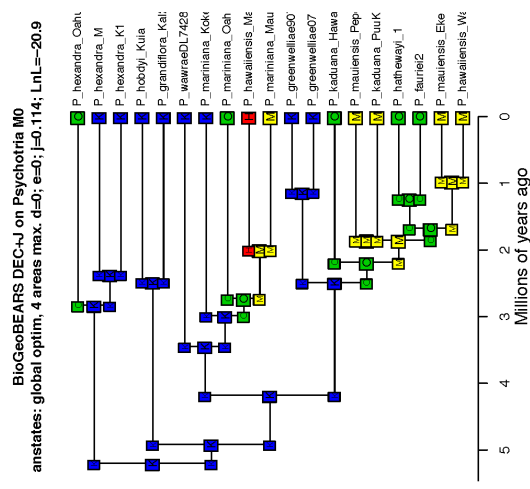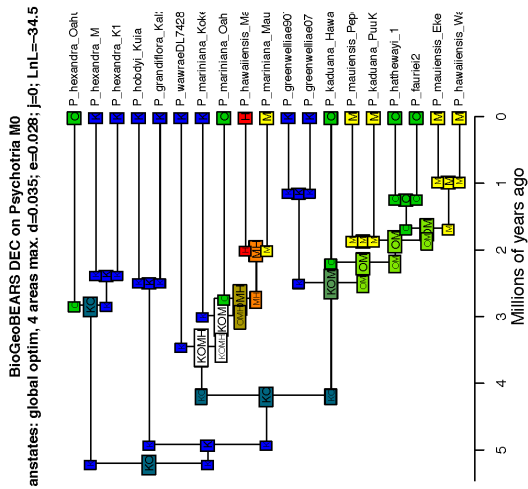anstates: global optim, 4 areas max. d=0; e=0; j=0.114; LnL=-20.9
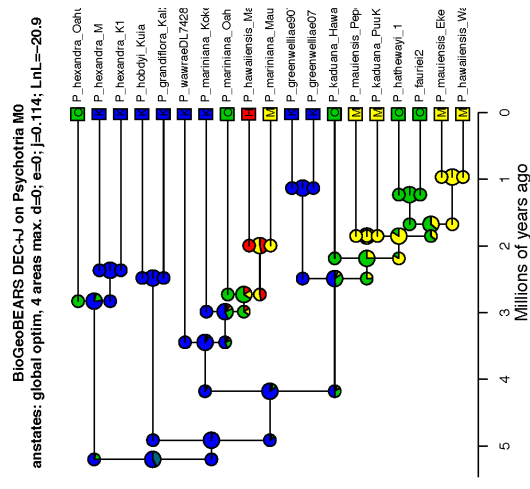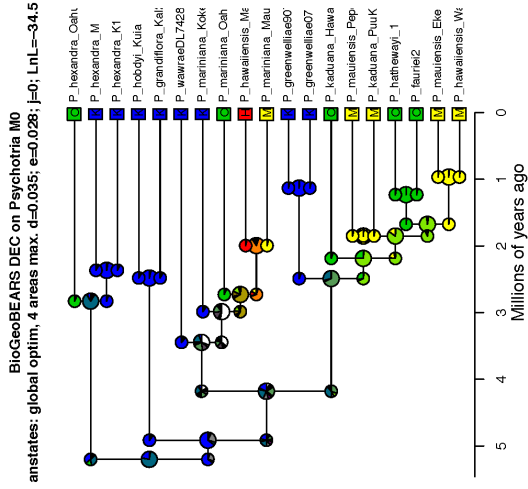
Table 1. Parameters that are customizable in BioGeoBEARS, with brief descriptions. Example parameterizations are shown that reproduce (1) LAGRANGE DEC, (2) DEC+founder-event speciation (DEC+J), (3) a likelihood version of a DIVA model, (4) a likelihood version of the BAYAREA model, and (5) a standard character evolution model. For details, see main text and BioGeoBEARS documentation.

	Model					
Parameter	DEC	DEC+J	DIVA-like	BAYAREA-like	Standard	Description
d	free	free	free	free	0	anagenesis: rate of 'dispersal' (range expansion)
e	free	free	free	free	0	anagenesis: rate of 'extinction' (range contraction)
a	0	0	0	0	free	anagenesis: rate of instantaneous range-switching (i.e. for a standard character)
b	1	1	0 or 1	1	1	anagenesis: exponent on branch lengths (set to 0 to approximate parsimony)
x	0	0	0	free	0	anagenesis: exponent on distance (modifies *d, j, a*)
u	0	0	0	0	0	anagenesis: exponent on extinction risk with area (modifies *e*)
j	0	free	0	0	0	cladogenesis: relative per-event weight of jump dispersal
y	1	(3-*j*)/3	1	1	1	cladogenesis: relative per-event weight of sympatric (range-copying) speciation
s	1	(3-*j*)/3	0	0	1	cladogenesis: relative per-event weight of sympatric (subset) speciation
v	1	(3-*j*)/3	1	0	0	cladogenesis: relative per-event weight of vicariant speciation
mx01	0.0001	0.0001	0.0001	NA	NA	cladogenesis: controls range size of smaller daughter (all)
mx01j	0.0001	0.0001	0.0001	NA	NA	cladogenesis: controls range size of smaller daughter (jump speciation)
mx01y	0.0001	0.0001	0.0001	0.9999	NA	cladogenesis: controls allowed range size of both daughters (sympatric-range copying)
mx01s	0.0001	0.0001	0.0001	NA	NA	cladogenesis: controls range size of smaller daughter (sympatric-subset)
mx01v	0.0001	0.0001	0.0001	NA	NA	cladogenesis: controls range size of smaller daughter (vicariance)
mf	1	1	1	1	1	detection model: mean frequency of truly sampling OTU of interest
dp	1	1	1	1	1	detection model: detection probability per true sample of OTU of interest
fdp	0	0	0	0	0	detection model: detection of OTU probability per true taphonomic control sample

Table 2. The features implemented in BioGeoBEARS, compared to the features currently available in previous phylogenetic inference software in historical biogeography.

Feature	Feature	Standard	DIVA	LAGRANGE (Python)	LAGRANGE (C++)	SHIBA	BayArea	BioGeoBEARS
					Software			
General	Inference method	parsimony, ML, Bayesian	parsimony	ML[1]	ML[1]	ABC-like	Bayesian	ML, Bayesian
	Geographic range concept	discrete	discrete	discrete	discrete	discrete	discrete	discrete
	Number of areas each species/lineage may inhabit	1	many	many	many	many	many	many
	Anagenetic range evolution	instantaneous range switching (a)	range expansion ("dispersal", d) range contraction ("extinction", e)[2]	d, e	d, e	d, e	d, e	d, e, a
Range evolution models / parameters	Cladogenetic range evolution	none[3]	sympatry, vicariance (fixed)	sympatry, sympatric subset, vicariance (fixed)	sympatry, sympatric subset, vicariance (fixed)	sympatry, sympatric subset, vicariance (fixed)	none[4]	sympatry, sympatric subset, vicariance, jump (all free or fixed)
	Includes founder-event speciation	no[3]	no	no	no	no	no	yes
	Range size of smaller daughter (non-vicariant events)	NA	1	1	1	1	NA	default 1, but parameterized
	Range size of smaller daughter (vicariant events)	NA	any	1	1	1	NA	default 1, but parameterized
	Optional branch-length scaling	no	NA	no	no	no	no	yes (non-time-stratified analysis only)
	Dispersal as function of distance matrix	no	no	no	no	yes	yes	yes
	Extinction as function of area size	no	no	no	no	yes	no	yes
	Model for imperfect detection	no	no	no	no	no	no	yes
Constraints	Time-stratification	no	no	yes	yes	yes	no	yes
	User-specified dispersal matrix	yes	no	yes	yes	yes	no	yes
	User-specified constraints on allowed area combinations	NA	no (check)	yes	yes	yes	no	yes
	User-specified dispersal matrix (time-stratified)	no	no	yes	yes	yes	no	yes
	User-specified constraints on allowed area combinations (time-stratified)	NA	no	no	no	yes	no	yes
Fossils	Fossils as tips (non-time-stratified analysis)	yes	yes	yes	yes	?	yes	yes
	Fossils as tips (time-stratified analysis)	no	no	no	no	?	no	yes
	Fossils as anagenetic ancestors (non-time-stratified analysis)	no[5]	no	no	no	no	no[5]	yes
	Fossils as anagenetic ancestors (time-stratified analysis)	no	no	no	no	no	no	yes
	Local optimization at each node	various	no	yes	yes	no	no	optional
	Probabilities at each node according to the globally optimum model	various	yes (parsimony)	no	no	yes (ABC)	yes (MCMC)	yes (default)

Ancestral state estimation	Joint ancestral states estimate -- i.e. single most probable history	no	no	no	no	no	no
	Plot ancestral states at nodes	various	no	yes[6]	no	?	yes
	Plot ancestral splits at phylogeny corners	no	no	no	no	no	yes
	Plot pie charts of state probabilities	no	no	no	no	no	yes
	Plot pie charts of split probabilities at phylogeny corners	no	no	no	no	no	yes
Limitations	Maximum # of areas (with maximum rangesize = 2)	max. # of states in program[7]	15 (hard-coded)	~15 (121 states)	~31 (497 states)	~5	~31 (497 states)
	Maximum # of areas (with maximum rangesize = # of areas)	max. # of states in program[7]	15 (hard-coded)	~7 (128 states)	~9 (512 states)	~5	~9 (512 states)
	Approximate computation time (small analysis, 5 areas)	<1 sec	<1 sec	<1 min	<1 sec	hours	<1 min
	Approximate computation (large analysis, 9 areas)	seconds	minutes	weeks or impossible	hours	impossible	hours

Numbered Notes:

1. "Bayesian" versions of LAGRANGE and DIVA are available via the procedure of running these programs on a large posterior sample of phylogenies from a Bayesian analysis. See main text.

2. DIVA technically allows extinction/range contraction events with cost 1, but because vicariance events have cost 0, extinction events will never be inferred by DIVA unless some non-default constraints or costs are implemented (Sanmartin ref).

3. In a sense, the instantaneous range-switching model including founder-event/jump speciation, if all branch lengths in a tree are set to length 1. In that situation, the probability of range-change events is proportional to the number of speciation events, rather than the branch length.

4. In BayArea, at each cladogenesis event, the ancestral range is simply copied to the descendant range. I.e., if the ancestor is ABC, the left and right descendant species will inherit this range: (left, right)=(ABC, ABC). In effect, speciation events are no different than any other part of the tree. This is just like any other continuous-time markov process, i.e., those used for DNA or amino acid substitution models. It is important to note that, biogeographically, this "no range evolution at cladogenesis" assumption actually is implementing a de facto cladogenesis model, namely complete copying of the ancestral range to the descendants at each speciation event. This is a very strong assumption.

5. In non-stratified analyses, attaching a very short branch to a phylogeny is a way of placing a "direct ancestor" on the tree. This is only legitimate if a continuous-time markov model of range evolution is being assumed; otherwise, the speciation model must be turned off for such ancestor tips (as is done in BioGeoBEARS).

6. C++ LAGRANGE produces output NEXUS trees with the names of the ML states or splits at each ancestral node attached as branch labels. This can then be loaded in a standard tree viewing package (e.g., FigTree).

7. The number of areas allowed in a model where biogeography is treated as a single standard character is equal to the maximum number of character states allowed by a program.

General Notes:

Programs/models not covered here include:

account lineage-splitting and lineage-loss events that are not recorded in the available tree.

ClaSSE, which implements a GeoSSE-like model for 3 areas.

Chapter 2: Formal Model Testing of the Dispersal-Extinction-Cladogenesis (DEC) Model Reveals that Founder-event Speciation is a Dominant Process Structuring the Biogeography of Island Clades

Abstract

Probabilistic modeling of geographic range evolution was a major advance in historical biogeography, making biogeographical problems accessible to model-based maximum likelihood (ML) and Bayesian methodologies. The most popular model is Dispersal-Extinction-Cladogenesis (DEC), implemented in the software LAGRANGE (Ree & Smith 2008). Standard DEC is a model with two free parameters specifying the rate of "dispersal" (range expansion) and "extinction" (range contraction). However, while dispersal and extinction rates are free parameters, the cladogenesis model is fixed, such that the geographic range of the ancestral lineage is inherited by the two daughter lineages through a variety of scenarios fixed to have equal probability. This fixed nature of the cladogenesis model means that it has been indiscriminately applied in all DEC analyses, and has not been subjected to any inference or formal model testing. I re-implement DEC in the R package BioGeoBEARS, which exactly reproduces LAGRANGE DEC parameter inferences and likelihoods. However, BioGeoBEARS also allows additional parameters controlling the probability of new cladogenesis processes. The example shown here is "founder-event speciation", in which one daughter jumps to an area completely outside the ancestral range. The new model, termed DEC+J, is tested and verified by simulation, and is then applied to 13 island clades under a variety of constraint scenarios. Standard statistical model comparisons show that DEC+J is vastly superior to

standard DEC for all clades, for the first time verifying the importance of founder-event speciation in island clades via statistical model choice, and falsifying vicariance-dominated models of island clades. Furthermore, the inferred ML estimates of parameters often differ radically under the DEC+J model, with the "DE" part of the model sometimes playing no role. Under DEC+J, ancestral nodes are usually estimated to have ranges occupying only one island, rather than the widespread ancestors often favored by DEC, and simulation results indicate that when DEC+J is the true model, DEC+J has 87% accuracy in inferring ancestral states, while DEC has only 57% accuracy. These results indicate that careful consideration of cladogenesis models in historical biogeography of island clades is recommended.

Introduction

Probabilistic modeling of geographic range evolution Ree *et al.* 2005; Ree & Smith 2008; Ree & Sanmartín 2009) was a major advance in historical biogeography, making biogeographical problems accessible to model-based maximum likelihood and Bayesian inference methodologies (reviewed by Ronquist & Sanmartin 2011; Sanmartín 2012), where previously only parsimony approaches (e.g., Ronquist 1997) had been available. The most popular model is Dispersal-Extinction-Cladogenesis (DEC), implemented in the software LAGRANGE in Python (Ree 2013) or C++ (Smith & Ree 2010). LAGRANGE has been has been widely used in the years since publication, with Ree et al. (2005) and Ree and Smith (2008) attracting a total of 566 citations to date (Google Scholar search, 8/4/2013).

DEC is an example of an "event-based" inference method (Ronquist 1996; Sanmartin 2007; Kodandaramaiah 2010), and on that basis may be distinguished from the area-based or pattern-based methods that were dominant in historical biogeography into the 1990s (Nelson & Rosen 1981; Cracraft & Shipp 1988; Page 1988; Crisci 2001). All such methods abstract geography into a limited number of discrete areas. At any point in time, a particular lineage will be present or absent in each of the areas, and this list of presences and absences constitutes the "geographic range" of the lineage at that point in time. Using this discrete-areas simplification (and sometimes it is a dramatic simplification, and should always be employed with caution), the problem then becomes how to make the best estimate of the history of

geographic range evolution down the phylogeny. This may be done by assuming some optimality criterion and some model for the evolution of geographic range. In a parsimony framework, the optimum history is the one with the minimum number steps (as in the parsimony-based DIVA; Ronquist 1996; Ronquist 1997) or the model parameters that confer the maximum likelihood on the geographic range data at the tips of the tree (typically the ranges of observed, living species), given the phylogeny and the model (as in LAGRANGE's DEC).

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). Dispersal and extinction are treated as continuous-time Markov processes, where the probability of change is a function of the amount of time that has passed along a branch.

DEC's cladogenesis model

Under DEC, geographic range is also allowed to change at cladogenesis events (the nodes on the phylogeny), with the geographic range of the ancestral lineage inherited by the two daughter lineages through a variety of equiprobable scenarios. These scenarios are (1) 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 daughter: A, right daughter: A). The second sympatric speciation scenario

occurs when an ancestor is widespread, inhabiting multiple areas. Here, under DEC, inheritance of the complete 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. The intuition behind this type of cladogenesis event is that new species often form by budding off from widespread ancestors, either because they colonize some new habitat (such as a mountaintop) or new ecological niche. The assumption that the new species must have a range size of one area is justified on the basis that speciation usually requires isolation of the gene pool, and this is most likely in a restricted subpopulation with a limited range. In this article, scenario #2 is termed "sympatric-subset" speciation, and #1 is termed "sympatric-range copying". It should be noted that "sympatry" vs. allopatry in the historical biogeography context is relative to the discrete areas chosen for the analysis. Within-range speciation is termed sympatric as the daughter lineages have overlapping ranges; however, this may have little or nothing to do with the true detailed process of speciation, which might still be allopatric, but at a smaller scale.

The third cladogenesis scenario allowed in the DEC model is vicariance, wherein a widespread ancestral range is divided up between the two daughter species. It is important to note that in the LAGRANGE DEC model, the assumption that one of the daughters must have a range of size of 1 area is kept for the vicariance process. Thus, an ancestor with range ABCD may divide into daughters with ranges (A, BCD), (C, ABD), etc., but the daughter ranges are not allowed to inherit the ranges (AB,

28

CD), (AD, BC), etc. DIVA makes a different assumption, allowing all of these vicariance scenarios; however, DIVA disallows sympatric-subset speciation (Ronquist & Sanmartin 2011). Comparison of the DEC and DIVA assumptions about cladogenesis highlights the fact that whatever cladogenesis model is implemented in an inference program is not necessarily the only choice theoretically available, even though it might be the only choice practically available to researchers.

Most researchers who use LAGRANGE DEC have focused on the inference of ranges at ancestral nodes, and on the estimation of the anagenetic parameters governing the rates of dispersal and extinction (parameters d and e). These are usually treated as free parameters to be optimized, sometimes under user-specified constraints based on distance or connectivity (e.g., Clark, Wagner & Roalson 2009; Webb & Ree 2012). The cladogenesis model often receives less attention, probably because it is a fixed, hard-coded model, and is therefore automatically part of any LAGRANGE analysis. The fact that the cladogenesis model is fixed in available software means that it has been applied indiscriminately, whether or not it is appropriate for the biology and geography of the group in question. Having a fixed model also prohibits model comparison, giving users no capability to detect when the cladogenesis model is appropriate or inappropriate. DEC might well be the correct model, or at least a wrong but useful model (Box & Draper 1987), but researchers will unable to check these assumptions unless additional plausible models are constructed and statistically compared to DEC.

Founder-event speciation

One well-known form of speciation that is left out of standard DEC analyses is

"founder-event speciation" (Paulay & Meyer 2002; Templeton 2008), sometimes

termed founder-event speciation through long-distance dispersal (Heads 2012) or

allopatric mode II speciation (Wiley 1981; Maguire & Stigall 2008; Lomolino,

Lomolino & Lomolino 2010). (Allopatric mode I speciation is vicariance.) 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 from a

population geneticists, going back to Mayr's "genetic revolution" theory for

peripheral isolate populations (Mayr 1954; Gould & Eldredge 1972), and it remains

a controversial question whether or not the population-genetic founder effect and

resulting genetic bottlenecks commonly contribute significantly to shifting of

adaptive peaks (Carson & Templeton 1984; Coyne & Orr 2004) or to other genetic

and morphological changes in evolutionary history. However, to most

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. The importance of founder-event speciation for explaining

certain biogeographical patterns remains true whether or not founder-event

speciation plays any major role in production of non-biogeographical

macroevolutionary patterns and processes (although there is some evidence that it

does; Carson & Templeton 1984; Moore & Donoghue 2007). Founder-event

speciation is particularly likely to be important in oceanic island systems (de

Queiroz 2005; Cowie & Holland 2006), where much evidence supports the

importance of rare dispersal events for structuring biogeography -- for example, the

high prevalence of self-fertilizing taxa on islands (Carlquist 1974), the progression

rule of Hennig (1966), or the strong relationships between organismal mode of

dispersal and the biogeographic origin of taxa on oceanic islands (Paulay & Meyer

2002; Gillespie *et al.* 2012).

Figure 1. Various models for the evolution of geographic range at cladogenesis events. The events allowed by LAGRANGE are highlighted in gray. Each allowed event is fixed to have equal probability in the LAGRANGE algorithm. LAGRANGE requires that, at cladogenesis, at least one daughter species must have a range size of 1. This means that, while some vicariance events are allowed, others are not (row 4, columns 3). Widespread sympatry is also disallowed (column 1, rows 2-4) and sympatric-subset speciation where the subset species has a range size larger than one area. Note: The trees shown are just examples of each type of event; an exhaustive enumeration of all of the possibilities would require a much larger figure: for a four-area ancestor, if the one-area restriction is maintained for one of the daughters, there are eight possible vicariance events, and eight possible sympatric-subset events. (The number is eight, not four, since the left and right daughters are not exchangeable.) If the one-area restriction is lifted, there are 28 possible vicariance events, and 28 possible sympatric-subset events, for the four-area ancestor. Founder-event speciation, with the dispersing lineage initially restricted to a single area, is implemented in this paper (column 4).

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. Part of the reason that the process's

absence has not been addressed until now may be that some think that founder-

event speciation is covered by the "dispersal" process of *Dispersal*-Extinction-Cladogensis (DEC) and *Dispersal*-Vicariance Analysis (DIVA). However, as noted above, these concepts of dispersal are anagenetic range-expansion events (e.g., A → AB or ABD → ABCD) and are not identical with founder-event speciation, in which, by definition, dispersal occurs simultaneously with lineage-splitting.

Figure 1 shows examples of the types of cladogenesis events allowed (grey highlighting) and disallowed (white) in the DEC model. From the figure, it can be seen that, computationally, implementing the founder-event process in an algorithm is not much more difficult than implementing any other cladogenesis process. However, it does require "thinking outside the box" in a fairly literal way, in that daughter species may jump to areas outside of the geographic range of their ancestors.

Some perceptive commentators (Ree & Sanmartín 2009; Kodandaramaiah 2010; Goldberg, Lancaster & Ree 2011) have in fact noted the absence of founder-event speciation from DIVA and DEC, and suggested that it should be taken into consideration in historical biogeography, at least in island systems. Here, I take their advice. The DEC model and ML inference procedure of LAGRANGE is re-implemented in the R package BioGeoBEARS (Matzke 2013b), and then the DEC model is extended by adding founder-event speciation, creating a DEC+J model. The 2-parameter DEC model is nested within the 3-parameter DEC+J model, and thus the likelihood ratio test (LRT; Burnham & Anderson 2002) can be used to compare

the models in a frequentist framework. Standard model selection procedures based on Akaike Information Criterion (AIC) are also used to compare models. Finally, a Bayesian approach to model comparison is also implemented.

These model tests are first run on the Hawaiian *Psychotria*, a clade of understory trees with an endemic subclade spread across the four main high islands. This is the same dataset used as the example in the original LAGRANGE paper (Ree & Smith 2008, derived from Nepokroeff *et al.* 2003) and available as the default example with the Python version of LAGRANGE (Ree 2013). All inference procedures yield strong support for DEC+J over DEC on the *Psychotria* dataset, despite the small size of the dataset. As a test of the sensitivity and specificity of the inference procedure and the distinguishability of the models, simulations are conducted under the DEC and DEC+J models. *Psychotria* is used as a typical small phylogenetic biogeography dataset, and the ML-estimated parameters from the DEC and DEC+J runs on *Psychotria* were used to produce 1000 simulations under each model. These simulated datasets were then each subjected to inference under DEC and DEC+J, and the results were compared with the LRT, allowing the measurement of the rate of false positives and false negatives under the LRT.

All of these analyses supported the identifiability and superiority of the DEC+J model on the *Psychotria* clade. Therefore the DEC and DEC+J models were compared on a large sample of phylogenetic biogeography datasets from published studies of island clades. Remarkably, DEC+J dominates in every case. In many cases, the DEC+J

model results in much different inferences of the anagenetic dispersal and
extinction rates (usually moving them closer, or all the way, to zero), suggesting that
jump dispersal is the dominant process structuring biogeography in these clades.
The inference of highest-probability ancestral states is also often changed, with
widespread ancestors inferred under DEC disappearing, and being replaced with
single-island ancestors. Finally, some conclusions are drawn about the possibility of
generalizing these conclusions to non-island clades, and about the possibilities that
the model-comparison approach opens up for model-based, parametric
biogeography (Ree & Sanmartín 2009).

Methods

Implementation of DEC model

In a DEC analysis, the investigator divides the geography into a small number of
discrete regions, such as islands, ecoregions, or continents. The OTUs (operational
taxonomic units) at the tips of the phylogeny are then scored for presence or
absence in each of these regions. The user may then impose constraints, such
dispersal limits, or a maximum range size. The latter is useful for limiting the
number of possible geographic ranges, and thus states, in the transition rate matrix.
Without constraints on the maximum size of ranges, the number of states rises
exponentially at $m=2^n$, where n equals the number of discrete areas in the analysis,
and m represents the number of possible geographic ranges, i.e. the number of

possible combinations of presence and absence in each area. In effect this limits the complexity of the geographic discretization to less than 10 areas in an analysis with no maximum range size, or about 20 areas with a maximum range size of 2 or 3 (Webb & Ree 2012).

DEC as typically used in LAGRANGE 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 & Smith 2008) which is exponentiated (Moler & Van Loan 2003) to calculate the probabilities of each state at the bottom of a branch, as a function of the branch length (typically in units of millions of years, Myr) and the probabilities of the states at the top of the branch. In most phylogenetic methods, when two branches meet at an ancestral node, the probabilities at the two branch bottoms are simply multiplied to produce data likelihoods at the ancestral node; this is the basis of the Felsenstein pruning algorithm (Felsenstein 1981; Felsenstein 2004). However, in applications where states change "instantaneously" at cladogenesis events, the downpass probabilities must be combined in a more complex way, via a cladogenesis model.

This cladogenesis model is essentially a three-dimensional rate matrix (Ronquist 1997) where the dimensions correspond to the states in the ancestor, the states in the left descendent daughter species, and the states in the right descendent daughter species, and the "time" is a discrete unit (a single cladogenesis event). However, it is easier to visualize what is going on by flattening the transition matrix

to two dimensions, 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.

In the DEC cladogenesis model, conditional on a particular ancestral geographic range, each possible cladogenesis event (Figure 1) has equal weight and therefore equal probability. This has the advantage of allowing rapid calculation, but the disadvantages (noted above) of assuming a fixed cladogenesis model.

Implementation of the DEC+J model

Examples of cladogenesis models for a three-area analysis are shown in Tables 1A and 1B, for DEC and DEC+J, respectively. It can be seen from Table 1 that implementing the DEC+J model is primarily a matter of assigning a parameter, *j*, to jump dispersal events in the cladogenesis matrix. Note that the parameter *j* represents the weight assigned to each individual jump dispersal event, not the total weight assigned to all jump dispersal events collectively. To calculate the probabilities of a particular daughter pair of ranges, conditional on an ancestral range, all that is required is that the weights of all of the allowed cladogenesis events from a particular ancestral range be summed; the per-event probability is just the per-event weight divided by the sum of the weights. It is imaginable that an alternative way to parameterize the cladogenesis model would be to assign parameters controlling the relative probability each class of cladogenesis event;

however, because the number of each type of cladogenesis event will vary for each ancestral range size (Table 1), and will vary further if users impose any constraints on allowed ranges, this would likely be difficult. Therefore, the strategy of assigning each individual event a weight, and then calculating the per-event probabilities by dividing by the sum of the weights, is the simplest and most generalizable calculation strategy.

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 spreadsheet in the Supplementary Data 1. (Beyond 4 areas, graphical depiction of the cladogenesis matrix becomes impractical: e.g., for 5 areas, the full matrix has $2^5-1=31$ rows for possible ancestral states, and $31^2=961$ columns for possible descendent daughter pairs.) The spreadsheet includes a worksheet that allows readers to vary the j weight parameter and see the resulting changes in the calculation of the probabilities of both founder and non-founder cladogenesis events.

BioGeoBEARS parameterizes the per-event weights of each type of cladogenesis event (Table 1). Parameters may be fixed to single values, may be set to be free, or may be set to be deterministic functions of each other. In this analysis, the weights of each allowed sympatric range-copying (y), sympatric-subset (s), and vicariance (v) event were set to be fixed to 1 in the DEC analysis (reproducing the LAGRANGE model), and were set to equal $(3-j)/3$ in the DEC+J model, to facilitate comparison with DEC. Users are warned that, due to the operation, described above, of dividing

the per-event weights by the sum of the weights, parameter identifiability must be taken into account when setting parameters. For example, the model $y=s=v=j=1.0$ is identical with the model $y=s=v=j=0.5$.

An important question arises when parameterizing a founder-event speciation model, namely, how should features like dispersal constraints be implemented? 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 state 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 analyst's model of the relative probability of dispersal between regions. It would probably be unfair to the DEC model, in initial testing, to impose such constraints on the dispersal (range-expansion) process, but not impose them on the founder-event speciation process. Therefore, in the current implementation, dispersal matrices and other constraints are applied identically to d and j as probabilities are calculated – i.e., if dispersal (range expansion) between two areas is prohibited, jump dispersal during cladogenesis is also prohibited. It is not actually a given that the two processes would share the same constraints. For events like island submergence, they would, but for distance-based effects on dispersal probability, range-expansion dispersal and founder-event dispersal should be parameterized differently. For now, this is left as a matter for future research.

The parameterization described above ensures that the DEC model is a special case of the DEC+J model. The models are identical when j=0. This enables usage of the Likelihood Ratio Test for model testing.

Computational implementation

Due to the exponential expansion of the size of the transition matrix with the number of areas, and, equally importantly, the similarly dramatic expansion of the cladogenesis matrix, special attention was given to the computational implementation of both the anagenetic and cladogenetic models.

The anagenetic calculations (matrix exponentiations) are performed using the FORTRAN EXPOKIT library (Sidje 1998), a library well-regarded for its accuracy and speed in dealing with large matrices (Moler & Van Loan 2003). EXPOKIT was made accessible to R by creation of the BioGeoBEARS accessory R package, rexpokit (Matzke & Sidje 2013), approved and freely available from CRAN via R's install.packages command (R Core Team 2013). Parts of some EXPOKIT wrapper functions were modified from the open-source implementations in C++ LAGRANGE (Smith & Ree 2010) and the R package expoRkit (Sidje & Hansen 2012) with assistance from the authors (S. Smith, personal communication; N. Hansen, personal communication).

Matrix exponentiations using the rexpokit implementation of EXPOKIT were faster

than other matrix exponentiations available in R, using EXPOKIT's dense matrix

exponentiation routine, for transition matrices of a variety of sizes (data not

shown). EXPOKIT also contains routines for rapid exponentiation of sparse

matrices, which should speed calculation times for very large matrices, and these

are implemented in rexpokit and BioGeoBEARS. However, preliminary tests

showed that sparse matrix exponentiation produces only approximately similar

probability calculations, resulting in differences in inference, so the sparse matrix

routines are only recommended for experimental use in BioGeoBEARS.

The cladogenesis calculations are implemented the BioGeoBEARS accessory

package cladoRcpp (Matzke 2013c), also available on CRAN. For any biogeography

analysis with more than a few areas, direct enumeration of the entire cladogenesis

matrix – checking each and every combination of ancestral range, left daughter

range, right daughter range to see if it matches an allowed cladogenesis event –

becomes onerous and soon, impossible (Ronquist 1997). To make cladogenesis

calculations as fast as possible, the searches through ancestor/left/right descendant

combinations are performed in C++, accessed from R via the Rcpp package

(Eddelbuettel & Francois 2013). Further speed is attained by searching for each

category of cladogenesis event individually, using the known constraints on the size

of daughter ranges, and the relationships between the descendant and daughter

ranges that are imposed by the cladogenesis model, to narrow the state

combinations that must be checked. This search through the state combination

space need be performed only once for each calculation of the likelihood of the data, given the tree and model parameters (or once for each stratum, in a time-stratified analysis). Therefore, to make the cladogenesis event probabilities accessible for the calculations at each node during a downpass, cladoRcpp stores a table of allowed events and their probabilities, where every row stores the index of the ancestral state, the index of the left descendant state, and the index of the right descendant state, and corresponding probability for that particular range inheritance scenario. Events with zero probability are not stored in the table, dramatically reducing memory costs and improving calculation times. This is essentially the coordinate-ordered-object (COO) format used to efficiently store large sparse matrices, but for a 3-dimensional matrix.

Maximum likelihood optimization and ancestral range estimation

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 & Varadhan 2011; Nash & Varadhan 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), taking care to propagate the probabilities differently when moving up or

down the tree, due to the asymmetric nature of both the anagenesis and

cladogenesis processes. 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 & Smith 2008). This is ancestral state estimation

under local optimization (Mooers & 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 (local optimization requires a number of ML searches equal to the number of

nodes times the number of states, whereas global optimization requires only one ML

search) and because Mooers (2004) found that ancestral states estimated under the

global model are more accurate.

Validation

BioGeoBEARS was validated against both the Python version of LAGRANGE

(http://www.reelab.net/lagrange/configurator/instructions, versions 20120508

and 20130526) and C++ version (version 0.20, commit 33,

https://github.com/blackrim/lagrange). Validation was conducted in two ways.

First, for each test dataset, analyses were run in both versions of LAGRANGE. The

ML estimates of *d* and *e*, and the resulting log-likelihood of 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 to the LAGRANGE estimates.

Study system: Hawaiian Psychotria

The clade used for the primary validation study was Hawaiian *Psychotria*. *Psychotria* is a hyperdiverse genus (~2000 species worldwide) of tropical shrub, with a subclade endemic to Hawaii (Nepokroeff *et al.* 2003). Hawaiian *Psychotria* approximately follows the common Hawaiian progression rule, with the deepest-diverging lineages located on the oldest high island, Kauai. In the paper introducing LAGRANGE, Ree and Smith (2008) simplified and ultrametricized the *Psychotria* phylogeny published by Nepokroeff (2003), and this is the default example dataset for Ree's LAGRANGE Configurator (http://www.reelab.net/lagrange/configurator/index). The phylogeny at the Configurator website was downloaded and scaled to the recommended age of 5.2 million years (Ma).

Ree and Smith compared several constraints models using LAGRANGE and the *Psychotria* dataset. Their model M0 was an unconstrained analysis. M1 was

43

unconstrained, except that the maximum range size was set to 2 for the analysis. M2

had the 2-area constraint, and added the constraint that dispersal could only happen

from west-to-east. Finally, their "stratified" model only allowed dispersal between

islands when the islands are known to have emerged above sea level based on

geological data (Hawaii, 0.5 Ma; Maui-Nui, 1.9 Ma; Oahu, 3.7 Ma; Kauai, 5.1 Ma;

Clague 1996; Nepokroeff *et al.* 2003; Ree & Smith 2008). Each of these models was

run in both versions of LAGRANGE, and implemented in BioGeoBEARS for

validation. Then, a 3-parameter version of each model was constructed and run in

BioGeoBEARS.

Simulation to test specificity and sensitivity in model choice, and accuracy of ancestral

range estimation

Although the DEC+J model is only slightly more complex than the DEC model, with

one additional parameter, it is possible that any likelihood advantage DEC+J has in

explaining geographic range data on phylogenies is due to some artifact. The DEC

cladogenesis model is already rather complex, the number of areas is limited, and

the amount of data available is roughly equal to that found in a single multistate

character (but see Landis *et al.* 2013 for a more complex evaluation). This differs

sharply from the situation in e.g. analysis of DNA alignments, which involve a small

number of states (four) and hundreds of columns that are treated as statistically

independent replicates of a common substitution process (for discussion and testing

of this sort of independence assumption, see Nasrallah, Mathews & Huelsenbeck

2011; Landis *et al.* 2013). Furthermore, issues have been raised about likelihood-based inferences on discrete characters, for example emphasizing their uncertainty (Cunningham, Omland & Oakley 1998) or the difficulty in achieving sufficient support to justify multiparameter models (Mooers & Schluter 1999) in the relatively small phylogenies available in the studies of most clades.

Of course, historical biogeography is at an advantage over other fields in this respect: the parameters in the models of discrete range evolution are directly inspired by known processes and known geography. This can be illustrated by contrasting biogeography models with the models for the evolution of discrete morphological characters. In the latter case, the assumed model of probabilistic character state change is assuming *de facto* that all the complexities of evolution of development, selective regimes, character description, character coding, and character discretization can be put in a black box and modeled with a simple stochastic process. Nevertheless, it is useful to test, via simulation, our ability to discriminate between the DEC and DEC+J models. As has been noted, "[i]f a method does not perform well under very simple models of evolutionary change or if a method does not perform well even when all of its assumptions are satisfied, that method probably has little hope of performing well for real character data" (Huelsenbeck 1995). Simulation tests provide a "first cut" for assessing the quality of inference methods (Huelsenbeck 1995), yet, despite the ubiquity of simulation studies in other fields where selecting the best model of character evolution is considered crucial (e.g., Posada & Crandall 2001) there has been almost no

simulation work on model selection in historical biogeography (the exception is

Landis *et al.* 2013), and none comparing DEC and DEC+J.

To remedy this, 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 support for each model was measured

via the log-likelihoods, and a likelihood-ratio test with one degree of freedom

(Burnham & 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, i.e. when the p-value cutoff used to reject the null

hypothesis is set to 0.05 (the value used here), the null hypothesis should be falsely

rejected only in 1 in 20 simulation/inference pairs. 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 datasets simulated under

the DEC+J model.

The accuracy of ancestral state estimates was measured for each of the 4000

simulation/inference pairs in two ways. First, the mean accuracy was calculated by

calculating the fraction of the nodes for which the most probable state under

ancestral range estimation matched the true simulated range. Second, the mean

absolute difference from truth, *D*, was calculated, by taking the inferred ancestral

range probabilities at each node, and calculating the absolute difference between

the inferred range probabilities and the true probabilities of the known ancestral

ranges at each node. As the true range history is known under simulation, the

probability of the true ancestral state is 1, and the probability of all other states is 0.

The value *S*, where *S*=1-*D*, represents the mean similarity to truth, and *S*=1 would

represent an exact match between the inferred and true range probabilities. *S* is

more convenient for comparisons to mean accuracy. S was plotted as a function of

node age for the different simulation/inference pairings, to assess how estimation

accuracy changes as the inference moves further from the tips.

The accuracy of parameter inference was assessed by taking the distributions of

inferred parameters for each model/inference combination, and comparing them to

the true model parameters. All statistical analyses and graphs were done in R 2.15

(R Core Team 2013).

Bayesian implementation

ML estimation only provides point estimates of parameters, and frequentist model

tests such as the LRT require certain theoretical assumptions which may not always

be met in real datasets. As a check on the likelihood-based model selection

procedures used for most of this study, a Bayesian version of inference and model choice under DEC and DEC+J inference was implemented on the test *Psychotria* dataset, unconstrained (M0) geography model.

The impressive R package LaplacesDemon (Statisticat 2013b, available at http://www.bayesian-inference.com/software) implements a wide variety of Bayesian techniques. Given only user-specified priors on free parameters, and a user-specified likelihood function, MCMC chains can be constructed with minimal overhead in LaplacesDemon, which also supplies sophisticated diagnostic functions and plots. LaplacesDemon implements more than two dozen MCMC algorithms, two of were used here. First, standard random-walk Metropolis MCMC (Link & Barker 2009; Statisticat 2013a) was employed. Uniform priors were 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). The same BioGeoBEARS function that was used to calculate the data likelihood in ML estimation was used for the Bayesian analysis. The DEC and DEC+J models were run separately, each for 10,000 generations. A 2- or 3-parameter problem is a simple one compared to many MCMC analyses, and 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 posterior samples were used plot the posterior distribution of each parameter under each model. To approximate the marginal log-likelihood of each model for Bayesian model choice, the nonparametric self-normalized importance sampling

algorithm of Escoto 2011) was used (LaplacesDemon LML function, NSIS option). The Bayes factor was calculated from the ratio of the approximate marginal likelihoods.

The second MCMC analysis used reversible-jump MCMC (Green 1995) with both DEC and DEC+J sampled in the same MCMC search. As is often found in RJMCMC analyses, due to the strong likelihood advantage of one model (in this case, DEC+J), the prior probability of the two models has to be extremely heavily biased in favor of the poorer model (DEC) in order to achieve any sampling at all of the weaker model (Link & Barker 2009). A prior that produced adequate sampling of both models was thus sought by trial and error, resulting in a prior probability of 0.000001 placed on the DEC+J model. Not coincidently, this prior almost exactly balances the log-likelihood advantage of DEC+J (~14 log-likelihood units) on the *Psychotria* dataset. Priors even slightly different resulted in runs that exclusively or almost exclusively sampled only one model. The functioning RJMCMC analysis was run for 50,000 generations and was assessed as described above. Here, the Bayes factor was calculated as the ratio of the posterior probability of DEC+J (which is simply the frequency at which the DEC+J model was sampled in the post-burnin posterior distribution) to the prior probability of the DEC+J model.

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 island systems. Datasets 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 branchlengths and sufficient dating information to allow estimation of an ultrametric tree using r8s (Sanderson 2003).

Where possible, original tree files were gathered from supplemental material or other online sources, but usually the only available source for the tree was the published graphic. As has been noted (Boettiger & Temple Lang 2012), the digital archiving of published, dated, phylogenies is still not standard practice, and the limited tree-digitization software available is typically platform-specific or otherwise difficult to use, driving some researchers to use calipers on the published trees in printed-out journal article pages in order obtain trees with branchlengths (Boettiger & Temple Lang 2012). Here, a relatively workable alternative approach to digitizing trees was discovered. GraphClick, a cheap and user-friendly manual digitization program (Arizona Software, $8, http://www.arizona-software.ch/graphclick/) was used to obtain the x and y coordinates of the tip nodes, the internal nodes, and the "corners" (the branch bottoms below nodes on a typical phylogeny with square corners). Then, an R script, TreeRogue (Matzke

2013f) was used to assemble the coordinates and tip labels into a standard Newick file. An experienced user can use this system to digitize a tree to Newick format in a matter of minutes; some of the trees used in this study were digitized by undergraduate research assistants, although this requires above-average computer skills, training in the reading and interpretation of phylogenies, and strong attention to detail, as missing even one node during digitization will cause the Newick conversion operation to fail. Accuracy of digitization is approximately 1 millimeter on a printed page or computer screen, which translates to less than 1% error in the digitized tree compared to the original, certainly well below the statistical uncertainty inherent in any dating analysis.

Island clades were gathered opportunistically by searching Google Scholar for historical biogeography studies, especially studies that examined Hawaiian clades or that used LAGRANGE. Some older studies were included, despite the weaker dating and biogeography methods, because they covered "classic" island radiations (e.g., the Hawaiian silversword alliance). The full accounting of datasets and their sources is given in Table 2. The clades and source studies are briefly summarized below, along with notes about the methods used, for more recent papers when methods more sophisticated than simple single-state parsimony character mapping were available. The Hawaiian understory tree *Psychotria* (Rubiaceae, wild coffee; Nepokroeff *et al.* 2003; Ree & Smith 2008) served as the example dataset for the original Python LAGRANGE (Ree & Smith 2008) and is used as the default example on the LAGRANGE Configurator (Ree 2013). The biogeographically famous

Hawaiian Drosophilidae, considered by many to be the paradigmatic examples of founder-event speciation (Carson 1974; Hampton & Kaneshiro 1976; Templeton 1980; DeSalle & Templeton 1988) were sampled via an older, but well-dated study on Hawaiian *Drosophila* Kambysellis *et al.* 1995) and a state-of-the-art study on *Scaptomyza* (Lapoint, O'Grady & Whiteman 2013); the latter study used LAGRANGE These studies do not cover the full diversity of Hawaiian Drosophilidae, which is immense (ca. 900 species), but they do cover the breadth of these genera in Hawaii. The Hawaiian honeycreepers (Fringillidae: Drepanidinae), one of the most famous examples of a spectacular morphological adaptive radiation, was sampled via the study of Lerner et al. (2011), which used a large multilocus dataset, including molecular data from extinct taxa, to produce a well-dated phylogeny. The Hawaiian endemic damselfly genus *Magalagrion* (Odonata: Coenagrionidae) was explored by Jordan (2003), in a study which concluded that most speciation in the genus was inter-island (Jordan, Simon & Polhemus 2003). The hyperdiverse Hawaiian leafhopper genus *Nesophrosyne* (Cicadellidae) was comprehensively treated in a massive study (6 genes, 191 species) by Bennett and O'Grady (Bennett & O'Grady 2011; Bennett & O'Grady 2013). This study relied heavily on LAGRANGE for its historical biogeography estimates of ancestral areas and process (Bennett, 2013 #825). The historical biogeography of the recently-discovered Hawaiian endemic spider genus *Orsonwelles* (Araneae; Hormiga 2002) was studied by Hormiga et al. (2003) who used area cladograms to determine that this genus had more within-island than between-island speciation events, and that while the genus might follow the progression rule, other scenarios were also possible. Hawaiian representatives

of *Plantago* (Plantaginaeaceae), a genus of small forbs, were investigated by Dunbar-Co (2008), who concluded that they followed the usual progression rule. The information to do a biogeographical analysis of the Hawaiian silversword alliance (Asteraceae), another famous Hawaiian adaptive radiation, was assembled from two papers (Baldwin & Sanderson 1998; Gillespie & Baldwin 2010).

To supplement the 9 Hawaiian clades described above, 4 non-Hawaiian island clades were included as well. The ubiquitous study-system lizard genus *Anolis* (Dactyloidae) was included by way of an extensive biogeographical and taxonomic review of the members of the genus in and surrounding the Caribbean (Köhler & Vesely 2010; Nicholson *et al.* 2012). This was the only case in which a digitizable, comprehensive tree was not included as a Figure in the study; therefore the dated tree was regenerated from scratch by taking the input sequence alignment (obtained courtesy of K. Nicholson, personal communication) and conducting a BEAST dating analysis (Drummond & Rambaut 2007; Drummond *et al.* 2012) using the settings and calibration points described by Nicholson (2012), and taking the resulting well-resolved maximum clade credibility tree (MCC) for use in biogeographical analysis. It should be noted that the dates obtained by Nicholson et al. (2012) are quite old (the root of the genus tree is 87 Ma), as a result of the input calibrations, which consist of two fossils of age 28 mya and 17-23 mya, each of which the authors place very high in the tree as sister to extant species or small clades. Given the recent documentation of extensive morphological convergence in *Anolis* (Mahler *et al.* 2013) these calibrations might be questioned, especially as

Nicholson et al. note that many workers in the group favor younger dates (Nicholson *et al.* 2012). Nicholson *et al*. premise their discussion of biogeography on ancient dates and vicariance interacting with the complex geological history of the Caribbean. However, their biogeographic inference method is parsimony character mapping combined with narrative interpretation; there is therefore no explicit dispersal matrix or other constraint that must be taken into account in the present likelihood analysis (such an analysis would a fascinating, but very large, project). Therefore, only an unconstrained, non-time-stratified analysis was run in the present study. For such analyses, the relative branch lengths are important, but the absolute timescale is just a scaling factor which will not influence model comparison. Therefore, Nicholson et al.'s dating is taken as a given here; however, group seems ripe for a more sophisticated tip-dating approach that directly includes fossil OTUs and morphological models in the dating analysis (e.g., Wood *et al.* 2013).

Another lizard clade used was *Microlophus* (Tropiduridae), the "lava lizards" of the Galapagos (Benavides *et al.* 2009). Benavides et al. (2009) constitutes a detailed dating study, but the reconstruction of biogeographic history is primarily narrative based, albeit with topological tests for the monophyly of certain postulated geographical clades. The next clade sampled was *Cyrtandra* (Gesneriaceae), a tropical herbaceous plant with species spread across the Pacific. It was previously examined by Clark et al. (Clark *et al.* 2008; Clark, Wagner & Roalson 2009) in their comparative study of ML-based DEC, Fitch parsimony (Bremer 1992), parsimony-based dispersal-vicariance analysis (DIVA; Ronquist 1996; Ronquist 1997), and

stochastic mapping (Nielsen 2002). The final clade employed was southeast Asian members of subgenus *Vireya* of the shrub genus *Rhododendron* (Ericaceae; Brown, Nelson & Ladiges 2006; Webb & Ree 2012). Like Caribbean *Anolis*, but unlike the other clades considered in this study, *Vireya* is not primarily found on oceanic islands; it has representatives spread from the Himalayas to Australia (Webb & Ree 2012). However, Malesia is its center of diversity, and it is interesting to test whether or not a clade distributed on continental islands will support the DEC or DEC+J models in a fashion similar to oceanic island clades. In addition, *Vireya* has become another study group, serving as the example dataset for the introduction of methods such as SHIBA (Webb & Ree 2012) and BayArea (Landis *et al.* 2013). Due to the lack of rigorous dating information for the clade, Webb & Ree (2012) explore two different scalings of the age of the clade, 11 and 55 my. As a time-stratified analysis is not being done in the present study, as discussed above this scaling issue is unimportant for model selection, so only the 11 my scaling was used.

Datasets could be analyzed once they had been processed into the input format required for BioGeoBEARS and C++ LAGRANGE, namely, a Newick file and a geography text file in PHYLIP format (Smith 2010). To achieve maximum comparability and broadness of results, if the source study in question made use of a dispersal matrix, a time-stratified dispersal matrix, or other constraints, these were also converted into the necessary input text 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 dispersal matrix) constraints models of Ree & 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 that 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 2 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 before Kauai (Clague 1996). Note that even though no extant tips in these phylogenies inhabit area Z, BioGeoBEARS will still infer it as the ancestor after all of the extant high islands have submerged. The constraints models are summarized described in Table 2, and the details are found in the geography files.

All input files (Newick files, geography files, and constraints files, as well as the scripts used to run each analysis) are available in Dryad (dryad.org). In addition to replicating the constraints used in published analyses, unconstrained analyses were run for comparison. DEC and DEC+J inferences were run on each dataset, and on each available dispersal model for datasets that included these. The DEC and DEC+J

models were compared with the likelihood ratio test, AIC, and AICc (Burnham & Anderson 2002) using custom functions implemented in BioGeoBEARS (Matzke 2013b).

Results

Validation

BioGeoBEARS's DEC model performed extremely well in validation tests against LAGRANGE for on the example *Psychotria* dataset for constraints M0-M3 (Table 3). In almost all cases, BioGeoBEARS was able to reproduce the exact log-likelihoods returned by LAGRANGE for the same input data and parameter estimates. In addition, the ML searches conducted by BioGeoBEARS returned exact or virtually exact matches to the parameter estimates and log-likelihoods returned by LAGRANGE. This is strong evidence that the fundamental logic of LAGRANGE has been understood and incorporated into BioGeoBEARS, even though BioGeoBEARS was rewritten from scratch in R, and contains additional features and parameterizations, including founder-event speciation and others not used here, that make the construction of BioGeoBEARS more complex.

The one exception to perfect matches in the validation tests was in the time-stratified analysis of *Psychotria*. Here, BioGeoBEARS gave slightly different parameter estimates and log-likelihoods than some of the LAGRANGE

implementations. However, the LAGRANGE implementations differed among each other as well. Programming a stratified analysis is a rather complex affair, as the phylogenetic essentially tree has to be sectioned, sometimes producing subtrees with root branches, and sometimes producing branch segments, and the likelihoods have to calculated, tracked, and correctly passed on to the connecting pieces during the downpass, and the uppass as well, if ancestral states are being estimated. The 2013 version of Python LAGRANGE repaired a bug in the stratified calculations in the 2012 version (Ree 2013), and BioGeoBEARS returns the same result as the 2013 version. C++ LAGRANGE returned the most divergent result, indicating some issue different from that in the 2012 Python LAGRANGE.

One important, though not itself comprehensive, check of stratification calculations is to run the same analysis in a stratified and non-stratified context. This "null stratification" simply uses the same dispersal matrix as the non-stratified analysis and repeats the identical matrix at each time step. All programs passed this check except 2012 Python LAGRANGE (Table 3).

Hawaiian Psychotria*: DEC versus DEC+J*

Comparison of the performance of the 2-parameter DEC model to the 3-parameter DEC+J model on the Hawaiian Psychotria dataset yielded dramatic results on all constraints models (Table 4). At the cost of only one additional free parameter, j,

the likelihood of the tip range data conferred by the model jumps by 11-15 log-likelihood units.

With one additional free parameter, approximately 2 log-likelihood units would constitute a statistically significant increase. As a result, formal model tests show that DEC+J is very strongly favored in all cases, with the Likelihood Ratio Test rejecting the nested 2-parameter model at the p<0.0001 cutoff level in all cases.

AIC weights gives a sense of the relative support that the data lend to the 3-parameter model. The ratio of the weights in favor of DEC+J is impressive, ranging from 1800 (for the highly constrained, time-stratified M3 model) to 10^6 (for the M2 model where dispersal is eastwards-only). As the dataset is small (only 19 OTUs, and thus 19 geographic ranges, in the *Psychotria* tree), AICc was also calculated. However, because of the small number of free parameters (2 or 3), the results with AICc were substantively the same, with only a slight decreasing in the overwhelming support for DEC+J.

In sum, with the *Psychotria* dataset, from a frequentist perspective, the DEC model is decisively rejected in favor of a model that includes the process of founder-event speciation, and from a model selection perspective, DEC+J earns virtually all of the model weight.

Consideration of parameter inference under DEC and DEC+J 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 ML inference 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, this is accounted for, with much higher probability, by a series of cladogenesis events, mostly sympatric range-copying within a single area, and founder-event speciation. The only 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 models are closer in the likelihood they confer on the tip data. In this highly-constrained model, the range-expansion and range-contraction processes have much less flexibility. In an unconstrained model, the dispersal-extinction process suggests substantial waiting times between dispersal (range-expansion) events, and subsequent cladogenesis events. These waiting times lower the probability of the data, as many other events could be happening instead, with approximately equal probability. When there are tight area constraints, by contrast, range-expansion only can occur after a new island has emerged, which will often also coincide with an appropriate cladogenesis event in the dated tree. Thus, the advantage of the founder-event speciation events – instantaneous change perfectly correlated with speciation events, and no need for long residence times in widespread states before cladogenesis occurs – is reduced, bringing the models closer together.

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 tendency of the model; Clark *et al.* 2008; Ree & Smith 2008; Buerki *et al.* 2011), which then break up through vicariance to produce daughter lineages with narrow ranges. Under DEC+J, this phenomenon completely disappears. It is replaced by founder-events at each node where geography changes. It appears that the entire geographic distribution of *Psychotria* in the Hawaiian islands can be explained by only two processes: founder-event speciation, and sympatric speciation within single islands. This fits the conceptual model held by many biogeographers of oceanic islands; however, notably, this is not the computational model implemented in DEC analyses.

The pie charts depicting the relative probability of each ancestral range (Figure 2, right side) also show that ancestral states are reconstructed with much less uncertainty under the DEC+J model than the DEC model. 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. One weakness of DEC, for this dataset, appears to be that it has difficulty finding the ancestral states that confer the highest likelihood on the data, and reconstructs several nodes with very high uncertainty. This important fact can be missed in plots that show only the single most-probable state (Figure 2, left side).

Two somewhat subtle points should be made about interpreting the ancestral range

estimates in depictions such as Figure 2. First, these depict the best estimates, under

the model parameters, of the ancestral range at each node and each "corner" (the

corner positions are used to represent the geographic range immediately after a

cladogenesis event). The best estimate at each node is not identical with the single

best joint history (Felsenstein 2004), so in some locations on a phylogeny, the

highest probability ancestral ranges at each individual node cannot be read naively

as a joint history. Second, unlike LAGRANGE, BioGeoBEARS by default estimates

ancestral states under the global ML model. LAGRANGE estimates ancestral states

by conducting an ML search for the best model, conditional on fixing each possible

state at each node in the tree. The resulting ancestral states will usually be similar

but not necessarily identical.

Simulation results: false positives and false negatives

For the 1000 simulations under each ML model, the difference in data log-likelihood

between the ML models estimated under DEC and under DEC+J was calculated. The

top histogram in Figure 3 shows the log-likelihood advantage of the DEC+J model

over the DEC model, when the true model is DEC. As DEC is a special case of DEC+J,

the additional free parameter of DEC+J means that it will always explain the data at

least slightly better than DEC. However, when the true model is DEC, this advantage

is very slight, usually less than 2 log-likelihood units; the mean advantage is 0.247.

This is exactly the desired behavior, indicating that when DEC+J has a large

advantage over DEC, this is not likely to be due to chance or to some systematic artifact.

This conclusion is confirmed by the lower panel in Figure 3, which shows the log-likelihood advantage of DEC+J over DEC when the true model is DEC+J. In this situation, the advantage of DEC+J is massive, with a mean advantage of 7.86 log-likelihood units. Only rarely is the advantage less than 2 log-likelihood units, which would indicate situations that produced a false negative, in which DEC would not be rejected in favor of DEC+J, in spite of DEC+J being the correct model.

These impressions are confirmed by calculation of the rates of false positives and false negatives under the Likelihood Ratio Test (Table 5). When data was simulated under DEC, the LRT rejected DEC in favor of DEC+J only 43/1000 times, a rate approximately equal to the p-value cutoff ($p<0.05$). This indicates that the LRT has the desired frequentist properties when comparing DEC to DEC+J. On the other hand, when data was simulated under DEC+J, false negatives (failures to reject DEC when DEC+J is true) occurred for 59/1000 simulations, or a false negative rate of 0.059. This too is acceptable, although the false negative rate might well decline further if a similar simulation experiment were conducted on a larger tree.

Simulation results: accuracy of ancestral state estimates

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 was

calculated. The distribution of this fraction is shown in Figure 4. The mean fraction

of nodes correct 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, 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).

Examining the similarity of all ancestral state probabilities to the true probabilities

through calculation of S for each node yielded very similar histograms (data not

shown), and the same ordering: DEC+J simulations, DEC inference: 0.62; DEC

simulations, DEC inference: 0.78; DEC simulations, DEC+J inference: 0.81; DEC+J

simulations, DEC+J inference: 0.84. These differences in mean S were also all

significant, using the same tests as above (all p-values < 2e-15).

It is somewhat surprising that DEC+J inference of ancestral ranges is more accurate

than DEC inference, when the true generating model is DEC. We seem to have a

moderate exception to the otherwise well-established dictum that "the performance

of a method is maximized when its assumptions are satisfied" (Huelsenbeck 1995).

In the present case, the explanation might be related to some interaction between

the of range sizes that each model prefers, and the fact that the starting ancestral

range for the simulations was usually either Kauai or Kauai+Oahu. For example, if

the simulations produce many nodes with small geographic ranges, due to the

starting state, but the DEC model tends to infer widespread ancestors, DEC+J could

have a slight advantage simply because it tends to prefer single-area ancestors.

The mean value of S for each node age in the *Psychotria* phylogeny, plotted against

age, is graphed in Figure 5 for each of the four simulation/inference pairs. Reduced

accuracy for older nodes is seen in all analyses, and the relative ordering of

accuracies remains the same for all node ages.

Simulation results: parameter inference

Figure 6 compares the ability of DEC and DEC+J to correctly infer parameter values

when biogeography was simulated under each model. As demonstrated for

LAGRANGE (Ree 2005; Ree & Smith 2008), DEC's ability to correctly infer d is good,

but it typically vastly underestimates e. DEC+J shows similar properties. It is

probable that very little information about extinct ranges is usually retained in the

tip geography of a phylogeny, especially in a relatively small phylogeny. DECJ+J

usually does correctly infer d, and correctly infers a j parameter near zero, when the

true value of j is zero.

When the true model is DEC+J, with d and e approximately zero, DEC of course

nevertheless infers a positive value for d, as some positive value is required in order

to get lineages distributed across the geography of the study area in the first place.

Interestingly, when the true model is DEC+J, DEC inference will sometimes infer a

positive value for *e*, even though the true value of *e* is zero. Inference of a positive *e*

is infrequent, but it still occurs much more commonly than when DEC inference was

performed on DEC simulations with a truly positive value of *e*! The answer to this

paradox is likely the fact that the jump dispersal process will sometimes create

distributions on the phylogeny that DEC can only explain by positing a range

expansion, followed by a range loss. This configuration of the data would

mistakenly appear to give some positive signal for the extinction/range-loss

process, and for a positive value of *e*. Excluding the process of founder-event

speciation from the inference, when it is a part of the true model, leads to

pathological behavior in DEC-based parameter inference.

On the other hand, when DEC+J is the true process, DEC+J inference usually recovers

zero values for *d* and *e* nearly 100% of the time, and also usually infers a positive

value for *j*. It appears that *j* is typically underestimated, although in the simulation

experiment conducted here, the true value is at the 90.6% percentile of the

empirical distribution of inferred *j* estimates, and so not significantly outside of the

distribution. On a small phylogeny, with only four areas, the number of jump

dispersal events that might occur in a forward simulation might be small, and the

probability of "convergence" (independent jump dispersal to the same island) would

not be an infrequent occurrence. Furthermore, there is no guarantee that every

forward simulation will produce tip species occupying all four islands. If these

factors are influencing the underestimation of *j*, they would be expected to weaken

on a collection of simulations that are conditioned through rejection sampling to

have tip species occupying all four islands, or in simulations conducted under constraints that prevent, for example, back-dispersal to older islands. Analyses with larger trees and more areas would also be expected to be less subject to underestimation of j.

Bayesian inference on Psychotria M0 dataset

Trace plots and histograms for the parameters and the model log-posterior probabilities are shown in Supplementary Figures 1, Supplementary Figures 1.1-1.6. The results of random-walk Metropolis MCMC are shown in Supplementary Figures 1.1-1.4. Inference of d appears to be unproblematic for both DEC and DEC+J, with the ML estimate of d appearing at the peak of the posterior distribution in both cases. However, for the e parameter, the ML estimate (0.028) is well outside the 95% highest posterior distribution (HPD) for the DEC model, a result further highlighting the weaknesses of DEC inference in estimating e. Even when the ML estimate of e is nonzero, the likelihood surface is extremely broad, and the likelihood advantage of a positive e value over $e=0$ is tiny. Inference of j appears unproblematic, with the ML estimate of j (0.11) within the 95% HPD, and with $j=0$ outside of the 95% HPD.

The estimates of the log-marginal likelihood for DEC and DEC+J were -44.6 and -36.4, respectively. This yields a Bayes Factor of 3595 in favor of DEC+J. On the Jeffreys' scale for the interpretation of Bayes Factors (Jeffreys 1961), anything above

100:1 is considered decisive support for a model. Under reversible-jump MCMC,

prior probability of DEC+J was set to 0.000001, and the posterior probability was

0.1524, resulting in a Bayes Factor of 152422. This Bayes Factor is similar in

magnitude to the ratio of model weights calculated from AIC, and indicates that the

likelihood advantage of DEC+J is dominating the result. The Bayes Factor calculated

from the posterior distribution random-walk Metropolis MCMC appears to be more

conservative, taking into account the uncertainty in parameter inference and

resulting uncertainty in model choice.

DEC versus DEC+J on island clades

Figure 7 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. In general, the

results are overwhelming support for the importance of including founder-event

speciation as an explanatory process in the biogeography of island clades, with

DEC+J being anywhere from 4 times more probable than DEC, to 10^{17} times more

probable. The same figure calculated for AIC weights is almost identical (not

shown) and Table 6 indicates that the AIC weight ratios are even more impressive.

However, as AIC has no correction for small sample size, AICc is recommended as a

more conservative metric (Burnham & Anderson 2002).

The only 2 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 (Ree & Smith 2008). In both cases, a more realistic implementation of the constraints model taken from the source studies, improved the situation for DEC+J. In the case of 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, O'Grady & Whiteman 2013), representing a permanent but very distant continent (Table 6).

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 datasets (a maximum of 198 OTUs).

Figure 7 and Table 6 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 in likelihood ratio test terms). However, this 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.

Turning to inference on the explanatory processes, Figure 11 shows the ML point

estimates of d, e, and j for each clade/constraint combination, under DEC and DEC+J.

In essence, when j is moderate or large, d and e drop to zero or are greatly reduced.

As noted above for *Psychotria*, this is evidence that the "DE" part of DEC is not even

needed in these clades. However, d and e do not always drop to 0; sometimes all

three processes are supported. There even 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, any phylogeny with

more than one tip with a widespread range will require a positive d to explain the

data. Inference of e is extremely problematic, as discussed previously, so no huge

weight should be put on the e estimates. The j parameter is always positive (except

for the two cases where DEC+J was not supported, discussed above), 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, i.e., subtrees with area arrangements such as

(A(A(A,D))). However, when j=3, this means that the values of y, s, and v

(controlling sympatric range-copying, sympatric subset, and vicariant speciation,

70

respectively) are 0. In this situation (here found in *Microlophus*), jump dispersal is the *only* process that needs to be invoked to explain distributions.

For illustrative purposes, two pairs of range estimates are shown (Figure 9 and Figure 10), comparing the inference under DEC versus DEC+J. All of the inferences for all of the runs are available in a large PDF available in Supplemental Figures 2. A useful way to compare the ancestral range estimates made under DEC and DEC+J is to size a window such that one page at a time is viewed, and then flip back and forth between pages of the PDF, using the Page Up/Page Down keys.

The general picture that emerges across all island clades examined is that the patterns found in *Psychotria* apply in general: under the DEC+J model, ancestors are usually inferred to have narrower ranges, this inference is made with higher confidence, and the estimated history can usually be interpreted quite simply as a 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 model that is apparently inappropriate in such situations, namely DEC, has come to be so widely applied, even in the island situation.

Conclusions

The probabilistic Dispersal-Extinction-Cladogenesis (DEC) model implemented in

LAGRANGE (Ree *et al.* 2005; Ree & Smith 2008) was revolutionary. It was the

advent of parametric biogeography (Ree & Sanmartín 2009), wherein rather than

simply relying on parsimony reconstructions, events are assigned probabilities, and

optimal estimates of history can be made by maximum likelihood or Bayesian

approaches. However, probably because it was the first major effort, LAGRANGE

did not fully explore all of the potentiality inherent in the parametric approach. One

of major advantages 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, which adds founder-

event speciation to the suite of processes available in the original DEC model. As

demonstrated here, that the BioGeoBEARS implementation replicates LAGRANGE's

DEC model, and the ML inferences from this model. A simulation experiment

showed that the two models have surprisingly good identifiability, and that when

the likelihood ratio test is used to test the models, the test has high specificity and

sensitivity, and that the false positive rate matches the *p*-value cutoff, a key desired

behavior of a frequentist test. The simulation also showed acceptable estimation of

d and *j*, but not *e*, similar to the results of Ree et al. (2008). Bayesian inference

backed this point up, by showing that the ML point estimate for *e* can be well outside

of the 95% HPD estimated for the parameter.

Finally, the simulation study 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 stark – when DEC+J was the true process, the accuracy of ancestral range estimation under DEC was 57%, but under DEC+J it was 87%.

When we turn to comparisons of DEC and DEC+J on real datasets, the results are quite stark. On a broad sample of island clades, there is massive support for DEC+J over DEC, over a wide range of constraints models. The differences in data likelihoods, and resulting AIC and AICc, are of approximately the same order of magnitude as those found when comparing DEC inference and DEC+J inference on the simulated DEC+J data. In other words, reality looks like it was 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 apparent elimination of the bias towards widespread ancestors sometimes found in DEC.

What is the mathematical cause of the dramatic improvement in likelihoods under the DEC+J model, on these datasets? A hypothesis is presented here (Figure 11). 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. Thus, a positive d is required to spread lineages out, but having a positive d means that branch histories that produce the correct ancestor at the correct time (the time of speciation) will be somewhat improbable, compared to all of the other possible histories. The waiting time between range expansion and an uncorrelated cladogenesis process is probably what lowers likelihoods in the DEC model. By contrast, on the DEC+J model, d and e are often zero, meaning the probability of no change anagenetically along a branch is 1. Cladogenesis and dispersal are perfectly correlated, and if the tip geography is suitable, a very simple history of founder-events can explain the distribution.

A pessimistic view would be to see the above result as a negative finding about the DEC model and its widespread use in apparently inappropriate situations such as island clades, but this would be to miss the most important point. The first probabilistic models that are applied to some phenomenon are always oversimple. DNA substitution models are a case in point – the model of Jukes and Cantor (1969) was a simple, equal-rates model, but it served as a starting point; later models incorporated violations of the equal-rates assumption, such as the transition-transversion rate ratio; eventually an entire family of models was produced, and model selection procedures are now used routinely pick the best substitution models and to improve phylogenetic inference as a result (Posada & Crandall 2001).

The optimistic view is to note that through utilization of multiple models and model selection tools, we may finally begin to get at the crucial question of causal processes in historical biogeography, in a statistically rigorous way. A great many island biogeographers have thought for a very long time that jump dispersal/founder-event speciation is a crucial process in the historical biogeography of island taxa. On the other, for just as long, the vicariance biogeography wing of the field has disagreed with such dispersalist interpretations, even for oceanic islands (Heads 2012). Probably most readers have found the generations-long vicariance-versus-dispersal debate interminable at times. The major problem has been that adherents of each approach have been stuck in the rut of simply assuming, as a starting point for analysis, that vicariance or dispersal will be used as the explanatory process. The vicariance biogeographers provide the most spectacular examples. For example, Heads (2012) writes, "Most modern biogeographers follow Mayr…in accepting that allopatry can be found by vicariance (dichopatry) or by founder dispersal (peripatry), but only vicariance is accepted here." (Heads 2012, p. 15) He writes elsewhere that "founder dispersal…is controversial and may not exist." (p. 12)

Such seemingly dogmatic *a priori* assumptions may strike many readers as unwise or even unscientific, but the statements of Heads do have the important virtues of honesty and clarity. Clarity may be lacking in other parts of the field, however, particularly when it comes to computational methods being employed in inference.

For example, molecular dating often returns dates too young for classic vicariance explanations (Wood *et al.* 2013 is one of the few exceptions). Heads dismisses molecular dating methods wholesale precisely because they seem to falsify classic vicariance. He is reacting against the dominant view in biogeography, which is that the dating is approximately correct, and that therefore biogeographers should conceptually adopt a dispersalist approach unless the dating supports vicariance. Strangely, though, the methods that are most commonly used in historical biogeography, namely DIVA and LAGRANGE DEC, only allow dispersal in the sense of range expansion, one of the two processes (along with vicariance) that Heads, the vicariance biogeographer extraordinaire, accepts! (Heads 2012, pp. 11-13). Nevertheless, Heads associates DIVA and DEC so strongly with his dispersalist opposition that he writes, (p. 311), "ancestral area analysis (using programs such as DIVA and Lagrange), is also based on center of origin/dispersal theory (as shown by Santos, 2007) and, in particular, endorses the key concept of speciation by dispersal." As we have seen (Figure 1; see also Ronquist & Sanmartin 2011), this is not the case.

If we consider for a moment the "phylogeny" of the methods, DEC descends from DIVA (Webb & Ree 2012), and DIVA descends from maximum-vicariance methods (Ronquist 1996; Ronquist 1997) popular in the heyday of vicariance biogeography. As a result of this historical contingency, neither of them ever incorporated founder-event speciation. However, because the methods work on phylogenies, and because they are almost the only methods available (apart from single-state character

mapping methods, the properties of which, as a biogeographic method, have not been sufficiently studied either), the most common situation today is that modern historical biogeographers happily mix a dispersalist conceptual and dating framework with a nondispersalist, vicariance- and sympatry-heavy inference method. No wonder historical biogeography sometimes seems so confusing!

I suggest that the situation would be vastly improved, both conceptually and practically, if model selection were instituted as a regular procedure in historical biogeography analyses, just as it is already is with DNA substitution models in molecular phylogenetics. Obviously this should not be restricted merely to the DEC and DEC+J models. Within the framework of BioGeoBEARS, it is a relatively trivial matter to convert the DEC model into a DIVA-like model (Ronquist 1997; Ronquist & Sanmartin 2011), or a BayArea-like model (Landis *et al.* 2013), and from there to create, for example, DIVA+J and BayArea+J models. Additional models, as-yet unnamed, can be implemented by, for example, fixing y, s, or v to 0 or other values, or allowing them to vary independently and see if the improvement in likelihoods warrants the increase in model complexity. One particularly obvious modification to the extant models that should be explored concerns the e parameter. As it appears that the likelihood surface for e is usually very flat, there may not be much difference in likelihoods for $e=0$, or $e=d$, compared to leaving e as a free parameter. These options are all readily accessible in BioGeoBEARS and should be fully explored.

Phylogenetic biogeography has always been dominated by a focus on inferring the history of individual taxa, usually assisted by some strong a priori assumptions about the processes that may be invoked in histories. While estimating the biogeographic history of individual clades is a worthy goal, another, higher goal of historical biogeography has always been to tease out the dominant causal processes behind observed phenomenon. Inference of cause gives rise to a fundamental understanding of a subject, and the histories of individual clades then may be seen as realizations of the underlying processes. However, the goal of inferring cause has been much more elusive, as successful inference of cause from pattern requires sophisticated modeling. Now that this has become accessible, rather than simply assuming which processes will be allowed up front (Heads 2012), we may let the data tell us which models are to be preferred (Huelsenbeck 1995; Huelsenbeck *et al.* 2001).

When causal processes where subject to inference in this study, a huge 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 (data not shown) indicates that whereas DEC+J realizations tend to have tip species with ranges of single areas, often allopatric, DEC realizations tend to have more species that are widespread, and a fair degree of overlap and sympatry. Therefore, study systems

with many species that are widespread and in sympatry may support the DEC model over DEC+J. Intracontinental clades in the tropics, for example, may fit the bill. On the other hand, studies that treat continents as areas, with few species widespread across multiple continents, may be fit better by the DEC+J model.

Given that the importance of founder-event speciation has long been conventional wisdom amongst many island biogeographers, some readers might consider the use here of elaborate modeling and statistical techniques to confirm conventional wisdom to be a case of beating a dead horse. However, the categorical dismissal of founder-event speciation still seems to be popular amongst certain biogeographers. Heads was quoted above; another recent and colorful statement may be found in Santos (2007): "The resurrection of dispersalism, as de Queiroz (2005) has envisioned, is reactionary and with little (or even no) empirical foundation." I conclude that the old advice of Zuckerkandl & Pauling (1965) may still be worth following: "Some beating of dead horses may be ethical, when here and there they display unexpected twitches that look like life."

Acknowledgements

All Supplementary Material (code, files, and figures) will be available on Dryad (dryad.org).

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). These transition tables assume a study system of 3 areas and a maximum range size of 3 areas; analyses with a different number of areas will have different transition matrices (see Supplemental 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, 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). I.e., 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. Here, for simplicity, all events have been given equal weight. 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 the ancestral range.

A. Parameterization of LAGRANGE'S DEC cladogenesis model for 3 areas. Transition probabilities are event weight / count of events. Blank cells have 0 probability.

Ranges of descendant pairs

Parameters — Ancestral ranges (event weight matrix), with summary:

Ancestral range	# of each event — y	y	v	j
A	1	0	0	0
B	1	0	0	0
C	1	0	0	0
AB	0	4	2	0
AC	0	4	2	0
BC	0	4	2	0
ABC	0	6	6	0

Weights — Ancestral ranges, Count of events:

Ancestral range	Count of events
A	1
B	1
C	1
AB	6
AC	6
BC	6
ABC	12

B. Parameterization of the DEC+J cladogenesis model for 3 areas. Transition probabilities are event weight / count of events. Blank cells have 0 probability.

Ranges of descendant pairs

Parameters — Ancestral ranges (event weight matrix), with summary:

Ancestral range	# of each event — y	s	v	j
A	1	0	0	4
B	1	0	0	4
C	1	0	0	4
AB	0	4	2	2
AC	0	4	2	2
BC	0	4	2	2
ABC	0	6	6	0

Weights — Ancestral ranges, Count of events:

Ancestral range	Count of events
A	5
B	5
C	5
AB	8
AC	8
BC	8
ABC	12

82

Table 2. Study clades, data sources, and constraints settings (maximum range size, dispersal matrices, time-stratification, areas-allowed matrices, etc.). In total, 13 clades and 53 constraints models were run.

Region	Clade	Dataset source	Modifications	# taxa	tree age	Abbreviation	# areas	max# areas	Constraints model
Hawaii	*Psychotria*	Ree & 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 & 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	M3areas	5	5	stratified dispersal & areas
				42	20	M3areas	5	5	stratified dispersal & 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 Honey-creepers	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 & 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
				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 & 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	*Orsonwelles*	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 NJM	16	5.1	M0	5	5	unconstrained
				16	5.1	M3_strat	5	5	stratified as with *Psychotria* (Ree & Smith 2008)
				16	5.1	M3_strat	5	5	stratified dispersal and areas
Hawaii	Silversword alliance	Baldwin & Sanderson (1998) phylo, Gillespie & Baldwin (2010) geog	Digitized by NJM; some polytomies arbitrarily resolved.	29	6.84	M0	5	5	unconstrained
				29	6.84	M3_strat	5	5	stratified as with *Psychotria* (Ree & Smith 2008), with areas
Caribbean	*Anolis*	geography data and date priors from Nicholson et al. (2012)	MCC tree regenerated from BEAST analysis	190	98.9	M0	12	3	unconstrained
Pacific	*Cyrtandra*	Clark et al. phylo & geog	None; digitized by NJM	59	41.5	M0	7	7	unconstrained
				59	41.5	M0	7	4	4 areas max
Galapagos	*Microlophus*	Benavides et al. (2009) phylo & geog	None; digitized by NJM	9	3.75	M1	10	2	2 areas max; otherwise unconstrained
SE Asia	*Vireya*	Webb & Ree (2012) phylo & geog	None; used 11 mya root	65	11	M1	20	2	2 areas, no other constraints
				65	11	M1+cc	20	2	2 areas + constrained connectivity

Table 3. Validation of the BioGeoBEARS implementation of the 2-parameter DEC model against the various versions of LAGRANGE. The full precision output by each program is included to allow detailed comparison.

Program	Taxon	Model	Description	LnL	d	e	Dataset
Python LAGRANGE (2012)	*Psychotria*	M0		-34.54	0.03505	0.02831	Ree & Smith
Python LAGRANGE (2013)	*Psychotria*	M0	unconstrained	-34.54	0.03505	0.02831	(2008) phylo &
C++ LAGRANGE	*Psychotria*	M0		-34.542	0.0350117	0.0282904	geog
BioGeoBEARS DEC model	*Psychotria*	M0		-34.5	0.0351	0.0285	
Python LAGRANGE (2012)	*Psychotria*	M0_strat	unconstrained,	-35.36	0.03818	0.04079	Ree & Smith
Python LAGRANGE (2013)	*Psychotria*	M0_strat	with null	-34.54	0.03505	0.02831	(2008) phylo &
C++ LAGRANGE	*Psychotria*	M0_strat	stratification	-34.542	0.0350117	0.0282904	geog
BioGeoBEARS DEC model	*Psychotria*	M0_strat		-34.5	0.03506479	0.02833348	
Python LAGRANGE (2012)	*Psychotria*	M1		-35.21	0.04271	0.04305	Ree & Smith
Python LAGRANGE (2013)	*Psychotria*	M1	2 areas max	-35.21	0.04271	0.04305	(2008) phylo &
C++ LAGRANGE	*Psychotria*	M1		-35.2111	0.0427272	0.0429474	geog
BioGeoBEARS DEC model	*Psychotria*	M1		-35.2	0.04270405	0.0427981	
Python LAGRANGE (2012)	*Psychotria*	M2	2 areas +	-31.92	0.1228	0.00000055	Ree & Smith
Python LAGRANGE (2013)	*Psychotria*	M2	eastwards	-31.92	0.1228	0.00000055	(2008) phylo &
C++ LAGRANGE	*Psychotria*	M2	dispersal only	-31.92	0.122973	4.7416E-07	geog
BioGeoBEARS DEC model	*Psychotria*	M2		-31.9	0.1227461	1E-15	
Python LAGRANGE (2012)	*Psychotria*	M3	time-stratified	-39.8	0.03687	7.233E-08	Ree & Smith
Python LAGRANGE (2013)	*Psychotria*	M3	dispersal	-39.83	0.03645	4.495E-08	(2008) phylo &
C++ LAGRANGE	*Psychotria*	M3	matrix	-36.6318	0.0342543	1.1488E-07	geog
BioGeoBEARS DEC model	*Psychotria*	M3		-39.8	0.03643885	1E-15	
Python LAGRANGE (2012)	*Cyrtandra*	M0		-46.43	0.001251	4.285E-09	Clark et al.
Python LAGRANGE (2013)	*Cyrtandra*	M0	unconstrained	-46.43	0.001251	4.285E-09	(2008, 2009)
C++ LAGRANGE	*Cyrtandra*	M0		-46.4624	0.00110037	2.9021E-05	phylo & geog
BioGeoBEARS DEC model	*Cyrtandra*	M0		-46.42506	0.00125161	1E-15	
C++ LAGRANGE	*Cyrtandra*	M1	4 areas max	-47.2735	0.00144745	3.0655E-05	Clark et al.
BioGeoBEARS DEC model	*Cyrtandra*	M1		-47.3	0.00153	1E-15	(2008, 2009)
Python LAGRANGE (2012)	*Lonicera*	M0		-125	0.005812	0.001049	Smith &
Python LAGRANGE (2013)	*Lonicera*	M0	unconstrained	-125	0.005812	0.001049	Donoghue
C++ LAGRANGE	*Lonicera*	M0		-124.977	0.00581186	0.00111119	(2010) phylo &
BioGeoBEARS DEC model	*Lonicera*	M0		-125	0.00581059	0.00104781	geog

Table 4A. Comparison of DEC and DEC+J on Hawaiian Psychotria, for various models. Models M0-M3 as described in text. M2b is M2 (eastwards-only dispersal) but with maximum rangesize 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.

Model	Const-raints	max # areas	LnL	Parameter estimates				Likelihood Ratio Test			
				num	d	e	j	LnLalt	LnLnull	D	pval
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	27.19	2E-07
DEC	M1	2	-35.2	2	0.043	0.043	0				
DEC+J	M1	2	-21	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	30.49	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	29.47	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	23.03	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	17.85	2E-05

Table 4B. 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	Const-raints	max # areas	LnL	Akaike Information Criterion					2nd Order Inf. Criterion (AICc)				
				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	294895	49.5	73.83	1	5.2E-06	191951
DEC	M1	2	-35.2										
DEC+J	M1	2	-21	47.9	74.4	1	1.8E-06	566941	49.52	75.17	1	2.7E-06	371386
DEC	M2	2	-31.9										
DEC+J	M2	2	-16.7	39.4	67.8	1	6.5E-07	1537750	40.94	68.59	1	9.9E-07	1009534
DEC	M2b	4	-31.4										
DEC+J	M2b	4	-16.7	39.4	66.8	1	1.1E-06	924075	40.94	67.57	1	1.6E-06	606220
DEC	M3	4	-39.8										
DEC+J	M3	4	-28.3	62.6	83.7	1	2.7E-05	36882	64.24	84.41	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	73.16	88.17	1	0.00055	1816.11

Table 5. False positive rates (A) and false negative rates (B) for the Likelihood Ratio Test applied to the DEC and DEC+J models, on 2000 biogeographical histories simulated on the Psychotria phylogeny using the parameters inferred under DEC and DEC+J ML estimation.

A.

True model	Inference model	# sims	mean LnL	mean LnL difference	# false positives in LRT	False Postive Rate
DEC DEC	DEC DEC+J	1000	-22.055 -21.805	0.247	43	0.043

B.

True model	Inference model	# sims	mean LnL	mean LnL difference	# false negatives in LRT	False Negative Rate
DEC+J DEC+J	DEC DEC+J	1000	-24.401 -16.542	7.86	59	0.059

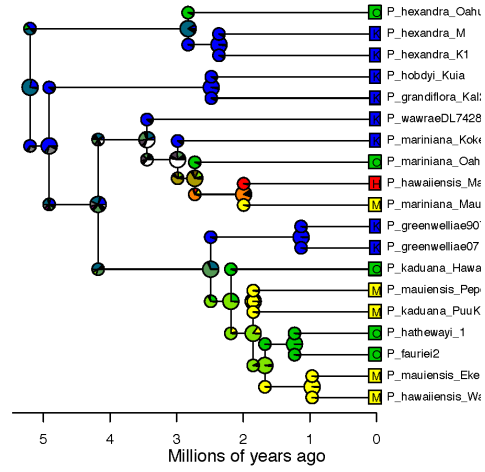| Clade	region	tree age	# tips	Constraints model ID	Description	DEC+J LnL	DEC LnL	LRT pval	sig.	AIC weight ratio	AICc weight ratio
Psychotria	Hawaii	5.2	19	M0	unconstrained	-20.9	-34.5	1.8E-07	***	294895	191952
Psychotria	Hawaii	5.2	19	M1	2 areas max	-21.0	-35.2	9.4E-08	***	566941	371387
Psychotria	Hawaii	5.2	19	M2	2 areas + east disp. only	-16.7	-31.9	3.4E-08	***	1.5E+06	1.0E+06
Psychotria	Hawaii	5.2	19	M2b	4 areas max, east-only disp.	-16.7	-31.4	5.7E-08	***	924075	606221
Psychotria	Hawaii	5.2	19	M3	strat. disp.	-28.3	-39.8	1.6E-06	***	36882	23981
Psychotria	Hawaii	5.2	19	M3B	strat. disp. & areas	-32.8	-41.7	2.4E-05	***	2771	1817
Hawaiian *Drosophila*	Hawaii	20	42	M0	unconstrained	-59.5	-97.1	4.6E-18	***	7.3E+15	6.2E+15
Hawaiian *Drosophila*	Hawaii	20	42	M3areas	strat. disp. & areas	-224.9	-226.2	0.11	ns	1.3	1.1
Hawaiian *Drosophila*	Hawaii	20	42	M3areas	strat. disp. & areas, w. Z	-547.6	-551.2	0.0068	**	14.3	11.5
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M0	unconstrained	-61.9	-103.0	1.2E-19	***	2.6E+17	2.3E+17
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M1	2 areas max	-62.2	-102.3	3.4E-19	***	9.4E+16	8.5E+16
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M1a	2 areas max; adjacency	-62.4	-121.7	1.4E-27	***	2.0E+25	1.7E+25
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M2	2 areas max; east disp. only	-116.4	-133.1	7.6E-09	***	6.5E+06	5.7E+06
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M3b_stratified_w_areas	strat. disp. & areas	-218.9	-224.0	0.0013	**	64.8	52.7
Hawaiian *Scaptomyza5*	Hawaii	12.4	49	M0	unconstrained, w. Z	-68.8	-108.9	3.4E-19	***	9.5E+16	8.3E+16
Hawaiian *Scaptomyza5*	Hawaii	12.4	49	M1	2 areas max, w. Z	-69.2	-108.1	1.2E-18	***	2.7E+16	2.4E+16
Hawaiian *Scaptomyza5*	Hawaii	12.4	49	M1a	2 areas max; adjacency, w. Z	-69.5	-128.5	1.7E-27	***	1.5E+25	1.3E+25
Hawaiian *Scaptomyza5*	Hawaii	12.4	49	M2	2 areas max; east disp. only, w. Z	-106.9	-119.6	4.9E-07	***	1.1E+05	1.1E+05
Hawaiian *Scaptomyza5*	Hawaii	12.4	49	M3b_stratified_w_areas	strat. disp. & areas, w. Z	-273.2	-273.5	0.41	ns	0.5	0.4
Hawaiian *Scaptomyza5*	Hawaii	12.4	49	M3c_stratified_w_areas	strat. disp. & areas, w. Z, low prob.	-505.6	-510.4	0.002	**	44.4	39.0
Hawaiian Honeycreepers	Hawaii	5.78	19	M0	unconstrained	-43.1	-55.2	8.6E-07	***	67028	43695
Hawaiian Honeycreepers	Hawaii	5.78	19	M3_strat	strat. disp.	-59.9	-64.5	0.0026	**	34.8	22.8
Hawaiian Honeycreepers	Hawaii	5.78	19	M3a_strat	strat. disp. + areas	-66.9	-78.7	7.4E-06	***	49078	32048
Megalagrion	Hawaii	9.6	34	M0	unconstrained	-43.4	-69.5	5.1E-13	***	7.8E+10	6.3E+10
Megalagrion	Hawaii	9.6	34	M1_2areas	2 areas max	-43.6	-72.0	4.9E-14	***	7.8E+11	6.3E+11
Megalagrion	Hawaii	9.6	34	M1a_2areas	2 areas max; adjacency	-44.0	-77.7	2.3E-16	***	1.5E+14	1.2E+14
Megalagrion	Hawaii	9.6	34	M2_2areasEastOnly	2 areas max; east disp. only	-46.4	-70.4	4.2E-12	***	9.8E+09	8.0E+09
Megalagrion	Hawaii	9.6	34	M3b_stratified_w_areas	strat. disp. & areas	-79.9	-90.0	7.4E-06	***	8518	6930
Megalagrion	Hawaii	9.6	34	M0	unconstrained	-43.4	-69.5	5.1E-13	***	7.8E+10	6.3E+10
Megalagrion	Hawaii	9.6	34	M1_2areas	2 areas max	-43.6	-72.0	4.9E-14	***	7.8E+11	6.3E+11
Megalagrion	Hawaii	9.6	34	M1a_2areas	2 areas max; adjacency	-44.0	-77.7	2.3E-16	***	1.5E+14	1.2E+14
Megalagrion	Hawaii	9.6	34	M2_2areasEastOnly	2 areas max; east disp. only	-46.4	-70.4	4.2E-12	***	9.8E+09	8.0E+09
Megalagrion	Hawaii	9.6	34	M3b_stratified_w_areas	strat. disp. & areas	-79.9	-90.0	7.4E-06	***	8518	6930
Megalagrion	Hawaii	9.6	34	M3c_stratified_w_areas	strat. disp. & areas, w. Z	-136.0	-145.3	1.5E-05	***	4303	3273
Nesophrosyne	Hawaii	3.2	198	M1 (UM-2)	unconstrained	-255.7	-371.6	2.4E-52	***	8.0E+49	7.7E+49
Nesophrosyne	Hawaii	3.2	198	M1 (UM-2)	2 areas max	-255.8	-376.9	1.3E-54	***	1.5E+52	1.4E+52
Nesophrosyne	Hawaii	3.2	198	M3_6max (TS-U)	unconstrained, strat. disp.	-360.5	-434.8	3.4E-34	***	7.0E+31	6.6E+31
Nesophrosyne	Hawaii	3.2	198	M3_2max (TS-2)	2 areas max; strat. disp. & areas	-497.4	-528.1	4.9E-15	***	7.5E+12	7.7E+12
Nesophrosyne	Hawaii	3.2	198	M3_6max (TS-U)	unconstrained, strat. disp. & areas	-500.2	-537.2	8.3E-18	***	4.1E+15	4.2E+15
Nesophrosyne	Hawaii	3.2	198	M3_6max (TS-Ua)	unconstrained, strat. disp. (nz) & areas	-506.1	-543.4	5.7E-18	***	5.9E+15	5.6E+15
Nesophrosyne	Hawaii	3.2	198	M3_2max (TS-2)	2 areas max; strat. disp. (nz) & areas	-499.8	-534.3	1.0E-16	***	3.5E+14	3.4E+14
Orsonwelles	Hawaii	4.35	12	M0	unconstrained	-13.1	-15.8	0.019	*	5.7	2.5
Plantago	Hawaii	5.1	16	M0	unconstrained	-15.8	-23.2	1.0E-04	***	625	366
Plantago	Hawaii	5.1	16	M3_strat	strat. disp.	-22.0	-25.3	0.01	*	9.8	5.8
Plantago	Hawaii	5.1	16	M3_strat	strat. disp. and areas	-41.5	-44.7	0.012	*	8.7	5.1
Silversword alliance	Hawaii	6.84	29	M0	unconstrained	-48.9	-51.3	0.027	*	4.2	3.3
Silversword alliance	Hawaii	6.84	29	M3_strat	strat. disp., with areas	-53.5	-57.9	0.0032	**	28.8	22.4
Anolis	Caribbean	98.9	190	M0	unconstrained	-242.0	-275.9	1.8E-16	***	1.9E+14	1.9E+14
Cyrtandra	Pacific	41.5	59	M0	unconstrained	-42.5	-47.3	0.0019	**	44.8	40.4
Cyrtandra	Pacific	41.5	59	M1	4 areas max	-42.5	-47.3	0.0019	**	44.8	40.4
Microlophus	Galapagos	3.75	9	M1	2 areas max	-20.5	-34.8	8.2E-08	***	6.5E+05	1.6E+05
Vireya	SE Asia	11	65	M1	2 areas, no other constraints	-214.1	-250.9	1.0E-17	***	3.3E+15	3.2E+15
Vireya	SE Asia	11	65	M1+cc	2 areas + constrained connectivity	-199.7	-219.4	3.6E-10	***	1.3E+08	1.2E+08

88

Figure 1. Various models for the evolution of geographic range at cladogenesis events. The events allowed by LAGRANGE are highlighted in gray. Each allowed event is fixed to have equal probability in the LAGRANGE algorithm. LAGRANGE requires that, at cladogenesis, at least one daughter species must have a range size of 1. This means that, while some vicariance events are allowed, others are not (row 4, columns 3). Widespread sympatry is also disallowed (column 1, rows 2-4) and sympatric-subset speciation where the subset species has a range size larger than one area. Note: The trees shown are just examples of each type of event; an exhaustive enumeration of all of the possibilities would require a much larger figure: for a four-area ancestor, if the one-area restriction is maintained for one of the daughters, there are eight possible vicariance events, and eight possible sympatric-subset events. (The number is eight, not four, since the left and right daughters are not exchangeable.) If the one-area restriction is lifted, there are 28 possible vicariance events, and 28 possible sympatric-subset events, for the four-area ancestor. Founder-event speciation, with the dispersing lineage initially restricted to a single area, is implemented in this paper (column 4).

89

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. Two important notes in interpreting these graphics: (1) These depict the best estimates of the ancestral range at each node and each "corner" (the corner positions are used to represent the geographic range immediately after a cladogenesis event). The best estimate at each node is not identical with the single best joint history (Felsenstein 2004), so in some instances, the ancestral ranges cannot be read naively as a joint history. (2) Unlike LAGRANGE, BioGeoBEARS by default estimates ancestral states under the global ML model; LAGRANGE estimates ancestral states by conducting an ML search for the best model, conditional on fixing each possible state at each node in the tree. The resulting ancestral states tend to be highly similar, but they are not necessarily identical. 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.

BioGeoBEARS DEC on Psychotria M0
anstates: global optim, 4 areas max. d=0.035; e=0.028; j=0; LnL=−34.5
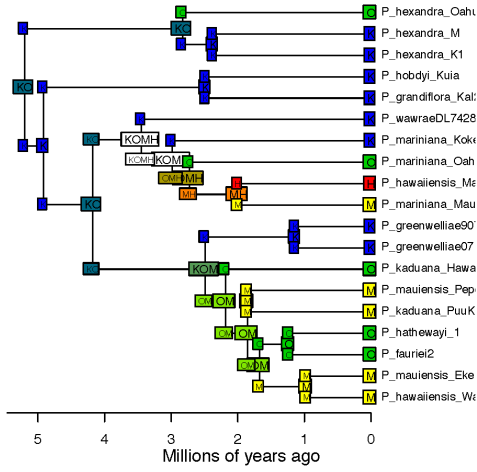
BioGeoBEARS DEC on Psychotria M0
anstates: global optim, 4 areas max. d=0.035; e=0.028; j=0; LnL=−34.5

BioGeoBEARS DEC+J on Psychotria M0
anstates: global optim, 4 areas max. d=0; e=0; j=0.114; LnL=−20.9
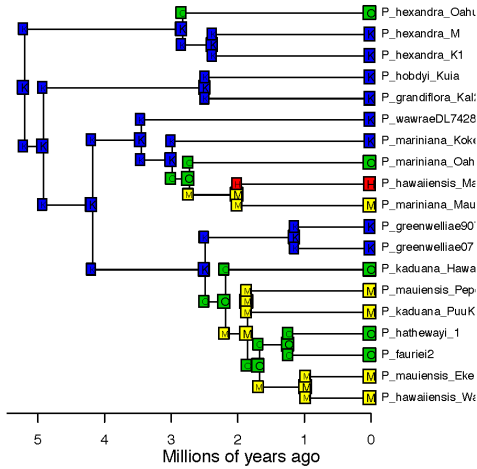
BioGeoBEARS DEC+J on Psychotria M0
anstates: global optim, 4 areas max. d=0; e=0; j=0.114; LnL=−20.9
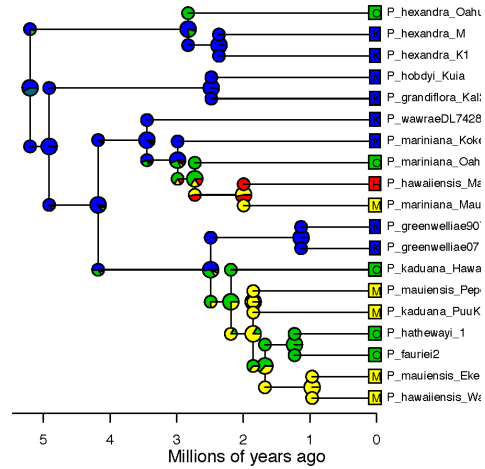
Figure 3. The log-likelihood advantage of DEC+J inference over DEC, when the true model is DEC (top) or DEC+J (bottom).

DEC+J advantage with DEC−simulated data
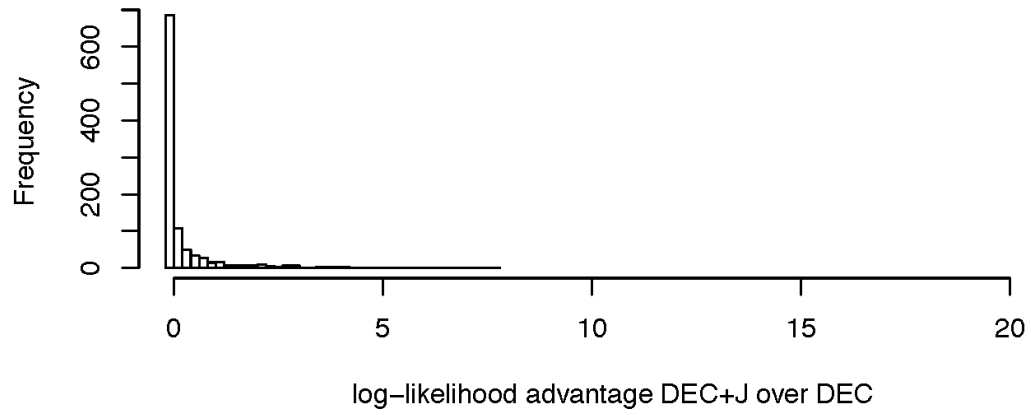
DEC+J advantage with DEC+J−simulated data
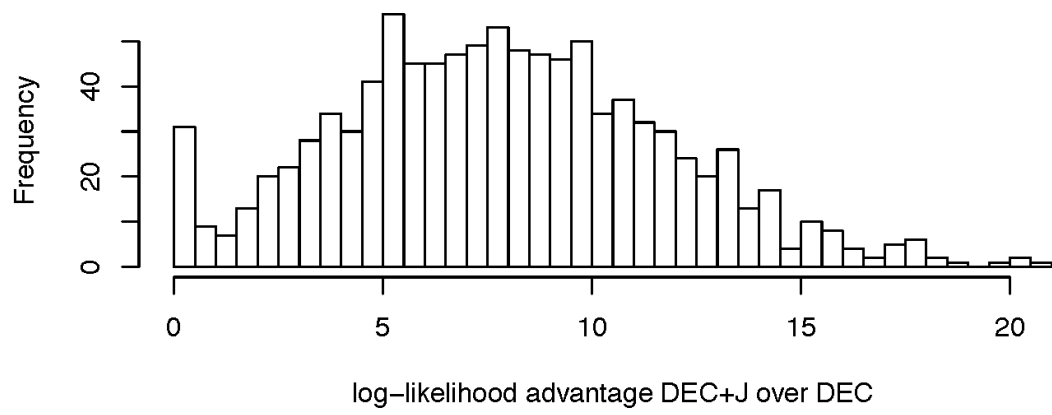
Figure 4. Histograms of the accuracies of ancestral state inference for each pair of simulation and inference under DEC and DEC+J. Accuracy equals the fraction of time that the most probable inferred ancestral state matches the true, simulated ancestral state.

simulations: DEC
inference: DEC
average fraction of nodes correct=0.78

simulations: DEC
inference: DECJ
average fraction of nodes correct=0.83

simulations: DECJ
inference: DEC
average fraction of nodes correct=0.57

simulations: DECJ
inference: DECJ
average fraction of nodes correct=0.87
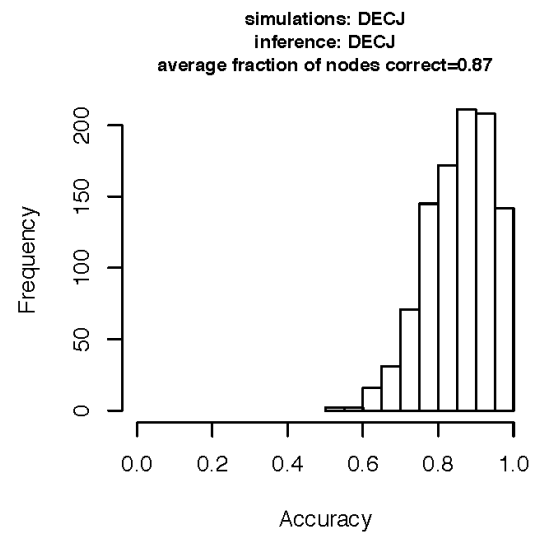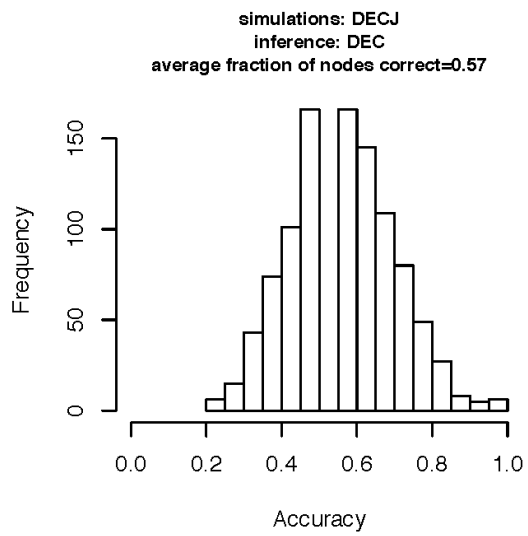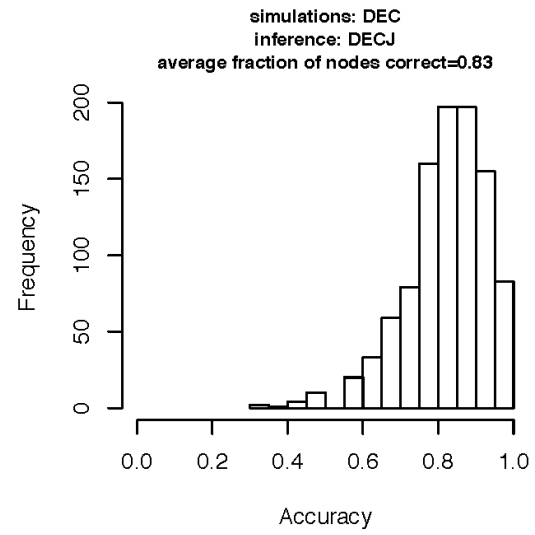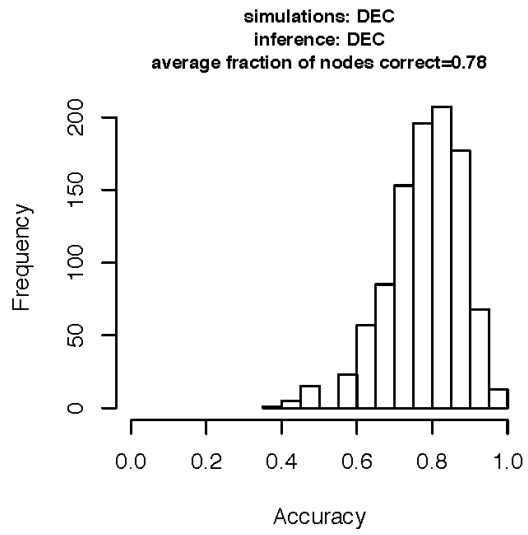
Figure 5. Mean S for each node age, plotted against node age. S equals mean similarity to truth, i.e. 1-(mean absolute difference in probability between inference and truth). The x-coordinates for the same node have been jittered to reduce overlap.

Legend
1. sim: DEC,   inf: DEC
2. sim: DEC,   inf: DEC+J
3. sim: DEC+J, inf: DEC
4. sim: DEC+J, inf: DEC+J

mean similarity to truth

Node age (Ma)
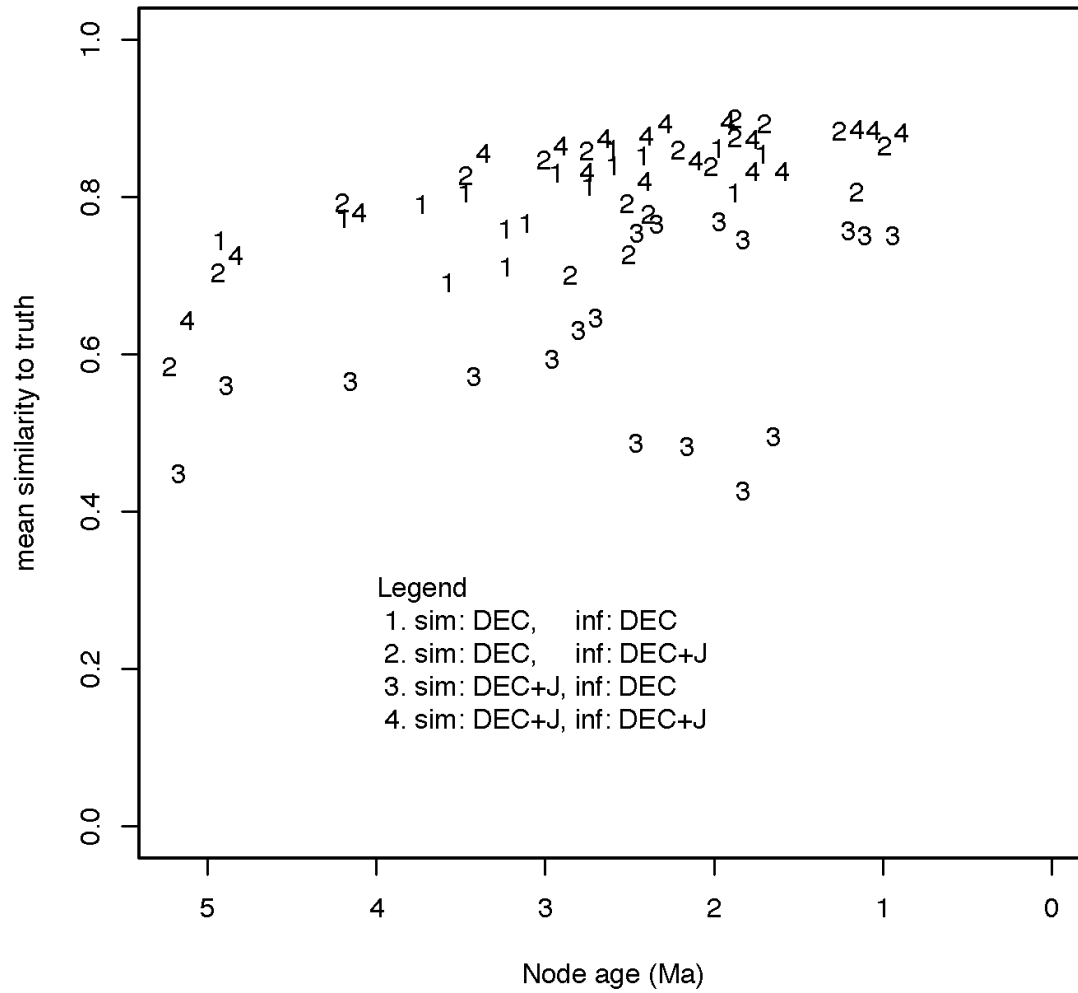
Figure 6. Accuracy of parameter inference under DEC and DEC+J, when the history was simulated under each of these models. The true parameter used in the simulations is shown by the dashed line. These parameter values were the ML estimates in the original M0 analysis of *Psychotria* under DEC and DEC+J.

sim: DEC, inf: DEC

Frequency (y-axis), inferred d (x-axis)

sim: DEC, inf: DEC

Frequency (y-axis), inferred e (x-axis)

sim: DEC, inf: DEC+J

Frequency (y-axis), inferred d (x-axis)

sim: DEC, inf: DEC+J

Frequency (y-axis), inferred e (x-axis)

sim: DEC, inf: DEC+J

Frequency (y-axis), inferred j (x-axis)

sim: DEC+J, inf: DEC

Frequency (y-axis), inferred d (x-axis)

sim: DEC+J, inf: DEC

Frequency (y-axis), inferred e (x-axis)

sim: DEC+J, inf: DEC+J

Frequency (y-axis), inferred d (x-axis)

sim: DEC+J, inf: DEC+J

Frequency (y-axis), inferred e (x-axis)

sim: DEC+J, inf: DEC+J

Frequency (y-axis), inferred j (x-axis)

Figure 7. 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 Supplemental Files); nz, zeros in the manual dispersal probability multiplier matrix have been replaced with a small nonzero value.

Psychotria: unconstrained
Psychotria: 2 areas max
Psychotria: 2 areas + east disp. only
Psychotria: 4 areas max, east–only disp.
Psychotria: strat. disp.
Psychotria: strat. disp. & areas
Hawaiian Drosophila: unconstrained
Hawaiian Drosophila: strat. disp. & areas
Hawaiian Drosophila: strat. disp. & areas, w. Z
Hawaiian Scaptomyza: unconstrained
Hawaiian Scaptomyza: 2 areas max
Hawaiian Scaptomyza: 2 areas max; adjacency
Hawaiian Scaptomyza: 2 areas max; east disp. only
Hawaiian Scaptomyza: strat. disp. & areas
Hawaiian Scaptomyza5: unconstrained, w. Z
Hawaiian Scaptomyza5: 2 areas max, w. Z
Hawaiian Scaptomyza5: 2 areas max; adjacency, w. Z
Hawaiian Scaptomyza5: 2 areas max; east disp. only, w. Z
Hawaiian Scaptomyza5: strat. disp. & areas, w. Z
Hawaiian Scaptomyza5: strat. disp. & areas, w. Z, low pro
Hawaiian Honeycreepers: unconstrained
Hawaiian Honeycreepers: strat. disp.
Hawaiian Honeycreepers: strat. disp. + areas
Megalagrion: unconstrained
Megalagrion: 2 areas max
Megalagrion: 2 areas max; adjacency
Megalagrion: 2 areas max; east disp. only
Megalagrion: strat. disp. & areas
Megalagrion: unconstrained
Megalagrion: 2 areas max
Megalagrion: 2 areas max; adjacency
Megalagrion: 2 areas max; east disp. only
Megalagrion: strat. disp. & areas
Megalagrion: strat. disp. & areas, w. Z
Nesophrosyne: unconstrained
Nesophrosyne: 2 areas max
Nesophrosyne: unconstrained, strat. disp.
Nesophrosyne: 2 areas max; strat. disp. & areas
Nesophrosyne: unconstrained, strat. disp. & areas
Nesophrosyne: unconstrained, strat. disp. (nz) & areas
Nesophrosyne: 2 areas max; strat. disp. (nz) & areas
Orsonwelles: unconstrained
Plantago: unconstrained
Plantago: strat. disp.
Plantago: strat. disp. and areas
Silversword alliance: unconstrained
Silversword alliance: strat. disp., with areas
Anolis: unconstrained
Cyrtandra: unconstrained
Cyrtandra: 4 areas max
Microlophus: 2 areas max
Vireya: 2 areas, no other constraints
Vireya: 2 areas + constrained connectivity

1   1e+10   1e+20   1e+30   1e+40   1e+50

DECJ:DEC AICc weight ratio

Figure 8. ML parameter estimates under DEC and DEC+J for parameters d (grey), e (white), and j (black). For each parameter, results are plotted in the same order as in Figure I and Table J. The * represents a bar that has been cut off for display purposes; the true value for this bar is j=3 (the taxon is Galapagos *Microlophus*), DEC+J analysis; a value of j=3 means that for this taxon, ML inference estimated that founder-event speciation events receive 100% of the probability at cladogenesis, and the other types of events (sympatric-range copying, sympatric-subset, and vicariance) receive 0%.

DEC   Model   DEC+J

d: rate of dispersal/
range expansion

e: rate of extinction/
range contraction

j: weight of jump dispersal/
founder–event speciation

*

2   1   0   1   2
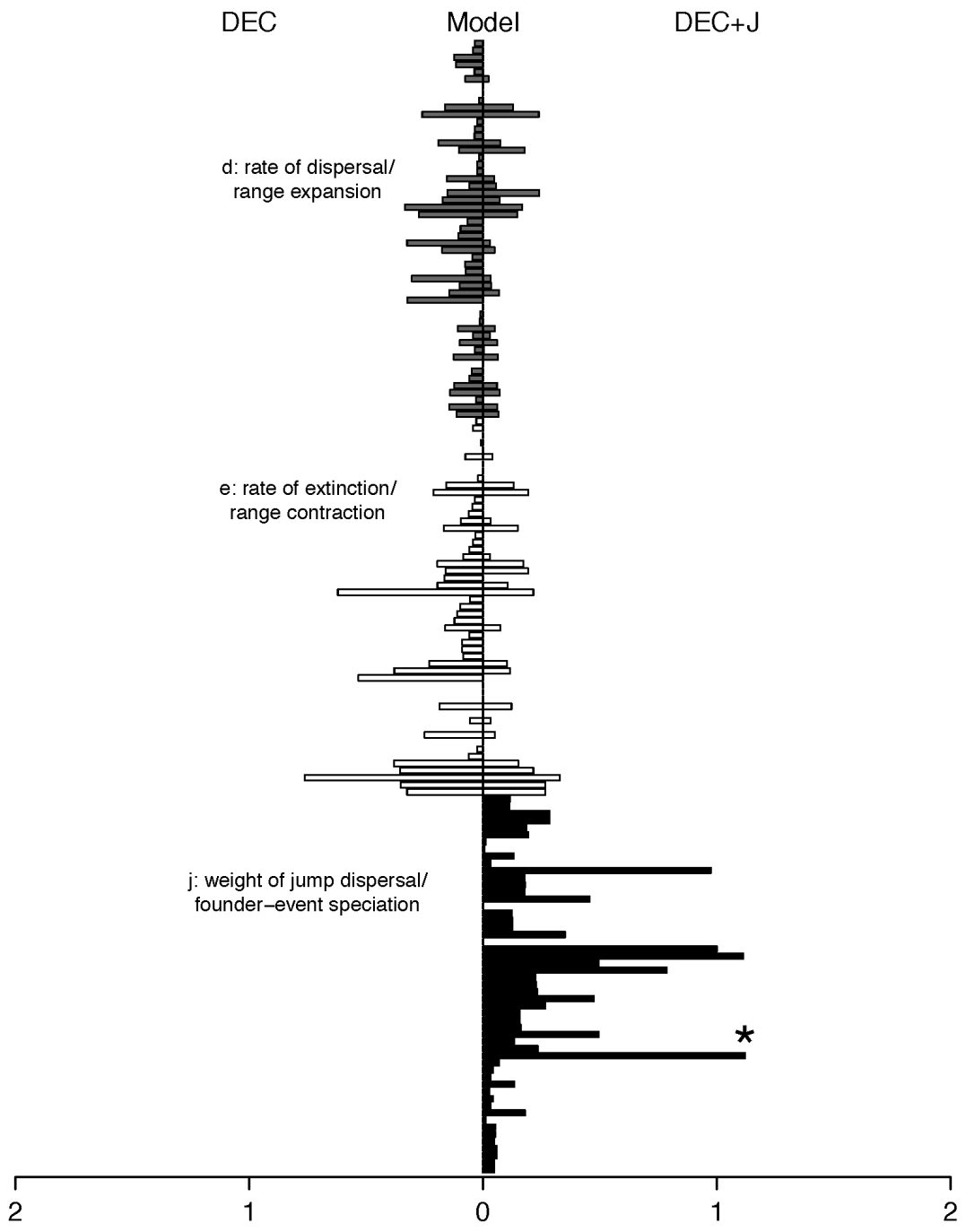
Figure 9. Global ancestral state estimates under M0 (unconstrained) for the DEC and DEC+J models run on Hawaiian Honeycreepers. Left: the single most-probable ancestral range. Right: pie charts show the relative probability of all 16 states/possible geographic ranges. For important notes about interpreting these figures, see the caption for Figure 2. 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.

BioGeoBEARS DEC on Honeycreepers M0
anstates: global optim, 4 areas max. d=0.174; e=0.164; j=0; LnL=−55.2
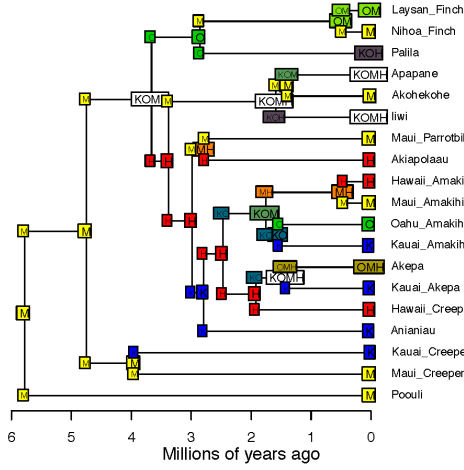
BioGeoBEARS DEC on Honeycreepers M0
anstates: global optim, 4 areas max. d=0.174; e=0.164; j=0; LnL=−55.2
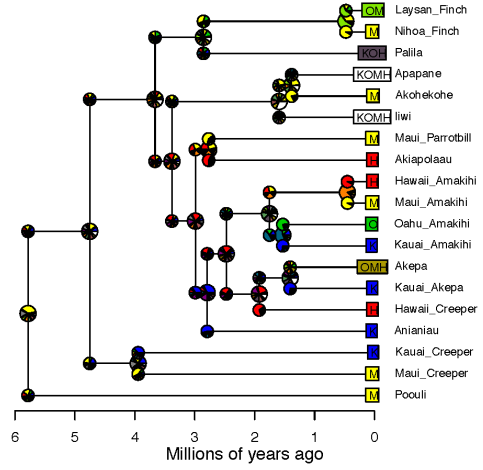
BioGeoBEARS DEC+J on Honeycreepers M0
anstates: global optim, 4 areas max. d=0.072; e=0; j=1.114; LnL=−43.1

BioGeoBEARS DEC+J on Honeycreepers M0
anstates: global optim, 4 areas max. d=0.072; e=0; j=1.114; LnL=−43.1

107
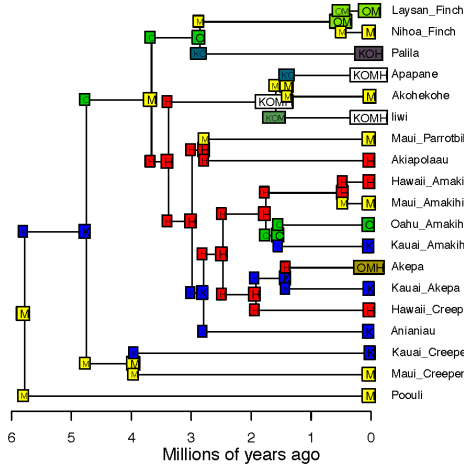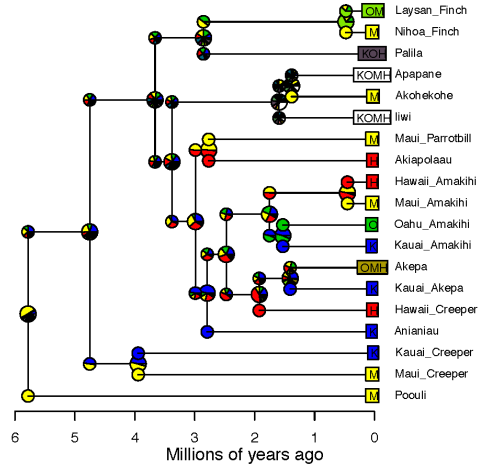
Figure 10. Global ancestral state estimates under M1 (10 islands, maximum range size = 2 areas) for the DEC and DEC+J models run on Galapagos *Microlophus*. Left: the single most-probable ancestral range. Right: pie charts show the relative probability of all 16 states/possible geographic ranges. For important notes about interpreting these figures, see the caption for Figure 2. Although *j*=3 and *d*=*e*=0, the fact that there is a single tip with a wider range (*M. albemarlensis*) is explained by the model as putting some probability on ancestors with that range, with the widespread species regularly budding new species via founder events. This is an example of how the single most-probable states (left side) cannot always be read naively as the best joint estimate of the history (Felsenstein 2004). The pie charts (right side) are more informative of how the probabilities are combining down the tree. Key: A, Espanola; B, San Cristobal; C, Santa Cruz; D, Pinzon; E, Floreana; F, Pinta; G, Santiago; H, Isabella; I, Marchena; J, Fernandina. Ranges that are combinations of these 10 areas have a color that is a mix of the component area colors.

BioGeoBEARS DEC on Microlophus M1_2areas
anstates: global optim, 2 areas max. d=0.324; e=0.533; j=0; LnL=−34.8
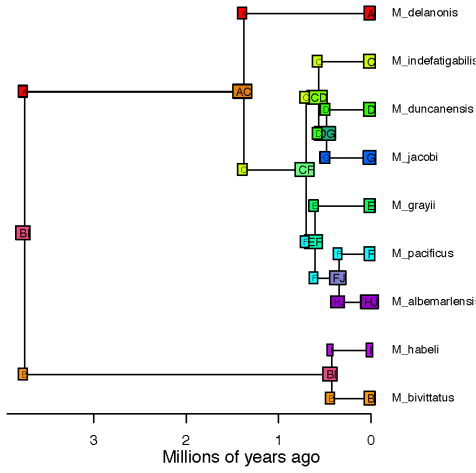
BioGeoBEARS DEC on Microlophus M1_2areas
anstates: global optim, 2 areas max. d=0.324; e=0.533; j=0; LnL=−34.8
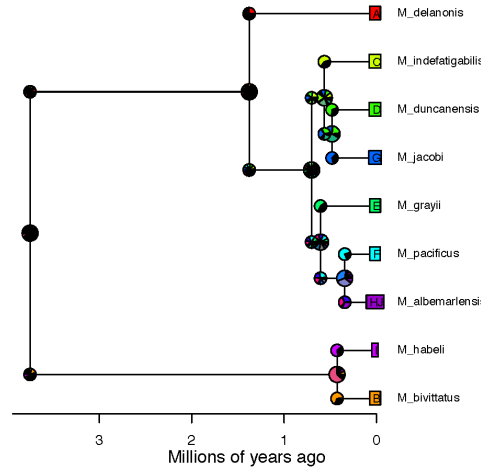
BioGeoBEARS DEC+J on Microlophus M1_2areas
anstates: global optim, 2 areas max. d=0; e=0; j=3; LnL=−20.5
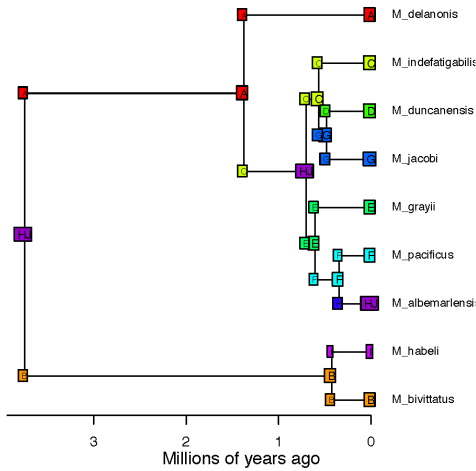
BioGeoBEARS DEC+J on Microlophus M1_2areas
anstates: global optim, 2 areas max. d=0; e=0; j=3; LnL=−20.5
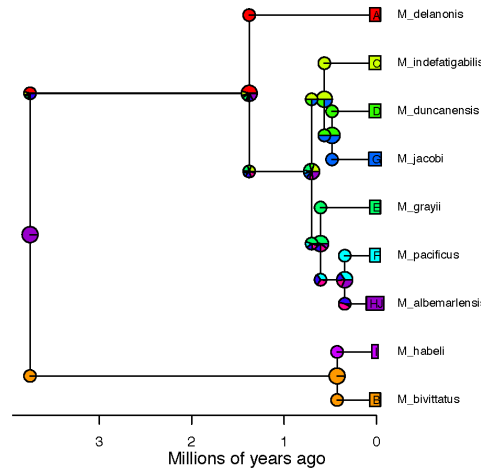
Figure 11. The hypothesized explanation for improved data likelihoods under the DEC+J model, for island taxa. Waiting times to get, and keep, the "correct" ancestor at uncorrelated cladogenesis events reduce probabilities, whereas under DEC+J with $d=e=0$, the probability of no change along branches is 1, and dispersal is perfectly correlated with speciation.

Example of a typical history, DEC+J model

Founder event

When *d=e=0.0* P(no change)=1

Example of a typical history, DEC model

Waiting times

Vicariance event

Waiting time

Range expansion event

Waiting time
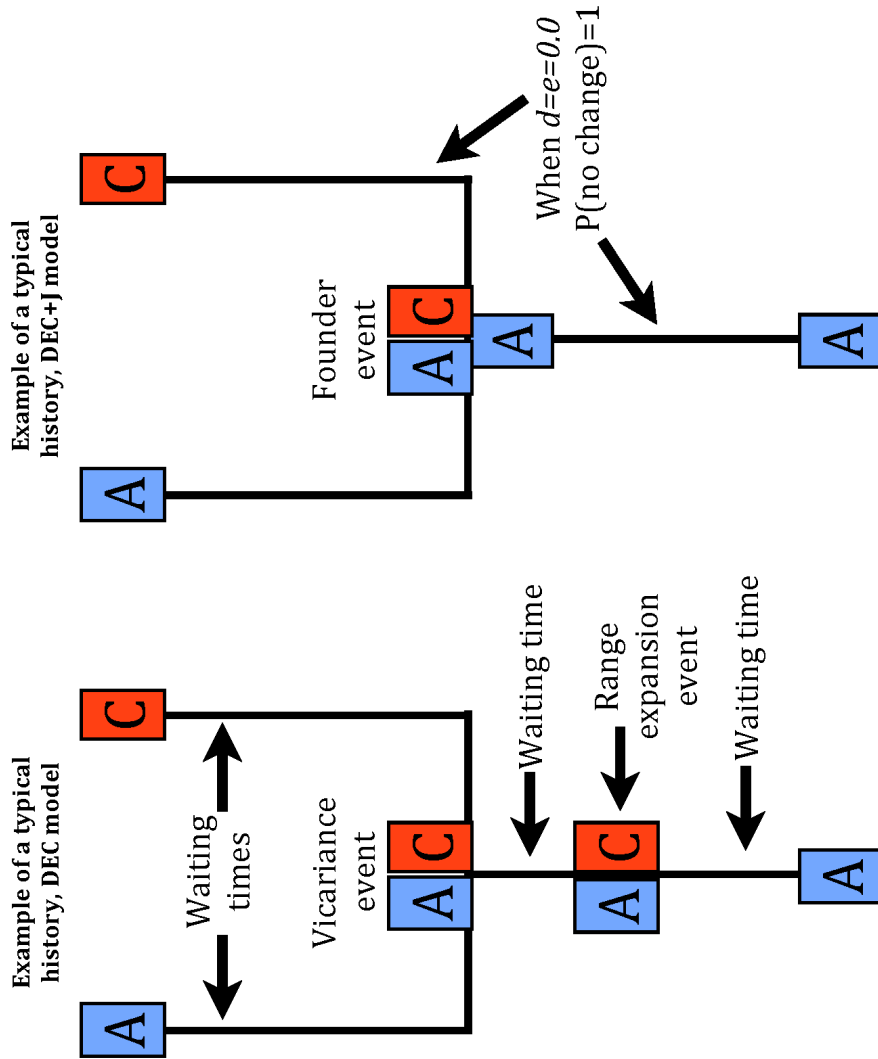
Supplementary Figures 1: Bayesian Analysis

Supplementary Figure 1.1. Inference of parameters for the *Psychotria* M0 dataset, under the DEC model, using random-walk Metropolis MCMC.

Supplementary Figure 1.2. Trace and histogram of the log-posterior probabilities (LP) of the DEC model, using random-walk Metropolis MCMC.

Supplementary Figure 1.3. Inference of parameters for the *Psychotria* M0 dataset, under the DEC+J model, using random-walk Metropolis MCMC.

Supplementary Figure 1.4. Trace and histogram of the log-posterior probabilities (LP) of the DEC+J model, using random-walk Metropolis MCMC.
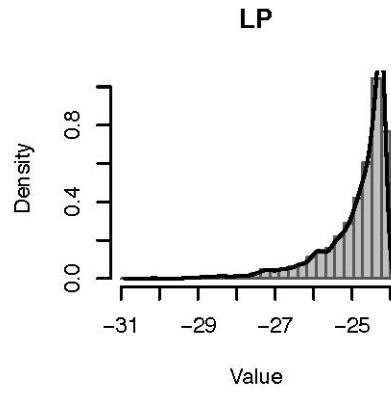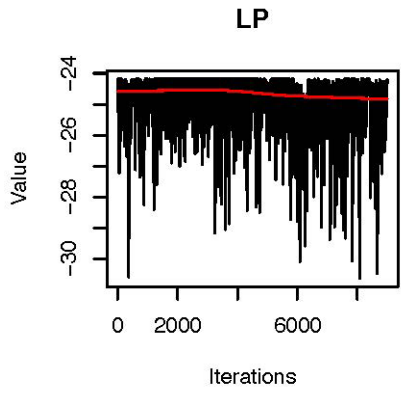
Supplementary Figure 1.5. Inference of parameters for the *Psychotria* M0 dataset, using reversible-jump MCMC to sample both the DEC and the DEC+J model.
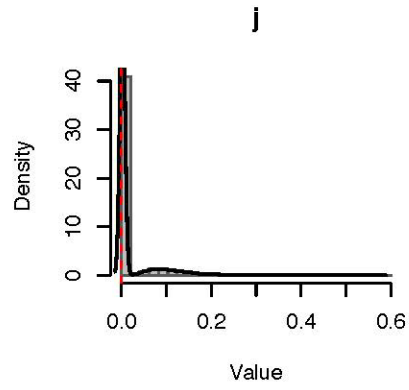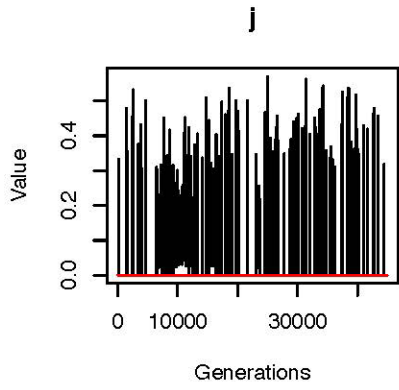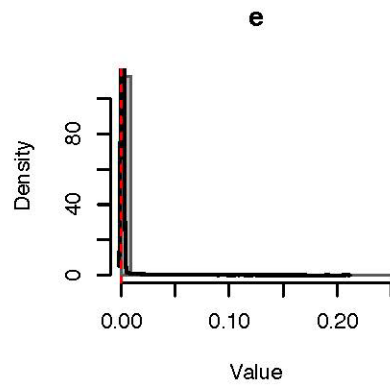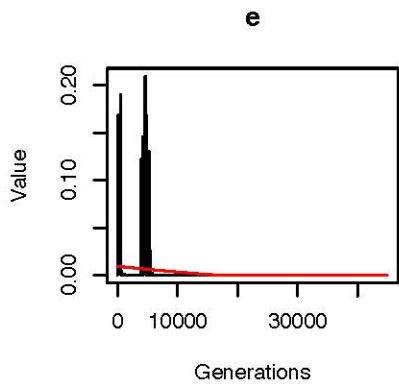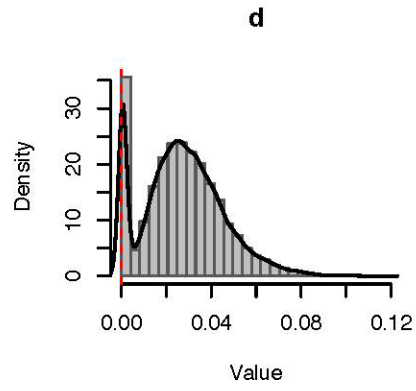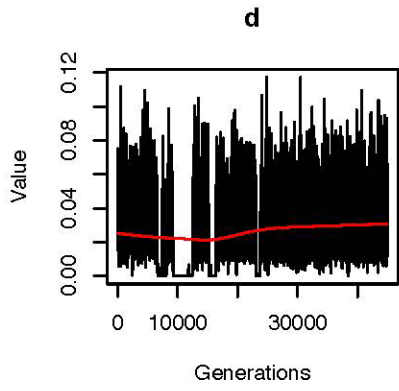
Supplementary Figure 1.6. Trace and histogram of the log-posterior probabilities (LP) when using reversible-jump MCMC to sample both the DEC and the DEC+J model.
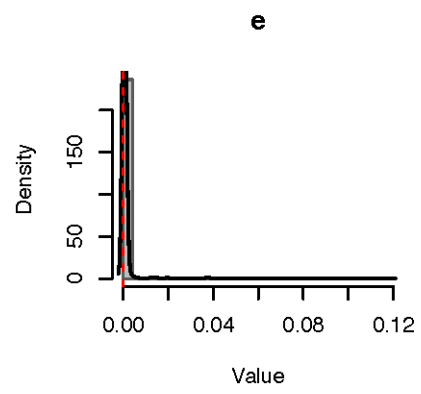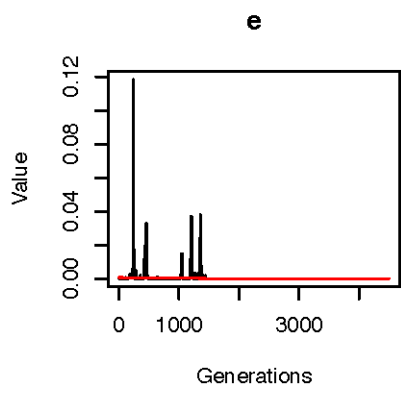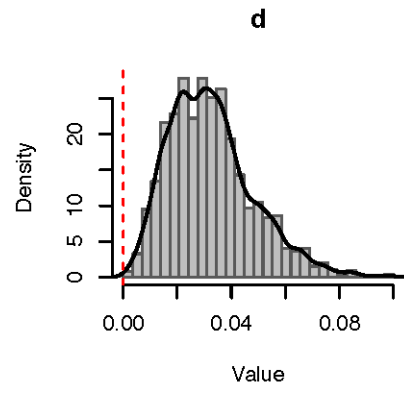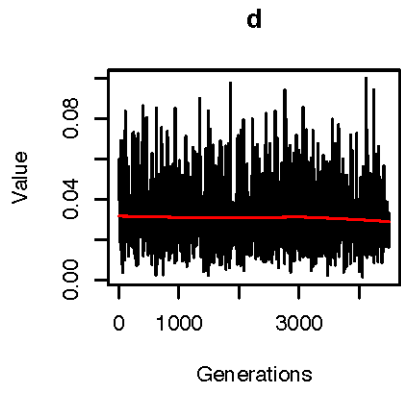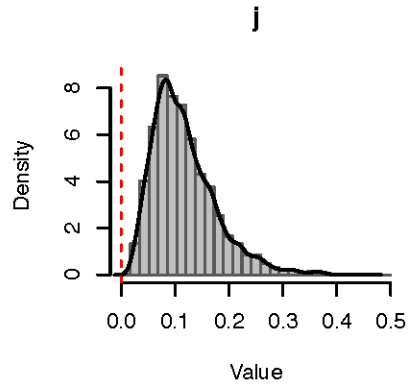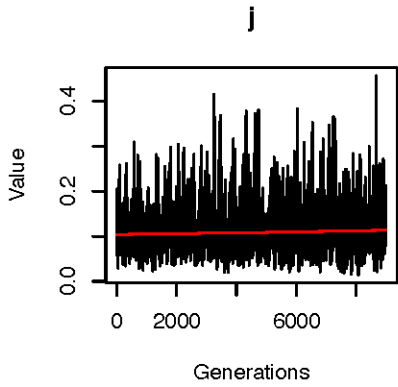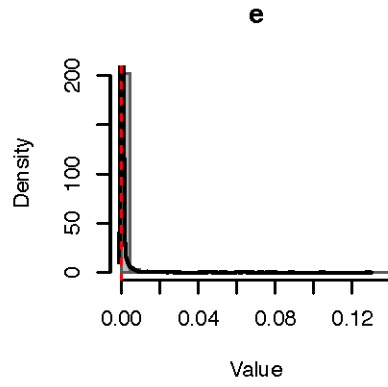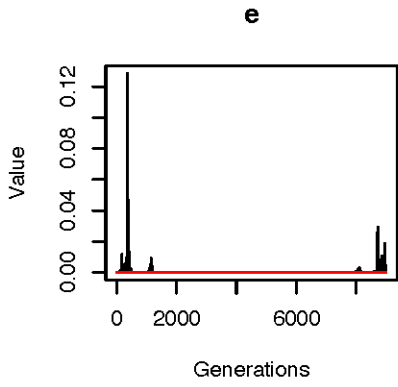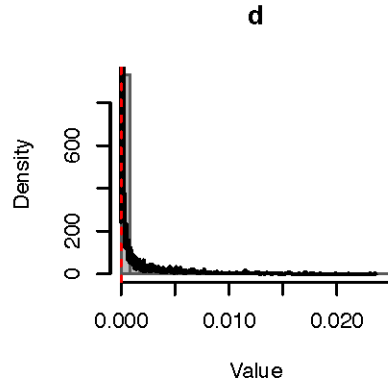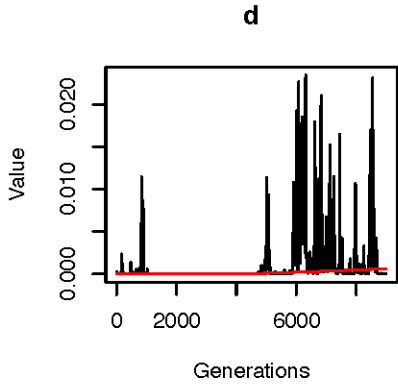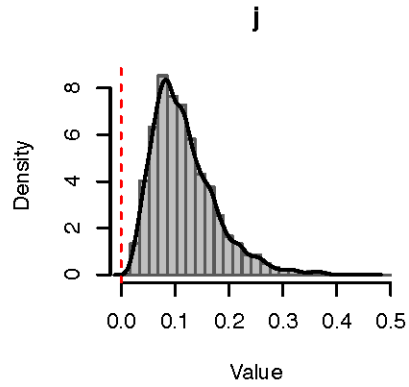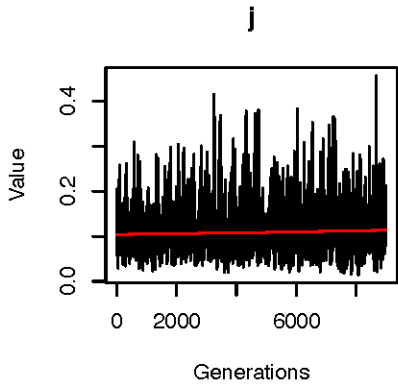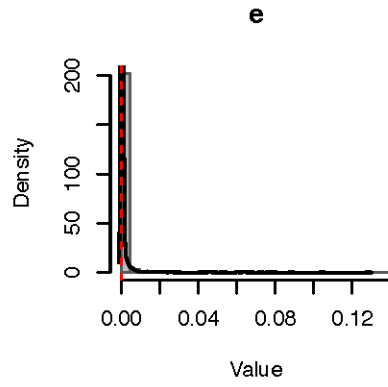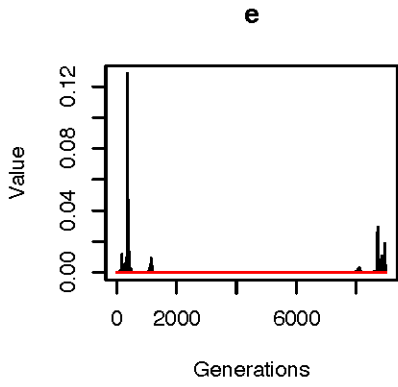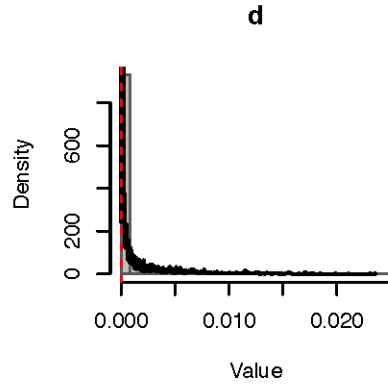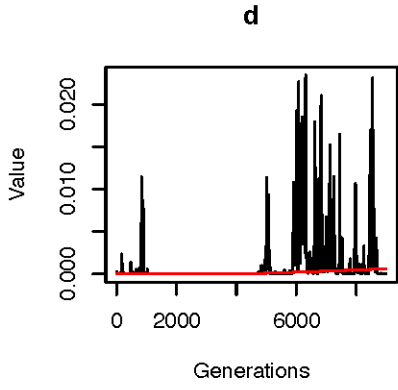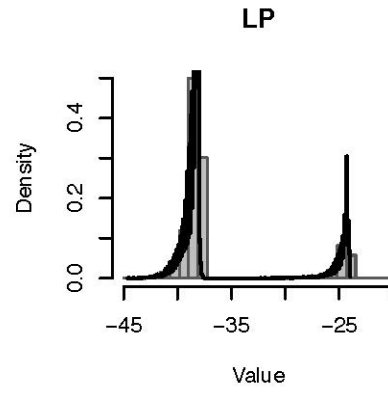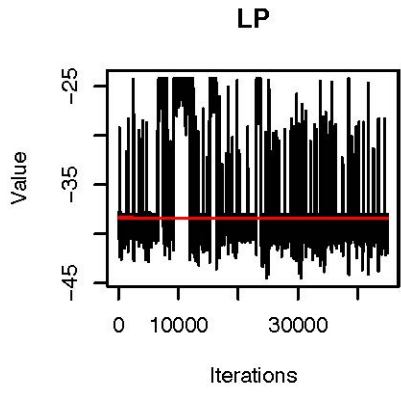
Chapter 3: Model selection reveals differences in cladogenesis processes operating in island versus continental clades

Structured Abstract

Aim. To compare the three major cladogenesis models used in the historical biogeography inference programs LAGRANGE, DIVA, and BayArea, to test the effects of adding founder-event speciation to these models, and to test for differences in cladogenesis processes between island clades and non-island (continental and oceanic) clades.

Location. Clades are sampled globally, but the regions emphasized are the Pacific islands, especially Hawaii; clades from sand-dune systems in the desert Southwest of North America; and clades from the Neotropics.

Methods. Likelihood versions of the DEC, DIVA, and BayArea cladogenesis models were implemented in the R package BioGeoBEARS, along with "+J" versions of these models. These six models were run on 63 clades/constraints combinations. Models were compared using the likelihood-ratio test and AICc model weights.

Results. Almost all analyses, including continental clades, strongly favored the "+J" models over the models without founder-event speciation. However, founder-event speciation was measurably less frequent in non-island analyses, being 2-4 times weaker than in analyses of island clades. One clade was found ("*Taygetis* clade" butterflies from the Neotropics) which favored the DEC model over all others.

Main conclusions. Formal model selection procedures can be applied in phylogenetic inferences of historical biogeography, and the relative importance of different processes can be measured. These techniques have great potential for strengthening quantitative inference in historical biogeography.

Keywords (6-10, alphabetical order)

Historical biogeography, LAGRANGE, DIVA, BayArea, cladogenesis, founder-event speciation

Introduction

It is a commonplace that the biogeographical processes operating on continents tend to be different than those operating on oceanic islands (Gillespie & Baldwin 2010; Lomolino, Lomolino & Lomolino 2010). Oceanic islands are geologically recent, of small area, and extremely isolated, and these facts likely restrict the processes that might be important. With few exceptions (e.g., Heads 2012) conventional wisdom is that oceanic islands are dominated by the effects of long-distance dispersal (Darwin 1859; Carlquist 1965; Cowie & Holland 2006; Gillespie *et al.* 2012). In particular, founder-event speciation (Paulay & Meyer 2002; Templeton 2008) seems to be fundamental on islands.

The historical biogeography of islands has received a great deal of attention, in part due to the relative tractability of analysis and geological reconstruction in these systems. The fact that significant generalizations, such as the progression rule (Hennig 1966), are often confirmed, gives the subject a degree of predictability unusual in historical evolutionary research. Islands have thus been extremely influential in historical biogeography (Lomolino, Lomolino & Lomolino 2010). However, given that continents are opposite from islands in the above-described traits (geologically ancient, large area, connected now and/or in the past), it is an open question whether or not the processes important in the historical biogeography of island clades will prove important in continental clades. The prior expectations of researchers on this question would probably correlate strongly with

121

their positions in the dispersal-vicariance debate. The debate is massive and cannot be reviewed here; readers are referred to recent surveys of the issue from both sides (de Queiroz 2005; Santos 2007; Heads 2012; Waters *et al.* 2013). Clearly, statistical methods that measure the degree of support that the observed data gives to each of the hypothesized processes would be valuable.

Parametric historical biogeography is a new analytical paradigm (Ree & Sanmartín 2009; Sanmartín 2012) that has great potential for advancing the discussion of the biogeographical processes important on islands versus continents. However, it has not yet been applied to this question. Parametric historical biogeography is based on probabilistic event-based models of how geographic ranges evolve anagenetically and cladogenetically. These models allow the likelihood of the geographic range data at the tips of the tree to be calculated under different model parameters, and the maximum likelihood (ML) estimate of the parameters made (Ree & Smith 2008). With the application of formal model-testing procedures to these models (Matzke 2013d; Matzke 2013a), numerous new questions can be asked. For example, Matzke (2013d) compared the Dispersal-Extinction-Cladogenesis (DEC) model implemented in the program LAGRANGE (Ree *et al.* 2005; Ree & Smith 2008; Smith 2010; Ree 2013) with a new model, DEC+J, which added founder event speciation as an additional cladogenetic process. Thus the computational tools now exist to ask questions such as, "What is the relative importance of founder-event speciation in island clades versus continental clades?",

and to obtain an answer that is quantitative and objectively repeatable by other analysts. This is one of two major questions addressed in this paper.

Another important use of model selection procedures is to assess inference methods that are in use or newly proposed. Any computational method for inference of ancestral states on a phylogeny, including states such as geographic range, is generating an event-based (Sanmartin 2007) model that makes a variety of explicit assumptions about what kinds of events will be allowed when estimating history. This is true whether the inference machinery is parsimony, maximum likelihood, or Bayesian. It is also true whether or not the user in question correctly understands what has been implemented in the computer code, and whether or not the assumptions in the code match the assumptions users are making in their conceptual model of the system (Matzke 2013d). Without model testing procedures, the best researchers can do is run the different available methods, note where results differ, and either conclude with the bare fact that when models differ the results differ, or reach a qualitative conclusion about which models are preferable, based on data or assumptions outside of those used in the computational analysis. (Examples might be fossil data, or assumptions about how widespread ancestral ranges may be.) Such studies are certainly worthwhile, but generalizing from them to make broad conclusions about which methods should be preferred is difficult. The results of any one study may be clade-specific, and usually there are few additional "ground truth" data about the true ancestral states, beyond that put

into the analysis in the first place. Examples of method-comparison studies include

Clark et al. (2008), Buerki et al. (2011), Springer et al. 2011, and Ali et al. (2012).

Existing methods in historical biogeography

Perhaps because of the difficulty in objectively choosing the best methods,

computational methods in historical biogeography are proliferating rather than

being whittled down. The three most popular methods are probably Dispersal-

Vicariance Analysis (DIVA; Ronquist 1996; Ronquist 1997), Likelihood Analysis of

Geographic Range Evolution (LAGRANGE; Ree *et al.* 2005; Ree & Smith 2008; Smith

2010; Ree 2013)), and treating biogeographic range as a simple unordered

morphological character under parsimony, ML, or Bayesian inference (discussed in

Matzke 2013a). The last of these is not really a biogeographic model, and in any

case it is often not useable in biogeography, as it requires all operational taxonomic

units (OTUs) in the phylogeny to have a range size of a single area. In addition to the

above approaches, a variety of pseudo-Bayesian methods run DIVA or LAGRANGE

on the posterior distribution of trees output by a Bayesian MCMC analysis (Nylander

et al. 2008; Yu, Harris & He 2010; Buerki *et al.* 2011; Ali *et al.* 2012; Wood *et al.*

2013). SHIBA (Webb & Ree 2012) uses a LAGRANGE-DEC-like model, but adds ideas

inspired by the theory of island biogeography (MacArthur & Wilson 1967), for

example, the hypothesis that dispersal varies inversely with distance, and that

extinction varies inversely with area size. The inference method is similar to

approximate Bayesian computation (ABC) in that the model is used to generate

forward simulations from the root of the phylogeny to the tips, and the simulations

that, by chance, produce tip species with ranges matching those of the real, observed

species ranges are treated as samples from the posterior distribution. This method

is naturally very inefficient and may only be used on a small number of areas (Webb

& Ree 2012; Landis *et al.* 2013).

Another new program is RASP (Yu, Harris & He 2013), which re-uses the DIVA and

LAGRANGE libraries to allow these algorithms to be run on a collection of trees from

the posterior sample of a Bayesian phylogenetic analysis. However, RASP also

implements a new biogeographic inference method, Bayesian Binary MCMC (BBM).

BBM has been used (Ali *et al.* 2012) but is not yet described in a publication (a

submitted manuscript is mentioned but not available at the RASP website). From

the software manual (Yu, Harris & He 2013), BBM appears to encode geographic

range as a series of presences and absences in each discrete area, i.e., binary

character encoding with multiple characters (Springer *et al.* 2011), also known as

presence coding ((Hardy & Linder 2005). Each area is then treated as an

independent presence/absence character, and the history of each is estimated

separately. This method of coding complex characters like biogeography has a long

history in parsimony studies of highly polymorphic characters (Mickevich & Mitter

1981) but has always been plagued by the fact that the independence assumption

means that impossible ancestors (i.e., an ancestor that is absent everywhere) may

be reconstructed at nodes on the tree, depending on the configuration of the tip data

(Hardy & Linder 2005). RASP attempts to deal with this by adding two "virtual

outgroups" to the study phylogeny, and giving these artificial taxa one of several

suggested ranges. Possible ranges for artificial taxa include (1) a special null range (absent everywhere, perhaps interpretable merely as a lineage being absent from the study region, rather than extinct), (2) present everywhere, which would minimize the chance of impossible ancestors, or (3) the user-specified known ancestral range. Each of these options has obvious difficulties, the greatest being that researchers rarely or never know the ancestral range; rather, the goal is usually to attempt to infer it. Another difficulty is that the concept of using outgroup states to fix the states of the ancestor is derived from older parsimony studies, where the outgroup was sometimes used as a stand-in for the ancestral states. Whatever its merits in parsimony inference, it is not applicable in an ML or Bayesian framework, unless the branch lengths of these outgroups are effectively zero. The RASP authors note in the manual that choices about the geographic range of the virtual outgroups have a large impact on ancestral range estimates, which is not desirable behavior in an inference method.

The main advantage of presence coding is that the independence assumption allows inference on a virtually unlimited number of areas, whereas the other methods are tightly constrained because of the exponential growth of the transition matrix with increasing number of areas (Matzke 2013a). A new method incorporating the best of both is BayArea (Landis *et al.* 2013), which samples branch histories as well as parameter values during MCMC sampling, using stochastic mapping (Nielsen 2002; Huelsenbeck, Nielsen & Bollback 2003). Histories are proposed under the independence assumption, but are evaluated under a dependence model (inspired

by Robinson *et al.* 2003), where probability of dispersal is a function of distance between areas, and where the null range is disallowed. This technique allows inference on a large number of areas, without the need to enumerate all of the possible ranges in computer memory, or to explicitly calculate the probability of all start and end states on branches with matrix exponentiation.

While revolutionary, the initial version of BayArea assumes a purely continuous-time model where all of the changes in geographic range occur via dispersal and extinction events along branches. Consideration of the cladogenesis process is left out; incorporating range evolution at cladogenesis may be possible, but it requires sampling possible range histories at cladogenesis, and this will require sophisticated proposal mechanisms. For example, to enable switching from a vicariance history at a node, to a sympatric history, will require proposal mechanisms to allow the MCMC algorithm to move between each possible histories; this is much more complicated than the proposal mechanisms required for sampling branch histories. As a result, for the moment, BayArea's cladogenesis "model" is, de facto, perfect sympatry: at every speciation event, the ancestral geographic range, whether narrow or widespread, is exactly copied to both daughter branches with probability 1.

Thus, BayArea's cladogenesis assumption differs strongly from the assumption made by the LAGRANGE DEC model, which assigns equal probability to several allowed cladogenesis events, namely, (1) sympatric range-copying only for ancestral areas of size one, (2) sympatric-subset speciation (one daughter species starts in

one area inside the ancestral range, the other daughter inherits the full ancestral range), and (3) vicariance, where one daughter occupies a single area, and the other daughter occupies the remainder. The cladogenesis models of BayArea and LAGRANGE are both different from the cladogenesis model assumed by DIVA, which disallows sympatric-subset speciation, and allows vicariance, but with no constraints on the range size of the daughters; DIVA allows equal splitting of ranges (e.g., ABCD → AB, CD), but LAGRANGE DEC does not (allowing only e.g. ABCD → A, BCD). Finally, the BioGeoBEARS (Matzke 2013d; Matzke 2013a) model option DEC+J allows

founder-event speciation to be added to any of these models. Further discussion of these models, with graphical depictions, is available elsewhere (Ronquist *et al.* 2012; Matzke 2013d; Matzke 2013a).

Graphical depictions of the cladogenesis events allowed by LAGRANGE DEC, DIVA, BayArea, and the "+J" model option of BioGeoBEARS are shown in Figure 1. Between them, these cover the range of cladogenesis models that are currently in use. SHIBA, RASP, etc., essentially re-use one of these cladogenesis models (RASP's BBM model makes the same cladogenesis assumption as BayArea). Which is the correct model? Or, more carefully stated, since all models are wrong -- paraphrasing George Box (Box & Draper 1987) -- which models are most useful for explaining the geographic range data of various clades?

Here, this question is assessed by implementing the DEC, DIVA, and BAYAREA

cladogenesis models in BioGeoBEARS, and using the "+J" option to also implement

DEC+J, DIVA+J, and BAYAREA+J models. These six models are then run on the

sample of island taxa used in a previous study comparing just DEC and DEC+J

(Matzke 2013d). To these taxa is added a sample of continental clades, to enable the

islands-versus-continents comparison, in order to assess (1) the relative importance

of jump dispersal in these two categories, and (2) the relative quality of DEC, DIVA,

and BAYAREA cladogenesis models, with and without jump dispersal added.

Materials and Methods

The implementation of the DEC and DEC+J models in BioGeoBEARS has been

described previously (Matzke 2013d). The BayArea cladogenesis model is trivial

and is described above. However, the DIVA cladogenesis model (Ronquist 1996;

Ronquist 1997; Ronquist & Sanmartin 2011) has only been implemented in a

parsimony framework.

Implementation of DIVA-like cladogenesis in BioGeoBEARS

Here, I develop a probabilistic interpretation of the DIVA model, essentially by

taking the DEC model and modifying it to allow only the cladogenesis events

allowed by DIVA. Strictly speaking this is a "DIVA-like" model; exact duplication of

all of the features of parsimony inference in terms of probabilistic models is very

involved (Huelsenbeck, Alfaro & Suchard 2011) and is not attempted here. In addition, as DEC and other probabilistic models of geographic range evolution use the important information found in phylogeny branch lengths (Donoghue & Moore 2003), and DIVA, as a parsimony method, does not, a likelihood implementation of DIVA may return different inferences than a parsimony implementation. (A close approximation could be made by taking all branch lengths to the exponent zero, rendering all branches of length 1 and discarding time information; this is an option in BioGeoBEARS; Matzke 2013b). However, here we are interested just in the cladogenesis model of DIVA; the anagenetic portion of the DIVA model will be identical with that of DEC, using two free parameters, *d* and *e*, to control the probability of dispersal (range expansion) and extinction (range contraction).

One of the goals in the programming of BioGeoBEARS (Matzke 2013b; Matzke 2013a) was to implement model assumptions as functions of parameters, rather than simply hard-coding assumptions into the computer code. Such hard-coded assumptions, although they may allow highly efficient calculations, can be difficult to change later, and can make it difficult for researchers to recognize, assess, or modify the models they are using, as they are usually in the poorly-document guts of the source code. This difficulty contributes to a research culture in which a single fixed model, being the only model available, gets naively applied to all situations and study systems; this practice is dangerous because of what might be being missed, as Matzke (2013d) demonstrated.

In BioGeoBEARS, the cladogenesis model is controlled by four parameters, y, s, v, and j, which control the relative per-event weight of each of the four cladogenesis processes, respectively, sympatry (range-copying), sympatry (subset speciation), vicariance, and jump dispersal/founder-event speciation. The details are described in Matzke 2013d. BioGeoBEARS users may fix any of these parameters, or to set them to be free parameters that are subject to inference. When $y=s=v > 0$, and $j=0$, the DEC model of LAGRANGE is reproduced, and each of the non-jump events is given equal probability, conditional on a particular ancestral range. To produce the BayArea cladogenesis model, j, s and v are set to 0, and y is set to 1, meaning that the only cladogenesis process allowed is sympatric speciation through range-copying.

To produce the DIVA cladogenesis model, then, one step is to start with the DEC model ($y=s=v=1$; $j=0$), and then set $s=0$, as DIVA does not allow subset speciation (Ronquist & Sanmartin 2011). However, this is not the only required modification, as DEC requires that one daughter species always have range size 1 during vicariance, and DIVA allows any daughter range sizes. Here, we interpret this to mean that, in our DIVA-like likelihood model, every possible vicariance result for the two daughter species receives equal probability, conditional on some ancestor. This maximizes similarity to the LAGRANGE-DEC assumption that each possible range-inheritance scenario has equal probability (Ree & Smith 2008).

Although allowing equal probability for every vicariance scenario could simply be hard-coded as an option, future research is enabled if the equal-probabilities model

is made to be the result of inputting some parameter value into some function. Many possible parameterizations could be imagined, including having a separate parameter describing the relative weight of each possible combination of (left daughter range, right daughter) given a particular ancestral range. However, given that the amount of data available in the form of tip ranges on a phylogeny will always be quite limited, efficient parameterizations should be sought.

In BioGeoBEARS, a single parameter, m_v, is used to control the relative probability, r, of the range size of the *smaller* daughter. (Once the range size of the smaller daughter has been established in a particular range inheritance scenario, the range size of the larger daughter species is just the remainder of the ancestral range.) The number of areas in the ancestor, for a particular range-inheritance scenario (i.e., a particular combination of ancestor range, left descendant range, and right descendant), is denoted by N. When vicariance occurs in an ancestor of range size N, the maximum range size of the smaller daughter species is L. L must equal floor($N/2$). I.e., for ancestral range sizes of 2, 3, 4, 5, 6, and 7, L for the smaller daughter is 1, 1, 2, 2, 3, 3, respectively. The actual range size of a particular daughter, l, thus may take a value from $1,...,L$.

The problem is thus how to use the single parameter, m_v, to determine $r_1, ..., r_L$, for any L, as L will vary throughout the rows of the cladogenesis transition matrix. It would be desirable if, when m_v is near 0, all of the weight is placed on cladogenetic range inheritance scenarios where the daughter has range size $l=1$. This would

132

replicate the assumption in LAGRANGE DEC. On the other handed, when m_v

approaches 1, all of the probability mass should be placed on range inheritance

scenarios where $l=L$. Finally, when $l=0.5$, the probability mass should be evenly

spread across all daughter range sizes, such that r_l is equal for all l.

A natural function that can be used to satisfy these requirements is the maximum

entropy discrete probability distribution of a series of ordered integers (Harte 2011,

equations 6.3 and 6.4), given the mean value, μ. This distribution is often used to

characterize the probability of the difference faces of a die. For example, for a 4-

sided die, with faces numbered 1-4, and a known μ of 2.5, the maximum entropy

distribution of the probability of each of the 4 die faces is flat, with each integer

having a probability of 0.25, indicating a fair die. If μ is lower or higher than 2.5, the

die will be biased towards lower or higher values, respectively. In BioGeoBEARS, m_v

is converted to μ by the following formula: $μ = m_v(L+1)$. The resulting probability

distribution on $1,...,L$ is used to weight the relative probability of different vicariance

cladogenesis events given a particular ancestral range size, based on the range size

of the daughter with a smaller area.

Implementation of BAYAREA-like cladogenesis in BioGeoBEARS

For the other three classes of cladogenesis, BioGeoBEARS implements a similar

parameterization of the relative probability of different range sizes for the smaller

daughter. This is done so that BioGeoBEARS users may alter the DEC assumption

that one daughter species always has a range size of 1 for a particular class of cladogenesis events. The only difference in the meaning of the *m* parameter comes with range-copying sympatric events. Here, there is only one possible cladogenesis event for a given ancestral range, and the question is whether or not sympatric range-copying will be allowed for an ancestral range of that size. The parameter m_y controls the per-event weight of a range-copying event at a particular range size (rather than the weight amongst events of that class). When m_y is close to zero, range-copying is allowed only for ancestral ranges of size 1. When m_y is close to 1, range-copying is allowed for ancestral ranges of any size without penalty. Notably, this means that when m_y=1, and $s=v=j$=0, the BayArea cladogenesis model, which allows only range-copying at cladogenesis events, is reproduced. Only BayArea's implied cladogenesis model is being used in this paper; other features of the BayArea inference are either not implemented in BioGeoBEARS (MCMC sampling of histories) or are not used here (dispersal probability as a function of distance).

Graphical representation of the cladogenesis weight matrix is difficult due to its size (Matzke 2013d; Matzke 2013a), but Supplementary Table 1 gives examples of the full matrix for a 4-area analysis for DEC, DIVA, and BAYAREA cladogenesis models, and with the "+J" versions of these models, with counts of the number of allowed events from each type of cladogenesis event.

Biogeography Datasets

The collection of island clades and constraint models is the same as in Matzke (2013d). To this dataset was added a sample of continental or marine clades from the literature for which a time-scaled phylogeny and geographic range data was available. The collection procedures were the same as in Matzke (2013d). As in that paper, when a constrained or time-stratified LAGRANGE analysis was conducted in the source study, it was re-implemented in BioGeoBEARS, along with an unconstrained version. Terrestrial clades included the Old World snake superfamily Elapoidea (Kelly *et al.* 2009) and the plant genus *Lonicera* (honeysuckles), distributed across the Northern Hemisphere (Smith & Donoghue 2010). In addition, four continental clades were added with centers of distribution in South and Central America. These were specifically the euglossine orchid bees (Ramirez *et al.* 2010), the wasp genus *Notiospathius* (Ceccarelli & Zaldívar-Riverón 2013), Matos-Maraví *et al.* 2013, the "*Taygetis* clade" (Matos-Maraví *et al.* 2013) of butterflies, a taxon consisting of *Taygetis* and related genera (Matos-Maraví *et al.* 2013), and the bat genus *Sturnira* (yellow-shouldered bats, Velazco & Patterson 2013). Each of these four Neotropical clades is widely distributed across South and Central America, sometimes including Caribbean islands. Possibly important biogeographical events in all the clades include the orogenesis of the Andes and the Great American Biotic Interchange with the formation of the Panamanian isthmus (Hoorn *et al.* 2010). The four source studies are comparable in that they use similar discrete regions and sometimes dispersal constraints, and in that all the clades are fliers. The analyses also all used LAGRANGE (Ree & Smith 2008), thus assuming the DEC model.

Two marine clades were included. The first was Cardiidae (cockles and giant clams; Herrera 2013), a bivalve family with global distribution and a center of diversity in the tropical Indo-Pacific. Herrera's time-stratified dispersal matrix was also replicated and compared to an unconstrained model. The other marine clade was the wrasse genus *Thalassoma* (Bernardi *et al.* 2004), distributed across tropical coral reefs worldwide.

To achieve additional sampling of clades with continental distributions, seven taxa from sand dune fields of the desert southwest of North America were added to the analysis. The dated phylogenies were produced using BEAST (Drummond *et al.* 2012) by Ph.D. student Matt Van Dam for his dissertation (Van Dam 2013, used with permission), based on a combination of GenBank sequences and his own sequencing. These datasets were developed as part of a collaborative project (Van Dam & Matzke 2013, in preparation) to use a multiclade analysis to test the hypothesis that clades from the desert southwest show similar biogeographical structure (Riddle 1996) due to (a) common constraints on connectivity between dune regions (e.g., during Pleistocene flooding events) and (b) common constraints due to path distance between dune regions. In the present paper, these clades are only being used to compare the six different cladogenesis models described above, all in geographically unconstrained analyses.

The island clades and their various constraints models are summarized in (Matzke 2013d); the additional oceanic and continental clades employed in the present analysis are described in Table 1.

Historical Biogeographical Analysis

Six models for the evolution of geographic range, dubbed DEC, DEC+J, DIVA, DIVA+J, BAYAREA, and BAYAREA+J, were run on all datasets in R 2.15 (R Core Team 2013) using the R package BioGeoBEARS (Matzke 2013b) and its accessory packages rexpokit (Matzke 2013c) and cladoRcpp (Matzke 2013c). BioGeoBEARS relies on two other phylogenetics packages for basic phylogeny manipulation (Paradis 2012; R Hackathon et al. 2012). The package optimx (Nash & Varadhan 2011; Nash & Varadhan 2012) was used to conduct the maximum likelihood search, using the quasi-Newton method with box constraints (Byrd *et al.* 1995). The maxent function of the package FD (Laliberté & Legendre 2010; Laliberté & Shipley 2011) was used in the above-described calculation of descendant daughter range size weights. The package parallel, now part of the R base distribution, was used to speed calculations via parallel processing.

Model comparison

Each model including founder-event speciation contains the non-founder-event speciation model nested inside of it (by setting j=0). For these pairs of nested

models, the likelihood-ratio test (Burnham & Anderson 2002) was used to test the null hypothesis, namely that the two models are equally good explanations of the geographic range data. However, the other model comparisons (DEC vs. DIVA, DIVA+J vs. BAYAREA+J, etc.) are not nested. For these, AIC and AICc (Burnham & Anderson 2002) were used to quantify how well each explained the data. These were used to calculate model weight and relative probabilities of each model. Depictions of relative model probabilities were graphed in R using bar plots, following similar depictions in previous publications (Harmon *et al.* 2010; Gong *et al.* 2012). In order to assess the relative importance of founder-event speciation in island versus non-island systems, histograms of the ML estimates j were plotted, and the null hypothesis of no difference in the mean of j between island and non-island analyses was tested using Welch's two-sample t-test. As the distributions of j estimates across clades and analyses were non-normal, the Komogorov-Smirnov test was used to test the null hypothesis of identical distributions. In order to assess whether or not the degree of support for the "+J" models tended to be larger in analyses of island clades, linear models were fit that predicted **Δ**AIC (the AIC advantage of the "+J" model) as a function of number of OTUs (since analyses with more data will tend to have more resolving power) and of the clade category (island versus non-island). The parameters of these regression models were tested for significance using the standard analysis of variance tools in R.

Results

The results of 378 BioGeoBEARS runs are shown in Supplementary Table 2. These constitute the result of running the 6 cladogenesis models on each of 63 clade-constraints combinations. The table reports the log-likelihood of the data on each model, the parameter estimates, and the *p*-values of the likelihood-ratio test. Overall, the major result is that the addition of founder-event speciation is a statistically significant improvement for almost all clades and models. Under the likelihood-ratio test, the "+J" models were significantly better, at the *p*<0.05 level, for 131/135 analyses of island clades, and for 58/69 analyses of non-island clades. The exceptions to the common finding of significant improvement are revealing. One of the Hawaiian *Drosophila* analyses and two of the Hawaiian *Scaptomyza* analyses failed to reject DEC, and DEC and DIVA, respectively. However, these clades are much older than the currently extant four Hawaiian high islands (Kambysellis *et al.* 1995; Lapoint, O'Grady & Whiteman 2013), and these failures occurred in time-stratified analyses in which only the four extant high islands were used; when a fifth area was added, namely "Z", which existed before the eruption of the volcanoes that created Kauai, DEC and DIVA were again rejected for these groups (this outcome is shown for the clade names for which "5" has been appended to the clade name; see Supplementary Table 2, and Matzke 2013d). Therefore, for island clades, the only exception to the rule that +J models outperform standard models occurred with the time-stratified analysis of *Plantago*, where DIVA was not rejected in favor of DIVA+J (*p*=0.41).

The exceptions for non-island clades are more numerous and more interesting. For the *Uma* clade (fringe-toed lizards of the desert southwest), DEC and DIVA could not be rejected in favor of their respective +J models, although BAYAREA was rejected in favor of BAYAREA+J (p=0.0015). However, the *Uma* phylogeny had only 5 OTUs, and with these few data, failure to distinguish models is to be expected. Another exception was the clam clade Cardiidae, for which DIVA could not be rejected in favor of DIVA+J for either the unconstrained analysis (p=0.49) or the time-stratified analysis (p=0.50). For the euglossine bees (Neotropical orchid bees), DEC and DIVA could not be rejected under a time-stratified analysis (p=0.37 and 0.44, respectively), although rejections were obtained under the unconstrained analysis (p=0.0014 and 0.035), and BAYAREA was decisively rejected in both cases (p=~0). The Neotropical *Taygetis* butterfly clade was similar in that neither DEC nor DIVA could be rejected (p=0.085, p=0.17) in unconstrained analysis, although BAYAREA was rejected (p=0.033); furthermore, in the *Taygetis* analysis with time-stratified constraints, none of the models lacking founder-event speciation could be rejected. Notably, the clades in which DEC and DIVA fair well against the +J versions of these models are those which have many OTUs that have widespread and sometimes overlapping ranges. Even most continental and marine clades sampled in this analysis have OTUs that are usually restricted to single areas, but the Cardiidae, *Taygetis* clade butterflies, and euglossine bees are exceptions (Ramirez *et al.* 2010; Herrera 2013; Matos-Maraví *et al.* 2013).

Model selection with only three models

140

Use of the likelihood-ratio test and *p*-values as the sole tool for model testing has been criticized due to the frequentist assumptions that must be made for such tests (Link & Barker 2009); in addition, the likelihood-ratio test can only be used to compare nested models. Here, the DEC, DIVA, and BAYAREA models are nested within their "+J" counterparts, but they do not have nested relationships with each other. Model selection with AIC and AICc obviates these difficulties, allowing comparison of all models regardless of nesting relationships. Below, the results for AICc results are shown, as AICc contains a correction for small sample size; the AIC results are very similar (data not shown).

Figure 2 presents, for analyses of island taxa, the relative probabilities calculated from AICc when just the three models lacking founder-event speciation are compared. Of the 45 comparisons of the three models for island clade/constraint combinations, AICc favors DEC for 15 comparisons, DIVA for 26 comparisons, and BAYAREA for only 4 comparisons. Figure 3 presents the same comparison for non-island taxa. Of the 23 comparisons of the three models for island clade/constraint combinations, AICc favors DEC for 7 comparisons, DIVA for 8 comparisons, and BAYAREA for 8 comparisons.

Model selection with all six models

It is tempting to highlight the decrease in support for the DIVA model among non-island clade analyses, and the corresponding increase in support for the BAYAREA model, particularly because when BAYAREA is favored, it usually favored overwhelmingly (7 of 8 cases; Figure 3). However, it is well-known that model selection procedures do not tell the researcher what the true model is; they only allow selection of the best model among the models under consideration (Burnham & Anderson 2002).

This fact is emphasized by Figures 4 and 5, where the "+J" models have been added, resulting in the relative probabilities of all 6 models for each clade/constraints combination for island taxa (Figure 4) and non-island taxa (Figure 5). For island clade analyses, some "+J" model is almost always overwhelmingly favored. One exception is the stratified analysis of Hawaiian *Scaptomyza* using only the 4 extant Hawaiian islands (discussed above) where models lacking founder-event speciation total approximately 15% of the total model probability. The other exceptions are 4 analyses from these clades: the Hawaiian *Orsonwelles* spiders, and the Hawaiian plant groups *Plantago* and the silversword alliance. In each of these analyses, DEC receives 10-20% of the total model support. This is likely partially explained by the fact that these are all relatively small clades, with 12, 16, and 29 species, respectively (Supplementary Table 2).

In contrast, analyses of the non-island clades (Figure 5) do not show univocal support for "+J" cladogenesis models. For two analyses (*Uma* lizards and the

stratified analysis of *Taygetis* clade butterflies), the model gathering the most support is a non-J model, namely DIVA and DEC, respectively. The *Uma* result can be chalked up to the small size of this clade (5 species), a conclusion supported by the fact that 5 of the 6 models have approximately equal support in the *Uma* analysis. However, the *Taygetis* clade has 61 species, and DEC receives about two-thirds of the model probability, with DIVA increasing AICc support for non-J models to over 80%. The unconstrained analysis of the *Taygetis* clade supports BAYAREA+J, but BAYAREA without J attracts about 20% support.

Nevertheless, most analyses of non-island clades strongly or overwhelmingly support models that include founder-event speciation (Figure 4). Notably, addition of founder-event speciation also influences which model classes are most dominant. For example, while BAYAREA was favored in only one-third of analyses in comparisons between the three models lacking founder-event speciation, in non-island clade analyses, BAYAREA+J wins the most support in the majority of analyses (14/23). Often this is overwhelming support (in 7 of the 15 analyses, BAYAREA+J beats other models by 10:1 or more). DEC+J wins in only 5/23 analyses, and DIVA+J in 2/23.

In comparisons of the six models in analyses of island clades, DEC+J is the most common winner (23/45 analyses), followed by DIVA+J (15/45) and BAYAREA+J (7/45). Unlike the situation in analyses of non-island clades, here it is uncommon for a particular model to gather overwhelming support. In 31/45 analyses, no

model gathers over two-thirds of the total model probability, typically because DEC+J, DIVA+J, and BAYAREA+J, while usually favored in that order, all have very similar amounts of support.

Results

An important caution in the above discussion is that, due to the diversity of clades and constraints models available in the island and non-island samples, no universal significance should be attributed to the counts of how many analyses support a particular model. The different analyses are not necessarily independent – for example when several analyses are done on the same clade -- and some study taxa have more constraints models available than others, and thus more analyses were performed for these clades. These difficulties can be ameliorated by comparing only one unconstrained analysis from each clade, but unfortunately in some clades, unconstrained analyses are not available, because the large number of areas used in the original source study necessitates that the maximum range size be constrained, in order to reduce the number of possible geographic range states to a computationally manageable level. Therefore, the counts are reported, but they should be regarded as only heuristics to indicate major tendencies in the analyses of island and non-island clades. Consultation of the figures and labels is recommended for a full assessment of the results, and the best statistical comparisons of the island and non-island analyses will remove non-independent analyses.

The average importance of founder-event speciation in island versus non-island clades

From the above, it is clear that models including founder-event speciation are generally strongly favored in both island and non-island clades, under a variety of constraints scenarios. However, it is still interesting to ask whether or not there is typically a difference in the *relative* importance of the founder-event speciation process in island versus non-island clades. This can be simply assessed by comparing the ML estimates of *j* for the two groups (Figure 6). As *j* represents the per-event weight of founder-event speciation at each cladogenesis event, it, unlike the *d* and *e* parameters, is not influenced by phylogeny branch length or root age.

The mean *j* across all analyses of island clades was 0.22, and across all analyses of non-island clades the mean *j* was 0.092. The non-island mean *j* was calculated after removing *Uma*, which likely gave unreliable results due to the extremely small size of the clade; however, the difference in mean j calculated with Uma is minimal (*j*=0.089). *Uma* is removed from further analyses below.

A statistical test of the null hypothesis that the mean *j* estimate is different between island and non-island analyses is complicated by the fact that the distribution of these estimates is strongly non-normal (Figure 6). Here, this results in a large standard deviation in the distribution, increasing the chance of non-significant results. As the mean value of *j* gives an easily-interpretable measure of the relative importance of founder-event speciation, the means were compared with statistical tests (Welch's two-sample t-test) despite the non-normality; non-parametric tests

that do not make the normality assumption were also performed and are given

below. The difference in means is significant at the $p<0.05$ cutoff ($p=7.1e-6$). Due to

the non-independence of some analyses, the test was repeated using just one

unconstrained analysis per clade, even though this meant excluding some clades;

the means remained similar and the null hypothesis was again rejected (mean j for

island clades: 0.21; for non-island clades: 0.062; $p=0.0040$).

As the j estimate is correlated between the DEC+J, DIVA+J, and BAYAREA+J

inferences for a particular clade/constraint combination (data not shown), the

significance tests were repeated for just DEC+J inferences between island and non-

island clades. This reduces the size of each sample by a factor of 3, and although the

means are similar (for all constraints analyses, $j=0.25$ for island analysis and 0.13

for non-island analyses; the corresponding mean j estimates are 0.23 and 0.071

when comparing only unconstrained analyses), the t-test no longer returns

significant results ($p=0.072$ and $p=0.13$, respectively). However, taking the natural

log of each j estimate under the DEC+J model produced approximately normal

distributions, and when Welch's two-sample t-test is applied (formally appropriate

in this case), the difference in mean log(j) is again significant for both all

clade/constraint combinations (mean log(j)= -2.1 for island analyses and -3.2 for

non-island analyses, $p=0.0094$), and for only unconstrained analyses (mean log(j) =

-2.3 for island analyses and -3.6 for non-island analyses, $p=0.017$). In summary, the

ML estimate of j is on average much lower in analyses of non-island clades, with the

mean value of j being only 25-50% of the mean estimate for island clades,

depending on the specifics of which models and constraints are used to test for differences between island and non-island analyses.

The Komogorov-Smirnov test for differences in distribution was also performed on the non-log-transformed distributions of j. The more conservative one-sided test (null hypothesis: the cumulative distribution function of island j estimates is not higher than the CDF of non-island j) was used. In all cases, the null hypothesis was decisively rejected. Comparing all analyses, p=5.2e-6. Comparing all unconstrained analyses, p= 1.9e-15. Comparing just unconstrained DEC+J analyses, p=3.4e-7.

Finally, if it is true that founder-event speciation is a more important process in island clades than in non-island clades, it should be true that the model selection advantage of the "+J" models, as measured by ΔAIC, is higher in island clades than non-island clades. Testing for this difference is complicated by the fact that the ΔAIC for the "+J" model will also be influenced the details of the constraints used in each analysis of each clade, and by the number of OTUs in the phylogeny. The latter is problematic as the non-island clades sampled in this study typically have more OTUs than the island clades. A linear model was estimated that predicted ΔAIC across all DEC+J analyses as a function of number of OTUs, and of the geographical category (island versus non-island). Analysis of variance under this model supported the influence of number of OTUs on ΔAIC (p=0.0019), as well as a weaker influence of the categorical predictor (p=0.044), the latter indicating that the +J model has a positive, nonzero advantage in ΔAIC model support in island systems. A

test for differences in the relationship between number of OTUs and **Δ**AIC showed

no difference in slope (*p*=0.99). These tests yielded similar results when repeated

for linear models fit to other combinations of models and constraints (data not

shown). A plot of **Δ**AIC versus number of OTUs for the DEC+J analyses (Figure 7)

confirms the general impression that island clades on average give higher model

support to the +J model, for a given number of OTUs, although there is much

variability.

Discussion

The first important result of this study is that founder-event speciation is shown to

be a crucial explanatory process, not just in island clades, but in most non-island

clades as well. The general importance of founder-event speciation will not come as

a surprise to biogeographers with dispersalist tendencies, but it should be

disturbing to those who favor a maximum-vicariance model for inference in

historical biogeography (reviewed by Matzke 2013d). Furthermore, while the

importance of founder-event speciation for clades inhabiting oceanic islands has

long been well-accepted amongst island biogeographers (de Queiroz 2005; Cowie &

Holland 2006; Gillespie *et al.* 2012), no similar consensus has existed about the

processes important for explaining the distributions of continental clades. This

study indicates that founder-event cladogenesis must be included as a likely process

in studies of the biogeography of continental clades. It should be noted that this is

not the same thing as claiming that founder-event speciation is the *only* process, or

the most important process, in explaining the distributions of continental clades. The DEC+J and DIVA+J models both include vicariance, and in some cases these models received the most support.

While support for founder-event cladogenesis is an important result, the fact that the methods employed here allowed *measurement* of the *relative* importance of founder-event speciation may be even more important. The study confirmed that, while founder-event speciation is an important process in most analyses of both island and non-island clades, the *j* parameter, measuring the per-event weight of founder-event speciation, is 2-4 times higher in island clades than in continental clades.

Another encouraging result is the fact that the "+J" models were not always the best models. For the time-stratified analysis of *Taygetis* clade butterflies, DEC was identified as the best model, confirming that in the original study (Matos-Maraví *et al.* 2013), which used LAGRANGE in a time-stratified analysis, the best model was in fact employed. This supports simulation studies (Matzke 2013d) indicating that DEC and DEC+J are easily distinguishable, and that the usual preference of the data for DEC+J over DEC is not universal, nor an artifact. A clue as to the reason that DEC is preferred in the *Taygetis* clade can be found by examining the geographic ranges at the tips of the tree: the *Taygetis* clade is unique amongst the studies sampled in that a majority (37/61) of the OTUs had geographic ranges larger than a single area (Matos-Maraví *et al.* 2013). The other clades for which support for the "+J" models

was relatively weak, and where sample size was large, also had many OTUs with widespread ranges, namely Cardiidae and euglossine bees (Ramirez *et al.* 2010; Herrera 2013). It appears that these results can be summarized with a simple rule: if the OTUs of a clade mostly or entirely inhabit single areas, then models including founder-event speciation will be heavily favored. Clades with many OTUs with widespread ranges will tend to give weaker support to "+J" models, or support models completely lacking founder-event speciation. This is in accord with common observation of the importance of range-expansion (or geodispersal) in explaining the distributions of widespread fossil taxa (D. Jablonski, personal communication).

Comparison of model probabilities in island and non-island analyses (Figures 4 and 5) revealed another interesting pattern. Many of the island clades gave almost equal support to DEC+J, DIVA+J, and BAYAREA+J models. This pattern was absent in non-island clades. My interpretation of this pattern is that these clades are ones where most or all OTUs are single-island endemics, and the dominant cladogenesis processes are just two: founder-event speciation, and within-island sympatric speciation. When these two processes are all that is required to explain distributions, there is very little difference between the likelihood of the geographic range data under DEC+J, DIVA+J, and BAYAREA+J models; these models each allow processes in addition to founder-event speciation and within-island sympatric speciation, but if they are never needed to explain the data, these models are all approximately equivalent. This hypothesis could be tested in more complex model-testing exercises, wherein each cladogenesis process is turned on or off; if my

hypothesis is correct, models allowing only within-island and between-island speciation will receive the most support for many island taxa.

A final interesting result is the surprising strength of the BAYAREA+J model in non-island clades (Figure 5). The cladogenesis models of DEC and DIVA were inspired by intuitions about the biological processes likely to operate during speciation. In contrast, the cladogenesis model in the BayArea program, which consists merely of copying the ancestral range unchanged at speciation events (Figure 1), was not inspired by any intuition about biological processes. Rather, it was just assumed as a simple starting point for a new computational technique that focused on the anagenetic processes of geographic range evolution (Landis *et al.* 2013). Two explanations of this result are possible. First, it may be the case that, contra the argument of Ree and Smith (2008), sympatric speciation of a widespread ancestor actually is a common occurrence, at least in some non-island clades. This seems particularly plausible in Neotropical clades with large ranges, overlapping sympatric ranges, and a great deal of specialization. Alternatively, it is possible that the fact that molecular phylogenies are missing many speciation events – those which led to species that are extinct or unsampled – somehow favors the BAYAREA+J model. (Sampling is an important consideration, e.g., for cardiids, many narrow-ranging species are missing, e.g. from the Central Pacific; D. Jablonski, personal communication.) For example, it might be the case that the true cladogenesis process resembles the DEC+J model, but that daughter species with small ranges have a high rate of extinction, such that both daughter lineages tend to survive to

151

the present only when they are the product of rare widespread sympatric events, or when a daughter with a small range rapidly expands to approximate sympatry with its sister species. The problem of missing speciation events might be particularly salient in relatively old, speciose clades with intracontinental distributions. This conceptual model explicitly links the evolution of geographic range with macroevolutionary processes of lineage diversification and extinction. Pioneering studies have produced probabilistic models that may be used to model geographic range evolution jointly with unobserved speciation and extinction events, however these models currently only exists for two-area (GeoSEE; Goldberg, Lancaster & Ree 2011) and three-area (ClaSSE; Goldberg & Igić 2012) systems. Addition of such models to BioGeoBEARS and other inference packages is clearly a priority, as it would enable the testing of the hypotheses described above.

A caveat should be mentioned. All inferences using the approach described here study depend on the discretization of biogeographic range into a relatively small number of regions. These regions are therefore usually quite large, such as entire islands, ecoregions, or continents. Thus, when "cladogenesis" processes are discussed and modeled within this framework, all statements about process are relative to the discretization scale of the biogeography. There is no guarantee, for example, that "sympatric" speciation or "founder-event" speciation within the framework of a coarse discretization of biogeography corresponds exactly to the detailed population-genetic mechanisms that might be described in a textbook. Speciation "sympatric" within one large area might be allopatric at a finer scale.

Conclusion

This study demonstrates that founder-event speciation is important not only in island clades, but in continental and oceanic clades as well. This conclusion across a broad sample of island and non-island clades, under a variety of constraints scenarios, and for each of the three main cladogenesis models currently in use in the historical biogeography literature (DEC, DIVA, and BAYAREA). The latter finding further supports the conclusions of Matzke (2013d). It also applies to both intra-continental and inter-continental clades, as well as globally-distributed oceanic clades. Future multiclade analyses should work to expand the representation of the latter two categories (represented by only four clades in the present analysis), but it appears that a confident prediction can be made that in most clades, founder-event speciation will play an important role.

While the qualitative result that founder-event speciation is more important in island clades than in non-island clades will likely match the intuitions of many dispersalist biogeographers, the ability to convert this intuition into a quantitative result of formal statistical inference is extremely useful. The fact that the relative importance of founder-event speciation in island versus non-island systems can be measured, and that founder-event speciation is 2-4 times more important in island systems, indicates that inferences of biogeographical history on phylogenies need

no longer be governed by fixed starting assumptions. With the probabilistic

modeling framework in place to enable quantitative inference of the importance of

different processes, advocates of different approaches and processes in

biogeography now have a common method by which hypotheses may be applied to

data and their relative importance assessed in statistically repeatable fashion.

It is easy to imagine the expansion of the approach presented here. For example,

different clades almost certainly operate by different rules; terrestrial amphibians,

for example, have much more limited long-distance dispersal capabilities than flying

birds. While this has been known since Darwin, the limited methods available to

historical biogeographers have forced a "one size fits all" approach when inferring

biogeographical histories on phylogenies. Clearly the time has come to assemble

large samples of phylogenies in clades with contrasting traits, and thereby measure

the relative importance of founder-event speciation and other processes in

explaining their biogeography.

Another avenue worthy of exploration is improvement in vicariance models. In this

study, following the assumption made in LAGRANGE's DEC model (Ree *et al.* 2005;

Ree & Smith 2008), all traditional cladogenesis events (sympatry and vicariance)

are given equal probability of occurrence, regardless of the configuration of areas.

This is surely a crude model, although arguably not any worse than the similar

assumptions that are made, in unconstrained analyses, for anagenetic dispersal and

local extinction processes, or for founder-event speciation. Biogeographers who are

convinced that vicariance dominates over founder-event speciation as an explanatory process, or at least those who wish to test such hypotheses, are encouraged to demonstrate this by putting forward vicariance models and datasets that statistically favor these models over the models currently available. Given the early stage of probabilistic model building in biogeography, there are surely better models out there, waiting to be found.

Acknowledgements

All Supplemental Files will be available on Dryad (dryad.org).

Conflicts of Interest

None.

Biosketch

Nicholas J. Matzke obtained his Ph.D. in the Department of Integrative Biology at the University of California, Berkeley, in Summer 2013. He is now a postdoctoral researcher at the National Institute of Mathematical and Biological Synthesis (NIMBioS, www.nimbios.org) at the University of Tennessee, Knoxville.

Table 1. Clades and data sources for non-island clades.

Common	Clade	Larger taxon	# taxa	tree age	region	island/continent	Dataset source	# areas	max # areas	Constraints	Constraint Model Abbr.
assimineid snails	*Assiminea*	Assimineidae	12	1.97	Southwestern North America	Intra-continental	Van Dam (2013) phylo & geog	11	4	unconstrained	M0
pupfish	*Cyprinodon*	Cyprinodontidae	23	7.1	Southwestern North America	Intra-continental	Van Dam (2013) phylo & geog	11	4	unconstrained	M0
spring snails	*Pyrgulopsis*	Hydrobiidae	100	22	Southwestern North America	Intra-continental	Van Dam (2013) phylo & geog	11	4	unconstrained	M0
giant flower-loving flies	*Rhaphiomidas*	Mydidae	26	67.7	Southwestern North America	Intra-continental	Van Dam (2013) phylo & geog	11	4	unconstrained	M0
sand dune weevils	*Trigonoscuta*	Curculionidae (true weevils/snout beetles)	25	14.6	Southwestern North America	Intra-continental	Van Dam (2013) phylo & geog	11	4	unconstrained	M0
thermal spring snails	*Tyronia*	Hydrobiidae	38	11.6	Southwestern North America	Intra-continental	Van Dam (2013) phylo & geog	11	4	unconstrained	M0
fringe-toed lizards	*Uma*	Phrynosomatidae	5	14.7	Southwestern North America	Intra-continental	Van Dam (2013) phylo & geog	11	4	unconstrained	M0
orchid bees	Euglossines	Hymenoptera: Apidae	137	41.3	South/Central America	Intra-continental	Ramirez et al. (2010) phylo & geog	7	7	unconstrained	M0
			137	41.3						time-stratified dispersal	M1
Neotropical wasps	*Notiospathius*	Braconidae: Doryctinae	142	40	South/Central America	Intra-continental	Ceccarelli & Zaldivar-Riverson (2013) phylo & geog	8	8	unconstrained	M0
			142	40					2	unconstrained, but 2 areas max	M1
			142	40					8	time-stratified dispersal	M3_stratified
			142	40					2	time-stratified dispersal	M3_stratified2
Neotropical butterflies	*Taygetis* clade	Nymphalidae: Satyrinae: Euptychiina	61	21.9	South/Central America	Intra-continental	Matos-Maravi et al. (2013) phylo & geog	8	8	unconstrained	M0
			61	21.9					8	time-stratified dispersal	L2
yellow-shouldered bats	*Sturnira*	Chiroptera, Phyllostomidae	22	13.4	South/Central America	Intra-continental	Ramirez et al. (2010) phylo & geog	8	8	unconstrained	M0
			22	13.4					4	4 areas max, adjacency constraint	M2
cockles	Cardiidae	Veneroida (bivalves)	119	134	global	ocean	Herrera (2013) phylo & geog*	11	3	stratified as in Herrera (2013)	M3
wrasses	*Thalassoma*	Labridae	28	10.5	global	oceanic	Bernardi et al. (2004) phylo & geog	8	2	2 areas max	M1
snake superfamily	Elapoidea	Caenophidia	45	5.58	global	inter-continental	Kelly et al. (2009) phylo & geog	7	7	unconstrained	M0
honeysuckles	*Lonicera*	Caprifoliaceae	59	41.5	Northern Hemisphere	inter-continental	Smith & Donoghue (2010)	7	7	unconstrained	M0

* Herrera's analysis was changed in that the maximum # of areas was set to 3, rather than 4.

157

Figure 1. The types of cladogenesis events allowed (gray) and disallowed (white) under the DEC, DIVA, and BAYAREA cladogenesis models. The "+J" version of each model adds founder-event speciation, as illustrated for DEC+J.

Figure 2. Island clades, relative probabilities of the three standard cladogenesis models: DEC (dark blue), DIVA (dark green), and BAYAREA (orange). The relative probabilities were calculated from AICc weights. The "+J" models are not included in this comparison. The clade/constraint combination is labeled at bottom, and the best model among the three is labeled at top.

Figure 3. Non-island clades, relative probabilities of the three standard cladogenesis models: DEC (dark blue), DIVA (dark green), and BAYAREA (orange). The relative probabilities were calculated from AICc weights. The "+J" models are not included in this comparison. The clade/constraint combination is labeled at bottom, and the best model among the three is labeled at top.

Figure 4. Island clades, relative probabilities of all six models: DEC (dark blue), DIVA (dark green), BAYAREA (orange), DEC+J (light blue), DIVA+J (light green), and BAYAREA+J (yellow). The relative probabilities were calculated from AICc weights. The clade/constraint combination is labeled at bottom, and the best model among the six is labeled at top.

Figure 5. Non-island clades, relative probabilities of all six models: DEC (dark blue), DIVA (dark green), BAYAREA (orange), DEC+J (light blue), DIVA+J (light green), and BAYAREA+J (yellow). The relative probabilities were calculated from AICc weights. The clade/constraint combination is labeled at bottom, and the best model among the six is labeled at top.

Figure 6. Histograms of the maximum-likelihood estimates of j for analyses of island clades (top) and non-island clades (bottom).

Figure 7. Scatterplot of the number of tips in the phylogeny of each clade, versus ΔAIC, the AIC advantage of the "+J" model, for all DEC/DEC+J analyses. Open triangles represent analyses of island clades; filled dots represent analyses of non-island clades.
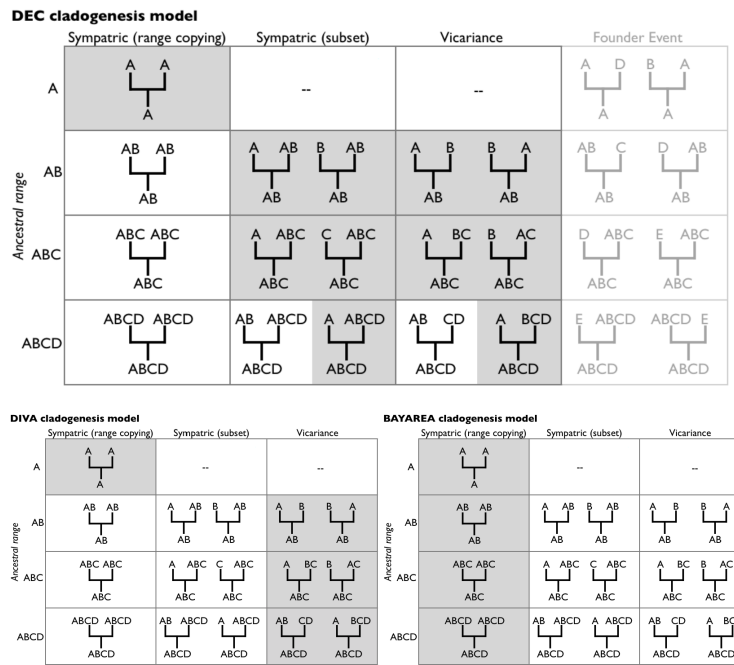
Figure 1. The types of cladogenesis events allowed (gray) and disallowed (white) under the DEC, DIVA, and BAYAREA cladogenesis models. The "+J" version of each model adds founder-event speciation, as illustrated for DEC+J.
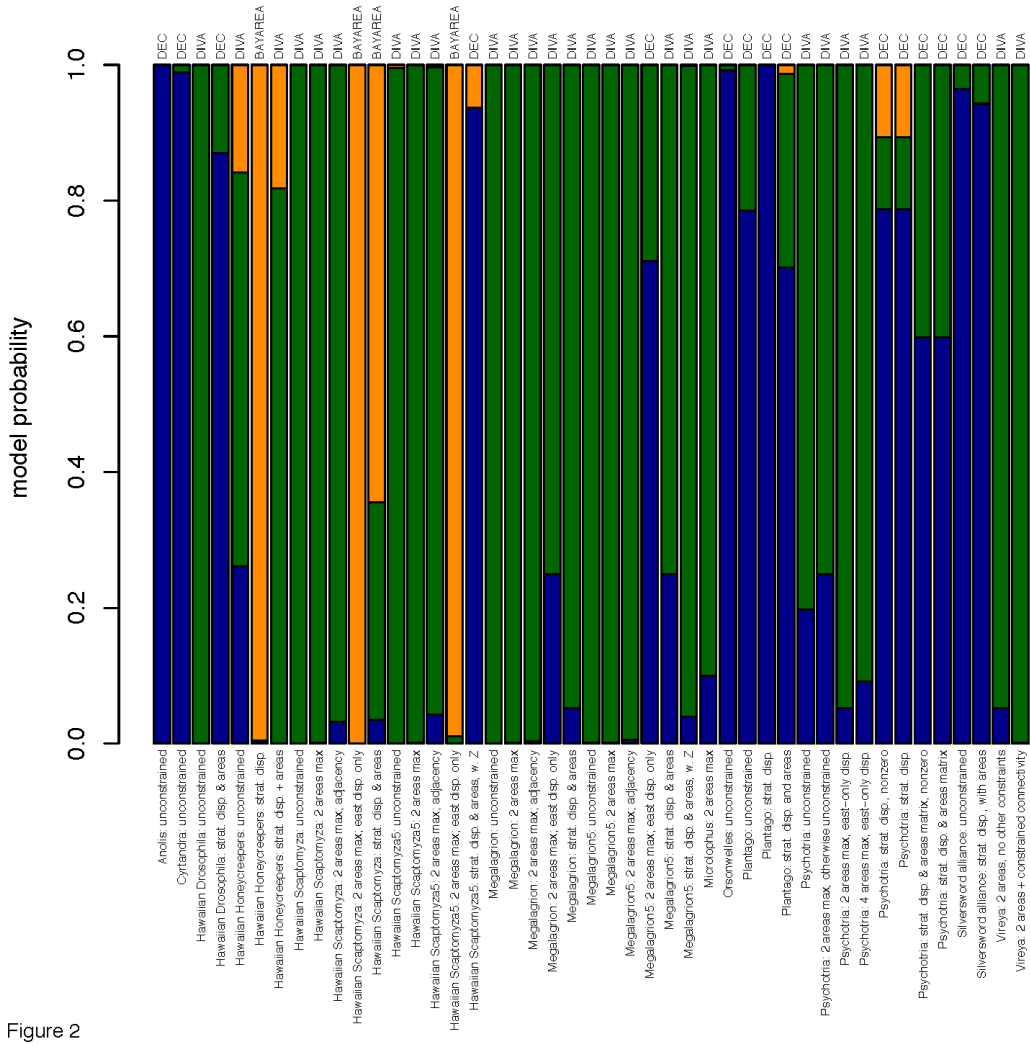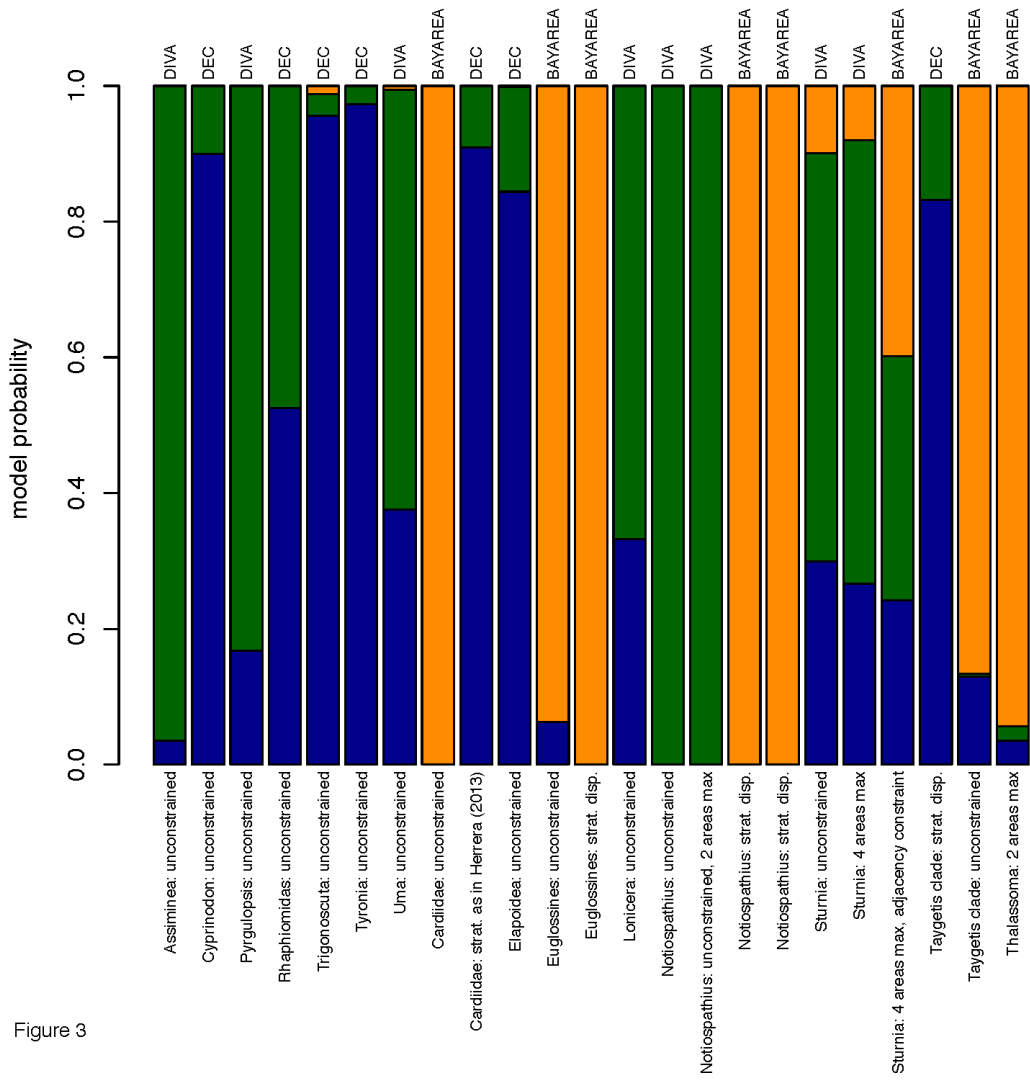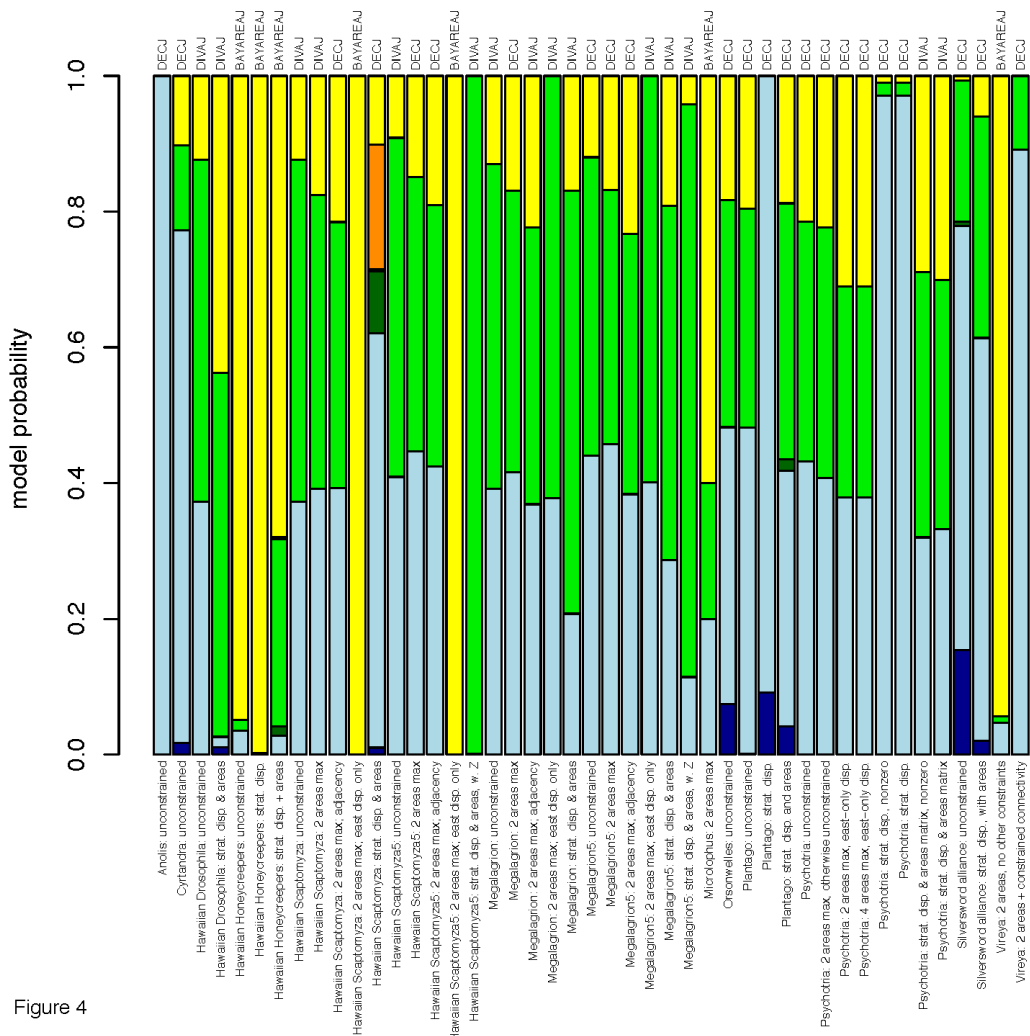
Figure 2

Figure 3

163

Figure 4

Figure 5

island clades

non−island clades

Figure 6

167

Figure 7

169
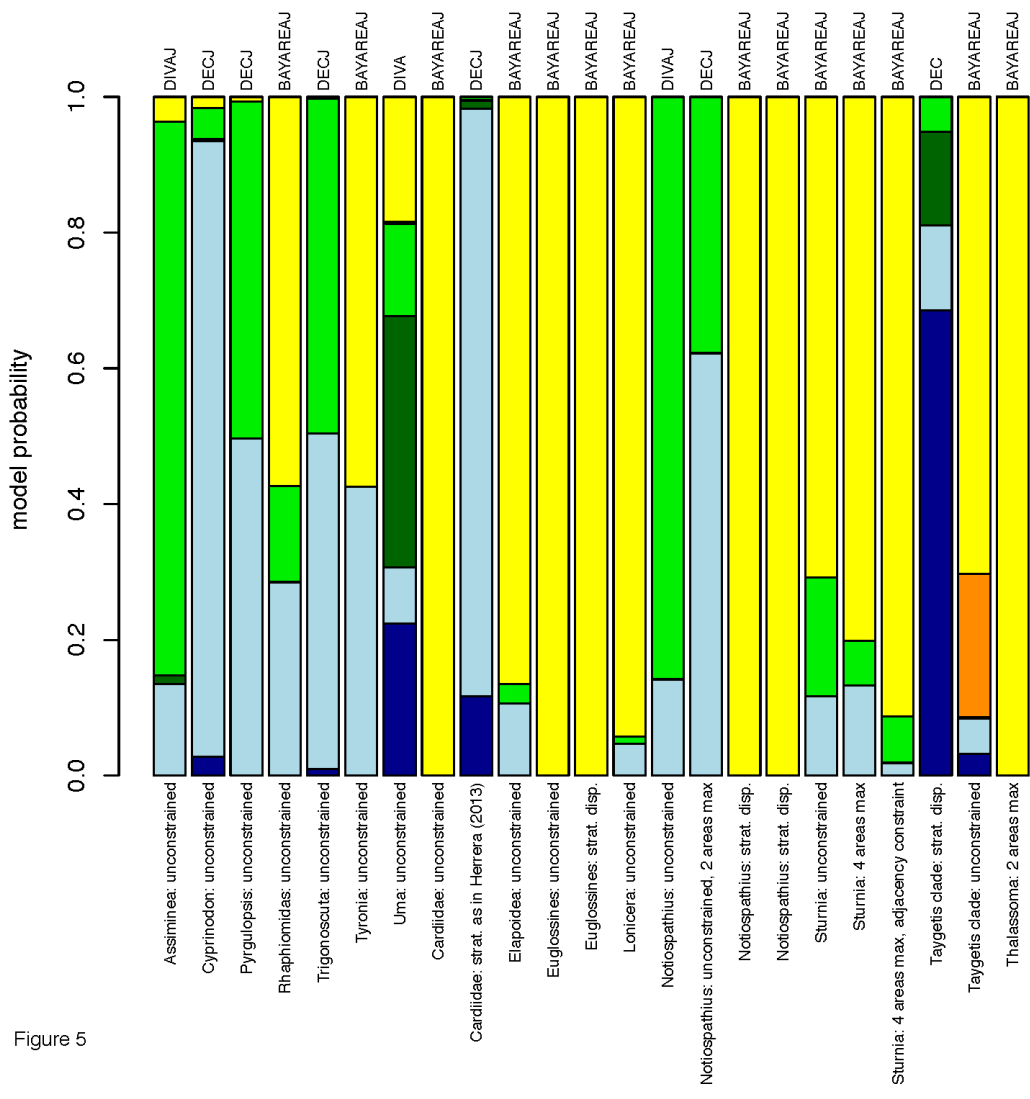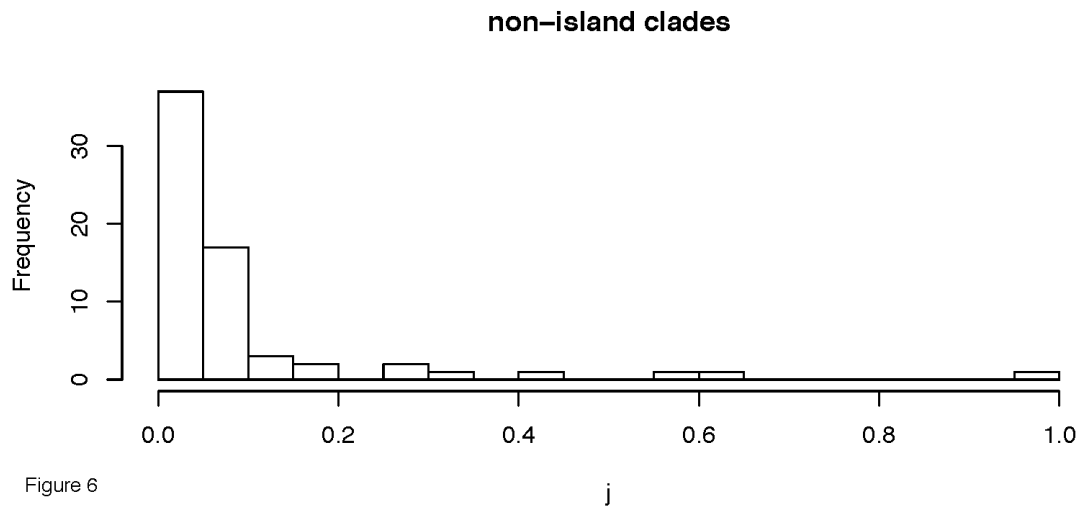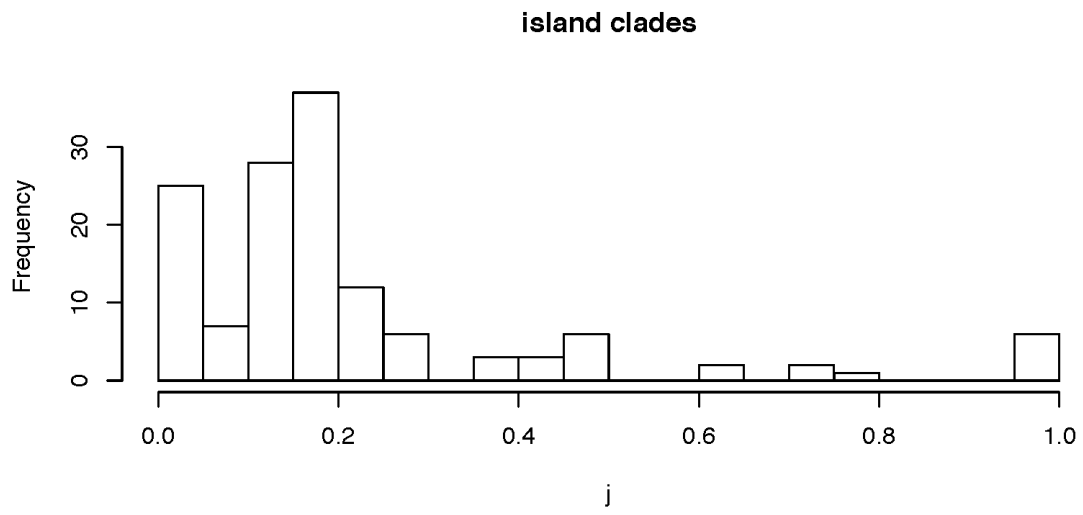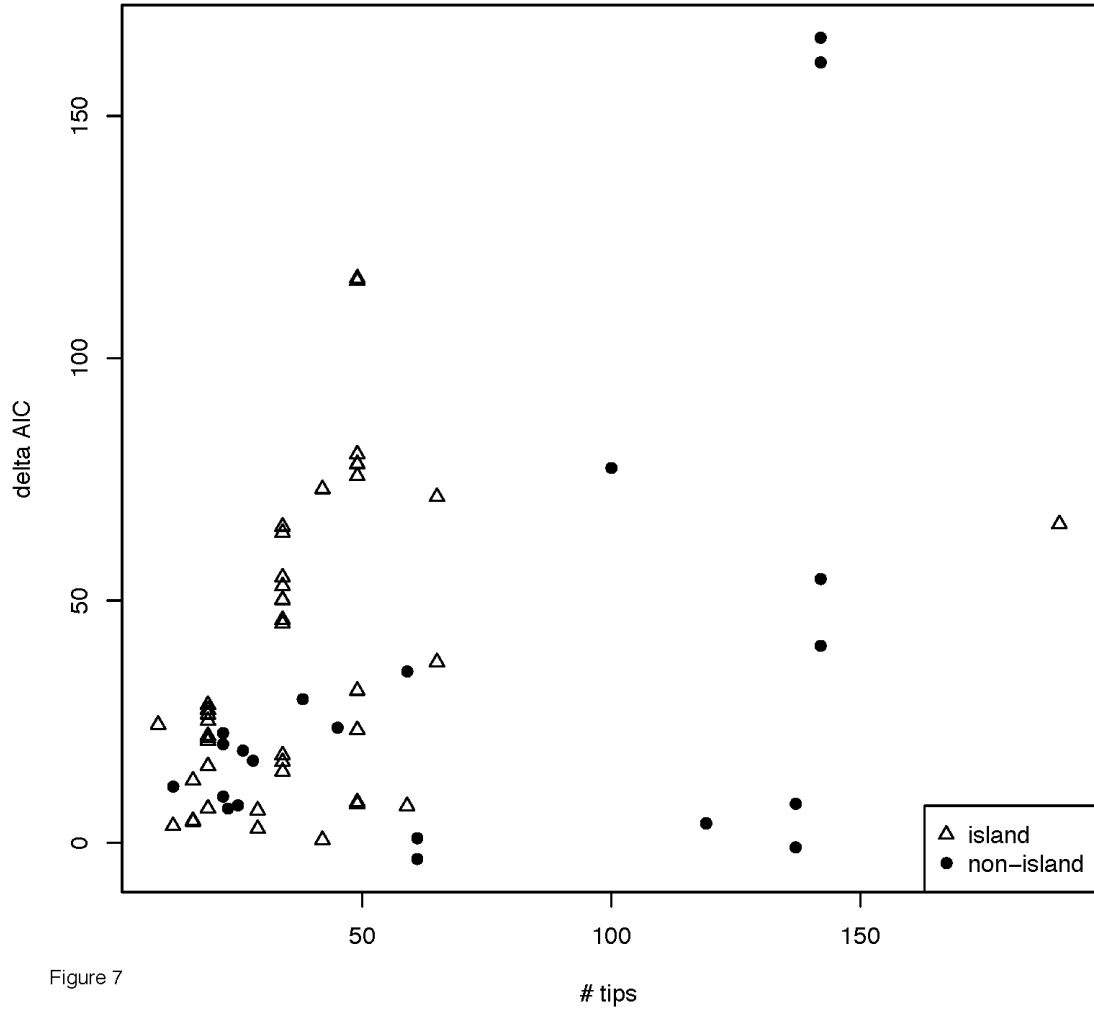
Chapter 4: Incorporation of fossils in likelihood analyses of historical biogeography using models for imperfect detection

Abstract

Likelihood analyses of geographic range evolution have become increasingly popular as analysis tools in historical biogeography. However, the models and software available have not had documented features allowing for fossils, or would allow the use of fossils but only in special circumstances. In addition, no methods have taken into account the fact that detection of presence and absence in regions will often be imperfect for fossil taxa. Here, both of these problems are addressed in the software BioGeoBEARS, and an example application to North American fossil Equinae and Canidae is presented. The NEOMAP database is used to provide occurrence data for both clades of interest, and for taphonomic control groups that are used to measure relative sampling effort in each region and time bin. The two clades are found to prefer different models for cladogenesis, a result found both with and without usage of the model for imperfect detection. Ironically, in test data chosen because of their high-quality fossil record, the record was so good that the model for imperfect detection had little impact.

Introduction

Fossil data are fundamental to increasing our understanding of evolution, as they constitute our most direct information about what existed, where, and when in evolutionary history (Smith 1994; Raff 2007; Sepkoski & Ruse 2009). Despite the various imperfections of the fossil record, themselves the subject of extensive study (e.g., Donovan & Paul 1998; Kidwell & Holland 2002; Benton 2009), "the inescapable conclusion is that the fossil record is an invaluable repository of information relevant to most of the major concerns of biological science" (Paul 1998). Furthermore, the neontological record has its own extreme imperfections, including the fact that most species that have ever lived are extinct (Nee & May 1997), and that the fossil record reveals morphologies that would never have been suspected or predicted solely from consideration of extant taxa (Edgecombe 2010).

The fossil record is similarly considered fundamental to our understanding of historical biogeography. No shortage of publications may be found voicing support (Jablonski, Flessa & Valentine 1985; Rosen 1988; Lieberman 2002; Lieberman 2003; Lieberman 2004; Roy, Jablonski & Valentine 2004; Lieberman 2005; Stigall Rode & Lieberman 2005; Stigall & Lieberman 2006; Fortey 2009; Lomolino, Lomolino & Lomolino 2010; Peterson & Lieberman 2012. Given this, it is peculiar that published analyses using the most-advanced, model-based, probabilistic methods for the inference of geographic range evolution on a phylogeny (Ree & Sanmartín 2009; Ronquist & Sanmartin 2011) typically leave out fossils. For example, out of

171

hundreds of published historical biogeography analyses using the most popular

program, LAGRANGE (Ree *et al.* 2005; Ree & Smith 2008), only two incorporate

fossils (Nesbitt *et al.* 2009; Wood *et al.* 2013). While many studies exist that use

undated cladograms and parsimony-based methods (e.g., DIVA, Ronquist 1997; or

Brooks' Parsimony Analysis and its relatives, Lieberman 2004) to include fossils in

historical biogeography analyses on cladograms, these do not take time explicitly

into account in the inference procedure. This is a major oversight which can lead to

problems such as pseudo-congruence between clades, and inference of vicariance

events which are falsified by dating analyses (Upchurch & Hunn 2002; Donoghue &

Moore 2003).

The problem of putting fossils in dated phylogenies

Much of the problem has stemmed from fact that, until recently, it was quite difficult

to include fossils in phylogenies dated with explicit statistical methods. First, of

course, many extant taxa lack fossil representatives, and thus all inference must be

done on phylogenies of living taxa. Second, even for groups with a fossil record, it is

often difficult to place fossils in the phylogenetic tree, due to limited character data

or incomplete specimens (Jablonski, Flessa & Valentine 1985). In this situation, the

role that fossils play in an analysis is often limited to providing calibrations for

dating internal nodes on a phylogeny (even though the best practice is to only use

fossils as dating calibrations after their position in a phylogeny has been determined

in a combined analysis; Parham *et al.* 2012). Third, even when sufficient characters

and specimens have been described to enable the inclusion of fossils in a phylogeny – either in a morphology-only analysis, or in a "total evidence" analysis combining morphological data with molecular data from living taxa -- until very recently, these studies were only conducted within the framework of undated phylogenetic analyses. Here, parsimony (Swofford 2007; Goloboff, Farris & Nixon 2008) or Bayesian (Ronquist & Huelsenbeck 2003) methods are used to infer the undated topology of fossil relationships, and if a dated tree is required, some *ad hoc* method is applied to scale the tree to time. The most common time-scaling method is to use the stratigraphic distribution of fossil species in the tree to set minimum ages for the clades containing these species. This always results in a ghost lineage for whichever member of a sister pair has the more limited stratigraphic range, and also often results in the compression of many cladogenesis events into an implausibly short period of time, with nodes separated by branches of length zero or some arbitrarily small value. If the result of minimum-age dating is treated as an inference result, rather than a merely a graphical display not meant to be taken literally, dramatically different interpretations of evolutionary history may result (O'Leary *et al.* 2013). Alternatives to pure minimum-age dating exist, for example evenly spacing out the branches below minimum-age constraints ((Nesbitt *et al.* 2009), and these are presumably improvements, but they do not constitute formal, probabilistic inference of a dated tree under an explicit model.

Recently, however, such models have become available in the form of Bayesian "tip-dating", a total evidence analysis of molecules and morphology, where fossils are

included as dated terminal taxa (Pyron 2011; Ronquist *et al.* 2012). The dates of the

fossil tips inform the dates of the nodes through a relaxed clock model applied to

morphological character data; traditional node-based dating calibrations may be

included if desired, but they are not required to produce an estimated dated tree.

Traditional dating based on node calibrations had a number of significant

weaknesses. In previous dating approaches, e.g., in the program BEAST (Drummond

& Rambaut 2007), an undated phylogenetic analysis was conducted first in order to

determine the phylogenetic position of fossils; these fossils were then converted

into prior distributions on the dates of calibration nodes in a molecules-only dating

analysis. Production of the node date calibration distribution was usually subjective

(Parham *et al.* 2012), fossils could not be used if their phylogenetic position was

uncertain (ironically, a particularly likely situation when "transitional fossils" found

near the branch points of major living clades are discovered), and after the date

calibration was produced, all additional fossil information (e.g., morphological

branch lengths and character data) was "thrown away" in subsequent dating

analysis. Worse, for purposes of estimating biogeographical history, the result of a

node-dating analysis was a dated tree with no fossil taxa. Tip-dating solves all of

these problems at once: subjective node calibrations are not required, fossil

character data and branch lengths are retained throughout the analysis, the

uncertainty in fossil relationships is directly factored into the joint estimation of

topology and divergence dates, and the result of a tip-dating analysis is a dated tree

which includes fossils. Tip-dating has already successfully been used to separate the

timing of genome duplication and the evolution of anadromy in salmonid fishes

(Alexandrou *et al.* 2013), and to demonstrate that the divergence of Palpimanoid

spiders is ancient enough to coincide with the breakup of Gondwanaland (Wood *et*

al. 2013), a rare exception to the common finding across many clades that

divergence times are too young to be explained by the breakup of continents before

the Cenozoic (de Queiroz 2005). Tip-dating as it currently exists certainly does not

solve *all* problems involved with the inclusion of fossils in phylogenies; notably, it

neglects the information that stratigraphic range and sampling frequency through

time may provide, and it, like almost all phylogenetic inference methods, does not

explicitly allow for species to be direct ancestors, even though we almost certainly

have sampled some direct ancestors in the fossil record (Foote 1996), and direct

ancestors may appear as short branches on the tree. For steps forward on these

issues, a recent special issue of *Methods in Ecology and Evolution* is recommended

(Bapst 2013; Ezard, Thomas & Purvis 2013; Hunt 2013; Slater & Harmon 2013).

The problem of imperfect geographic range data

Even once a dated tree including fossils has been obtained, additional difficulties

remain for the inclusion of fossil data in the estimation of biogeographic history on a

phylogeny. The primary difficulty is that the available fossil specimens

corresponding to an operational taxonomic unit (OTU) on a phylogeny may not

constitute a complete record of the geographic range of the taxon at the time point

or stratigraphic time-bin occupied by the OTU. Presence of a specimen is

informative (assuming accurate identification, although this cannot always be

assumed), but, as Rosen (1988) noted, "absence of a taxon from a particular time and place is ambiguous, and cannot be used as constructive or falsifying evidence" (p. 441). Specimens of an OTU may be missing from a particular region because the taxon really was absent from that region, but they may also be missing because sedimentary rocks of the appropriate age, type, and paleoenvironment are missing from the region, or are not exposed, or have not been explored for political or economic reasons. In addition, even if relevant fossils exist and have been discovered, they may not be known to the researcher, perhaps because they have not been published in an accessible journal, or not been entered into whatever database the researcher is using. Alternatively, the fossils might be available, but under a taxonomic name different than the one the researcher is searching for in the literature or database.

Some of these problems might be improved by the hard work of taxonomists and paleontologists – visiting museum collections, revising taxonomies, adding new field collections, etc. But some problems (unavailable rocks, political blockages) are completely beyond the power of hard scientific work to solve. And in any event, human limitations, the global nature of the research community, and the fact that taxonomy, museum records, the literature, and databases are all continuously moving targets mean that perfection will always be impossible, and thus biogeographic range data for fossils will always have the risk of gaps and errors.

I suggest that we will improve analysis and inference if we expand parametric

historical biogeography (Ree & Sanmartín 2009) to include parameterized,

probabilistic models of the imperfect detection process that produces our range

data. Below, I propose a simple model, which is intended as a starting point for

future research, and compare inferences made with and without the detection

model.

Study system

In order to provide a proof-of-concept of the method, and to give an example of how

it may be used for hypothesis testing, two study clades were chosen for which the

basic requirements of an imperfect detection model were available, namely: (1)

well-resolved, time-scaled species-level phylogenies; (2) abundant fossil occurrence

data in several different regions and time-bins; (3) similarly abundant fossil

occurrence data on taphonomic control groups with approximately equal

detectability; and (4) a standardized database containing the data which could be

downloaded and then queried via R scripts. The taxa chosen were Canidae (dogs

and relatives) and Equinae (horses), both from the well-studied Cenozoic of North

America. Both of these clades predominantly evolved in North America, and were

speciose for much of their history (although both are much less so in at present).

A biogeographical analysis of fossil Equinae was previously conducted by Maguire

and Stigall (2008). They used Lieberman-modified Brooks Parsimony Analysis (BPA,

an area cladogram approach; Lieberman 2000; Lieberman 2003; Lieberman 2004)

on a time-scaled cladogram assembled from the literature in order to assess the

relative roles of vicariance, dispersal, and climate in explaining the biogeography of

Equinae. They divided North America into four discrete regions (Southwest, Great

Plains, Gulf Coast, and Southeast). They mapped presence and absence in each area

onto the phylogeny, and from this area cladogram coded a vicariance matrix and a

geodispersal matrix, with gains and losses of areas in particular clades being

recoded as characters in these matrices. These matrices were then subjected to

parsimony inference, producing trees representing the best-supported relationships

between the four areas. The vicariance tree and geodispersal tree were largely

congruent, indicating that both processes had similar spatial structure, which the

authors suggested was due to fluctuations in climate influencing both processes.

The purpose of this paper is not to make a detailed assessment of the methods and

conclusions of BPA, which is an entirely different analysis paradigm than likelihood

inference of geographic range evolution. However, the general question of what

cladogenesis mechanisms are most important for explaining the biogeographic

ranges of North American fossil taxa can be addressed by the likelihood method, and

it would be useful to know if the conclusions reached depend upon whether or not

the likelihood method is making use of a model for imperfect detection of fossils.

This question is addressed in this paper by assessing the fit of a number of different

cladogenesis models (Matzke 2013e) to the Canidae and Equinae data, and

assessing how model selection results different with and without the inclusion of a model for imperfect detection.

Methods

Fossils and imperfect detection: previous work

Regarding the inclusion of fossils in likelihood analyses of historical biogeography, the original team of authors of LAGRANGE (Ree *et al.* 2005) did consider the use of fossils as constraints on ancestral ranges. The source code of the Python and C++ versions of LAGRANGE (Smith & Ree 2010; Ree 2013) contains brief references to use of fossil constraints; these features however remain undocumented at present. However, a manuscript (Moore *et al.* 2006) describing the use of fossil constraints in an early, Java-based version of a likelihood-based historical biogeography program, AreA, has been available online since 2006. The manuscript is apparently unpublished but is nevertheless an important contribution and has been cited multiple times in the literature (Moore & Donoghue 2007; Havill *et al.* 2008; Nylander *et al.* 2008; Nürk 2011; Harris, Wen & Xiang 2013). The AreA program, as with the first version of LAGRANGE (Ree *et al.* 2005) used a simulation approach to measure the probabilities of changes in geographic range along branches. For each branch, a large number of simulations were run under a set of model parameters,

and the distribution of resulting ranges was used as a probability distribution. Using

these distributions, the likelihood of the data under the model could be calculated.

Within this framework, use of fossils is relatively straightforward: if a fossil is placed

as a tip on the tree, and the available geographic range for that fossil OTU is thought

to represent the complete range, then only simulations which produce the observed

range are kept. If the fossil is to be considered only as a positive constraint,

indicating presence in the regions where it is observed, but indicating nothing about

presence or absence in regions where it is not observed, then simulations which

produce the observed presences are retained in the probability distribution,

regardless of whether the simulations have the OTU present or absent in regions

where the fossil has not been observed.

AreA also implemented use of fossils when a fossil could only be placed within a

clade, rather than on a specific branch. In this case, the model effectively assumed

that the fossil had an equal probability of being on any branch within the clade at

the time point in question, and simulations which failed to produce at least one

lineage occupying the area occupied by the fossil were penalized by a scaling factor

according to the number of branches. This "clade-specific" model of a fossil reduces

to the "lineage-specific" model described above, in the case when a fossil is found on

only one branch (Moore *et al.* 2006).

The simulation approach to calculation of likelihoods is computationally

cumbersome, and was superseded by the much more efficient matrix

exponentiation, pruning algorithm approach (Felsenstein 2004) in later versions of

LAGRANGE (Ree & Smith 2008). Nevertheless, a method treating fossils as positive

constraints on ancestral ranges is relatively easy to implement within the

framework of matrix exponentiation calculations of the data likelihood, at least for

the lineage-specific method. The situation is very similar to that of ambiguous base

calls in DNA sequencing (Felsenstein 2004). In maximum likelihood and Bayesian

phylogenetics packages, likelihood calculation begins by declaring the data

likelihoods under each of the 4 possible bases (A, C, G, and T) at each tip in the tree.

If the base at a particular site for a particular OTU is known to be "A", then the input

likelihood for that site at that tip is (1,0,0,0). If the base is "R", indicating a purine (A

or G), then the input likelihood is (1,0,1,0). In other words, the probability of the

sequence data is 1 under the hypotheses that true state is A or G, and 0 otherwise. If

the base is a "?", meaning total ambiguity, then the likelihood at the tip is (1,1,1,1),

i.e., the probability of the data is 1, under any hypothesis about the true identity of

the base.

The same approach can be used when the discrete states are geographic ranges,

rather than DNA bases. Consider a biogeographic analysis with two areas, "A" and

"B". There are four possible geographic ranges, if the null range, Ø (indicating

absence in all regions) is included. The four possible geographic ranges are Ø, A, B,

and AB. If a fossil is observed in region A, and this is taken to represent the

complete, true range, then the likelihood of the fossil data under each hypothesized

true state is (0,1,0,0). If the fossil observation in region A is taken as merely a

positive constraint, then two possible ranges are equally good explanations of the fossil data: A and AB. The data likelihoods are thus (0,1,0,1).

The R package BioGeoBEARS implements both of the above options (Matzke 2013b; Matzke 2013a). However, treatment of fossil range data as either complete range data, or as merely positive constraints which say nothing about geographic ranges where the fossils are unobserved, are positions at the two extreme poles, representing maximum confidence, and maximum agnosticism. There is a large, interesting grey area in-between these poles, where fossil geographic range data is partial and incomplete, but where absence of a fossil OTU from a region can be evidence of absence, if there has been sufficient sampling.

Modeling imperfect detection in historical biogeography

Models for imperfect detection have recently become popular in ecology (MacKenzie *et al.* 2002; MacKenzie *et al.* 2003; Mackenzie 2005; Royle, Nichols & Kéry 2005; Bailey *et al.* 2007; MacKenzie *et al.* 2009) and conservation biology (Kéry & Schmidt 2008; Link & Barker 2009). Species occupancy at a study site is important data for studying diversity, population size, conservation importance, etc. Many species will only be detected in a proportion of visits to the site, because they are only seasonally visible, or they migrate between habitat patches, or they are rare on the landscape. Information about the frequency of sampling over repeated visits can be used to estimate the parameters of a detection model, the key parameter of

which is the probability of sampling the species of interest on a per-visit or per-observation basis (Link & Barker 2009).

However, in historical biogeography, the "sites" are actually large, discrete regions – usually islands, ecoregions, continents, or ocean basins. Obviously, biogeographers and paleontologists do not usually conduct repeated visits to a large region and count how many times they detect a particular taxon. Furthermore, unlike many ecological or conservation applications, which may be focused on single species or on a community of unrelated taxa, biogeographers are interested in a collection of taxa that are all members of a monophyletic clade and placeable on a phylogeny. The data source is therefore usually some disparate collection of fossil observation records gathered over decades by many different researchers. The records usually refer to specimens, each collected at a particular fossil locality with some age or age range. The specimens are deposited in museums, sometimes described in a publication, and through one or both of these avenues, may (or may not!) make their way into a database, either one assembled for a research project by specialists, or a more general database built from the literature or museum records. This database therefore represents some imperfect, partial sample of the fossil specimens that have been discovered. These in turn are some imperfect, partial sample of fossils that are discoverable at exposed localities, and these are some partial, imperfect sample of all the fossils that exist. And, of course, the fossils that exist are some partial, imperfect sample of the original, true distribution of biodiversity in time and space.

One could imagine building a complex model that attempts to take separately into account each of these individual sampling processes, with parameters controlling each, but such a highly parameterized model would almost certainly lack sufficient data to estimate all the parameters, and undoubtedly many of the parameters would interact with each other and thus lack identifiability. In any event, usually the only data available is the database.

The key to building a model for imperfect detection is to have some measurement of sampling effort. Here, in the paleontological context, I use "sampling effort" to refer not only to the amount of sampling done by paleontologists, but also to the "sampling effort of Nature", due to the geological processes that preserved fossils, produced rock exposures of the correct age and type, placed these exposures in countries and on property where paleontologists can access them, etc.

A reasonable attempt to measure sampling effort – at least relative sampling effort – can be made through use of the concept of *taphonomic control groups*. Taphonomic control groups were originally defined by Bottjer and Jablonski as "higher taxa with ecological, morphological, and mineralogical properties similar to those of the group under study" (Bottjer & Jablonski 1988; see also Bottjer, Droser & Jablonski 1988). They noted that use of a taphonomic control enables the researcher to assess the significance of negative data – i.e., the observation of "a taxon's absence from a particular time and environment" (p. 540). Bottjer and Jablonski were concerned

with macroevolutionary patterns of the persistence of invertebrate taxa in onshore versus offshore environments. However, the concept can be repurposed for likelihood-based inference on phylogenies with a model that, given parameter values and hypothesized geographic ranges, allows the probability of the observations of the OTU of interest to be calculated.

The model is set up as follows. We start with a crucial assumption, one which is definitely arguable, but which at least provides a reasonable starting point for building models for imperfect detection. Imagine that a fossil species of interest, S, is a member of a clade of interest, C. A dated phylogeny exists for C, and the researcher plans to infer the history of geographic range evolution down the phylogeny, but the researcher wishes to include the possibility of imperfect detection. The researcher's database contains counts of occurrences of S in each discrete region of interest, in each time bin of interest. These occurrences could be presence/absence records at each of many fossil localities, or they could be actual counts of individuals at each locality, e.g. minimum number of individuals (MNI; Klein & Cruz-Uribe 1984). The database also contains counts of observations of many other species, both from clade C, and from D, a paraphyletic assemblage of many other clades. The crucial assumption is that, when S is present in a region, individuals from species S are approximately equally detectable as non-S individuals from clades C and D. Note that this assumption of approximate equal detectability is made about *individuals*, not about species; in all probability, there are many fewer individuals of the single species S than there are individuals of all of the other species in C and D. Obviously,

the taphonomic control groups in *D* should be chosen so as to maximize the plausibility of this assumption – i.e., approximately similar body size, preservational environments, etc.

The other assumption that must be made is that the database-derived counts of *S* and non-*S* individuals in each region of interest approximately reflect the true relative abundance of *S* and non-*S* in each region and time-bin of interest. With these two assumptions in place, as the counts of non-*S* individuals increase in a particular region and time bin, then the probability of counting an individual from species *S* will increase as well, if *S* is truly present in a region. If *S* continues to fail to be detected as the count of non-*S* increases, then this is evidence of absence. All that is required to make the probability calculations explicit is a value for the parameter *f*, the relative abundance of *S*.

How plausible are these assumptions? At first glance they may seem overly ambitious, but several considerations make them arguably valid, at least as a first approximation. First, there is a rich literature comparing species abundances in fossil and living assemblages (e.g., Kidwell 2001; Kidwell 2002; Kidwell & Holland 2002; Vermeij & Herbert 2004; Moore, Norman & Upchurch 2007; Terry 2010; Turvey & Blackburn 2011); if a summary may be hazarded, it is that abundances in fossil assemblages often are correlated with living abundances. In one well-known case, it has been shown that the time-averaging in death assemblages of marine molluscs may actually better reflect the long-term average community abundances

than short term samples of the living community (Kidwell 2001; Kidwell 2002).

Another study, comparing recent and Holocene fossil abundances from the avifauna

of Sweden, found that fossil species abundances were a good measure of the true

abundances within a particular size class, although not between them (Turvey &

Blackburn 2011). It is true that individual fossil localities can have any number of

site-specific taphonomic biases, due to local depositional environment, regional

climate, local physical features, sorting and trapping mechanisms, etc. (Moore 2012)

Even organism behavior can play a role – a famous instance is the prevalence of

carnivores in tar-pit sites. However, when we are dealing with very large, discrete

regions, as we are in historical biogeographic inference on phylogenies, and when

the stratigraphic time bins are also large, covering millions of years, it is reasonable

to hope that many of the biases of individual localities will average out.

Likelihood calculations

If the above assumptions are accepted for purposes of a particular study – especially

if the *C* and *D* taxa have been selected to maximize the chance that the above

assumptions are decent approximations, and the discrete regions and time bins

chosen for the analysis are large, encompassing hundreds of taxa and localities –

then the likelihood of the data can be calculated as follows.

For a particular species *S* in a discrete region R_m during time bin *t*, a database query

provides n_S, the count of observations of *S* across all databased localities in $R_{m,t}$. It

also provides $n_{!S}$, the count of the taphonomic controls, i.e. the non-S from all of the

other species in C and D. This query is repeated for all species in C, for all regions

and time-bins within the domain of analysis.

We are interested in the likelihood of the observation data in R_t under the

hypotheses that the species is truly present (T) or truly absent (F) from R_t. The

likelihood under hypothesis T is:

$$P(\text{data}|\text{T}) = f^{n_S}(1-f)^{n_{!S}} \qquad (1)$$

where f is the fraction abundance of S individuals in $R_{m,t}$, when S is present. Under

hypothesis F the likelihood of the observations is:

$$P(\text{data}|\text{F}) = 0^{n_S}(1)^{n_{!S}} \qquad (2)$$

In other words, under the hypothesis F that a species S is absent from R_t, the

likelihood of the data is 0 if S has in fact been observed, and 1 otherwise. This is

implicitly assumes that there are no false positives, i.e. that every databased

occurrence of S was correctly identified and the identification was correctly

transferred to the database. It is not difficult to construct a more complex model

with parameters controlling probability of false positives and false negatives; such a

model is implemented in BioGeoBEARS (Matzke 2013b; Matzke 2013a). However, it will take further study to determine what sort of data is required to make such parameters identifiable and subject to inference.

With the ability to calculate the likelihood of the observation data for a particular species S in a particular region $R_{m,t}$ under hypotheses T and F, all that is required to calculate the likelihood of the data for a particular hypothesized *geographic range*, G_t, is to atomize the range into a series of presences and absences, where $R_{m,t}$= T if S is hypothesized to be present in $R_{m,t}$, and F if absent:

$$G_t = \{R_{m=1,t}, \ldots, R_{M,t}\} \tag{3}$$

where M is the number of discrete areas being used in the analysis, and m is the index identifying a particular area. The likelihood of the count data for a particular G_t is then:

$$P(\text{data}_{G_t}|G_t) = \prod_{m=1}^{M} P(\text{data}_{R_{m,t}}|R_{m=1,t}) \tag{4}$$

In a standard likelihood analysis of geographic range evolution on a phylogeny, likelihood 1 would be assigned for the known correct range at each tip on the tree, and likelihood 0 would be assigned for all other possible ranges. However, with the above model for imperfect detection, the likelihood of database count data is

calculated for each possible geographic range for each tip on the tree. These likelihoods are then input into the likelihood calculations as in the standard analysis. Equation 4 is implemented in the function calc_obs_like in BioGeoBEARS (Matzke 2013b).

In the above, we have assumed that every species S in the clade of interest C has the same fraction abundance f compared to non-S when it is present in a region. Obviously this is a large simplification, as species have different relative abundances, and furthermore as relative abundances may vary across time and space. However, if C and D are chosen such that there is a diverse pool of non-S species, and no S in C ever dominates the relative abundance distribution in the broad region/time bin $R_{m,t}$, then it may be a fair approximation to say that the abundance of each species S in C has the same, small, fractional abundance f. Again, more complex models can be envisioned, but they could require many more parameters (e.g., a different f for each species in each region and time-bin) and therefore should explored in future studies about the data required for identifiability and inference. The simple model presented here can also serve as a null model against which to test more complex models.

The only issue remaining is to determine an appropriate value for the parameter f. There are three options. First, a researcher could pick a fixed value for f, based on ancillary data about the typical abundance of species in clade C. Alternately, the researcher could make a reasonably good estimate of f from the data before

conducting the likelihood analysis. An obvious option is for every species S in C, take

each region $R_{m,t}$ in which S is observed to be present and calculate the empirical

fraction of individuals that are S versus non-S. The mean of all of these empirical

fractions would constitute a decent approximation of f. However, this strategy

ignores that there is some information about the abundance of species even in

regions where they have not been observed, because there is some probability that

they were truly present, but were not sampled in the database. It also ignores the

fact that the phylogenetic relationships and the biogeographic model will confer

some information about the true presences and absences of each species.

Therefore, the final alternative is to make f a free parameter that is jointly estimated

along with the d, e, and j parameters of the DEC+J model (dispersal-extinction

cladogenesis, with founder-event speciation added; Matzke 2013d) or any other

model during maximum likelihood optimization of the model on the data.

A graphical representation of a traditional likelihood-based inference of historical

biogeography is shown in Figure 1. A depiction of the modification that occurs when

a model for imperfect detection is added is also shown.

Empirical demonstration

The time-scaled phylogeny of Equinae used by Maguire and Stigall (2008) was

obtained in Newick format from K. Maguire (personal communication). The Canidae

phylogeny was assembled by digitizing and merging the time-scaled trees published

by Wang et al. (1999) and Tedford et al. (2009) on Borophagines and Caninae, respectively. Digitization was conducted with GraphClick (Arizona Software, http://www.arizona-software.ch/graphclick/) and TreeRogue (Matzke 2013f) as previously described (Matzke 2013d). The phylogenies for both Equinae and Canidae are of unusual quality, in that they combined cladograms derived from parsimony analysis of character data with expert knowledge on the stratigraphic range and geographic distribution of species; thus, unlike most parsimony analyses, directly ancestral species are considered to be known in many cases, and the branch lengths in time can be considered meaningful.

The source of occurrence data for the two clades was the NEOMAP database (http://www.ucmp.berkeley.edu/neomap/) of North American fossil mammal occurrences from the late Oligocene to the present. NEOMAP consists of the merged FAUNMAP (FAUNMAP Working Group 1994; Graham & Lundelius 2010; http://www.ucmp.berkeley.edu/faunmap/) and MIOMAP (Carrasco *et al.* 2005; http://www.ucmp.berkeley.edu/miomap/) databases. NEOMAP was also used as the data source for taphonomic control groups, which were other medium and large mammal taxa present in North America during the Neogene. These were families Agriochoeridae, Amphicyonidae, Antilocapridae, Bovidae, Camelidae, Canidae, Cervidae, Chalicotheriidae, Dichobunidae, Entelodontidae, Equidae, Felidae, Feliformia, Gelocidae, Gomphotheriidae, Hyaenodontidae, Hypertragulidae, Hyracodontidae, Mammutidae, Moschidae, Mustelidae, Nimravidae, Oreodontidae,

Oreodontoidea, Palaeomerycidae, Procyonidae, Protoceratidae, Rhinocerotidae, Ruminantia, Tapiridae, Tayassuidae, and Ursidae.

The four regions used in this study (Figure 2) were defined during preliminary collaborative studies with K. Maguire (Maguire & Matzke 2009; Matzke & Maguire 2011), and are somewhat different than those used in Maguire and Stigall (2008). This decision was made due to the better sampling in the NEOMAP database compared to the Paleobiology Database used by Maguire and Stigall (2008), particularly in the intermountain West (K. Maguire, personal communication). The four regions are the Western U.S., the Rocky Mountains, the Great Plains, and the Gulf Coast (including the Southeast and Florida).

GIS shapefiles containing the outline of the four discrete regions were obtained from K. Maguire (personal communication) and loaded into the R programming environment (R Core Team 2013) for reprojection and analysis with the geospatial data analysis package rgdal (Keitt *et al.* 2012).

The time bins used in the analysis were North American Land Mammal Ages (NALMAs) utilized by NEOMAP. The time-scaled trees were linked to the NEOMAP database data as follows. First, the midpoint of each NALMA was calculated. Then, for each midpoint, each phylogeny was searched for branches crossing that time point. Where branches were present, "hooks" – arbitrarily short branches -- were digitally added to the tree, and given names identifying the species to which they

belonged, or a code if no name was available (i.e., ghost lineages). The hooks represent the existence of a species in a particular time bin, and the points at which geographic occurrence data may be attached to the phylogeny.

Each hook in the resulting trees was henceforth treated as an OTU, i.e., a species sampled at a particular timepoint. For each OTU, the NEOMAP database was searched for occurrences within the appropriate NALMA, using the MNI (minimum number of individuals) field. For localities containing occurrences of the species, the locality latitude and longitude were extracted, and a GIS point-in-polygon operation was used to place the occurrences in a region. The result was that every hook for which data was found contained a count of occurrences of the OTU. In addition, for every NALMA, a count was done in each region of all of the occurrences of the taphonomic control groups. Finally, the phylogeny was pruned to remove all original phylogeny tips, and also to remove any remaining hooks for which no occurrences were found in the database within the study area. R scripts performing these operations are available upon request.

Using the occurrence counts of each OTU and the total counts of taphonomic controls, maximum likelihood (ML) inference of ancestral ranges and model parameter values was conducted for each of the six cladogenesis models described in Matzke (2013e): DEC, DEC+J, DIVA, DIVA+J, BAYAREA, and BAYAREA+J. These six models were run under two different detection models: (1) model M0, with no detection model, i.e., each OTU is treated as present where occurrences were found,

and absent where not; and (2) model D1, where the detection model was used, with *f* estimated as a free parameter.

All models were run in the R package BioGeoBEARS (Matzke 2013b; Matzke 2013a) making use of the required accessory packages rexpokit and cladoRcpp (Matzke 2013c; Matzke & Sidje 2013) as well as the dependencies APE (Paradis 2012) and phylobase (R Hackathon et al. 2012). ML optimization was conducted with optimx (Nash & Varadhan 2011; Nash & Varadhan 2012). One important feature of BioGeoBEARS not introduced previously, but important in the present analysis, is the capability of turning the cladogenesis process on and off for certain nodes. In particular, OTUs that are "hooks" in a tree represent direct ancestors, not sister species (of course, during the pruning process described above, the terminal hooks at the end of pruned branches are converted into standard tips). Direct ancestors are not connected to the tree via a cladogenesis event, instead, they directly represent the ancestral state. Therefore, BioGeoBEARS was set to treat remaining hooks in the pruned phylogenies as direct ancestors in the likelihood calculations.

Models were compared with the likelihood ratio test, AICc, and AICc model weights (Burnham & Anderson 2002), as in previous studies (Matzke 2013d; Matzke 2013e). Ancestral state estimates were plotted on the time-scaled phylogenies for visual comparison.

Results

195

In the pruned phylogeny linked to the geographic occurrence data, 52 OTUs were retained in the Equinae phylogeny, resulting in a tree spanning 17.2 Ma. In the larger Canidae phylogeny, 93 OTUs were retained in a tree spanning 34.9 Ma. An empirical estimate of the fractional abundance of individual OTUs against the background counts of taphonomic control occurrences, averaged across all OTUs, yields an empirical f of 0.031 for Equinae, and 0.0086 for Canidae. This suggests that a typical species from Equinae is about four times more abundant than a typical species from Canidae. This is in accord with the common observation that herbivores are usually much more abundant than carnivores.

Table 1 shows the likelihood scores and parameter inferences for the different models and clades. Interestingly, unlike all but a few clades in previous analyses of cladogenesis models (Matzke 2013e), both Equinae and Canidae fail to reject DEC and DIVA in favor of DEC+J and DIVA+J in likelihood ratio tests with a significance cutoff of 0.05. This was true for both the standard M0 analysis (p=0.33 and 0.48 for Equinae, and p=0.17 and 0.12 for Canidae) and the D1 analysis with imperfect detection (p=0.43 and 0.64 for Equinae, and p=0.13 and 0.32 for Canidae). However, in all analyses, the BAYAREA cladogenesis model was rejected in favor of BAYAREA+J (for Equinae, p=0.012 under M0 and 0.0096 under D1; for Canidae, p=0.0009 under M0 and 0.0036 under D1).

Table 2 shows the AICc scores for the different models and clades. Note that the "data" is actually formally different between the M0 and D1 analysis – M0 used simple OTU presence and absence as the data, where as D1 uses counts of occurrences and taphonomic control occurrences. Model likelihoods and AICc can be directly compared on the same dataset, but are not meaningfully compared between different datasets. However, the relative model probabilities (Table 3) may be directly compared, and the result appears to be high consistency in the weight assigned to each cladogenesis model under M0 and D0. For Equinae, BAYAREA+J is the strongly favored model, with BAYAREA without founder-event speciation trailing. For Canidae, both DEC and DEC+J receive approximately equal support, which in effect means that the DEC+J model has converged closely on the DEC model, with the j parameter close to zero and supplying only a slight advantage.

Although model choice is highly consistent, the two clades tell somewhat different stories in terms of parameter inference (Table 1; Figure 3). For Equinae, all parameter inferences are very similar between M0 and D1 (R^2=0.90 for d, R^2=0.97 for e), with the exception of the inference of j under DIVA+J discussed below. However, for Canidae, there is more difference in the parameter estimates under M0 and D1 (R^2=0.74 for d, R^2=0.89 for e).

The inference of j is also usually similar, with the major exception of DIVA+J under detection model D1. Here, optimx diagnostics warned of convergence problems, and j jumps to over 10 times the value estimated with other models. A hint about what is

going on is shown by examine the inference under DIVA without founder-event

speciation under D1. Here, the f parameter is elevated by an order of magnitude or

more over either the empirical estimate, or the ML estimate under the other models.

It appears that the average relative of abundance of OTUs interacts strongly with the

DIVA cladogenesis model and with the founder-event speciation process in DIVA+J.

The issue needs further exploration; until then, usage of DIVA-derived models in

combination with this model for imperfect detection is discouraged. In any event,

DIVA and DIVA_J are consistently the lowest-scoring cladogenesis models in all

analyses. This is an important result, as the DIVA program, being a parsimony

method not requiring a time-scaled tree, has been one of the most popular programs

for inference of historical biogeography on cladograms that include fossils.

It appears that inference of j and f interact in most analyses (Table 1); when j is a

free parameter, f tends to be much lower, and closer to the empirical estimates of f.

This is reminiscent of one of the results of a detailed comparison of DEC and DEC+J

models (Matzke 2013d), namely, when founder-event speciation is a real process,

but is left out, the model is more likely to infer a positive value for e, even though the

true value is 0. Increasing the value of f means that the chance of failing to sample an

OTU, when the OTU is in fact present, is lower. This means that the model will take

absence data more literally as strong evidence of true absence. Perhaps with the

jump dispersal process added to the cladogenesis models, there is less phylogenetic

conservation of geographic range, and thus it is more difficult for the model to be

confident that an absence of occurrences represents a true absence.

Discussion and Conclusion

The fact that Equinae strongly favor the BAYAREA or BAYAREA+J model, and that Canidae favor DEC or DEC+J (Table 3) suggests that different processes were important in the biogeography of the two groups. Notably, the BAYAREA model completely excludes vicariance, only allowing complete sympatric speciation, where the entire ancestral range is copied to both daughters. This should raise concerns about Brooks' Parsimony Analysis and similar area cladogram-based methods, which rely on a strong prior assumption of the importance of vicariance for explaining distributions. Founder-event speciation does add an allopatric form of speciation to the cladogenesis process in Equinae, but one fundamentally different than vicariance. The founder-event speciation process in the Equinae analyses, while a statistically significant improvement, is much weaker than in most of the island and non-island clades surveyed in previous work (Matzke 2013e).

Canidae, on the other hand, fits traditional models much better, favoring the DEC model used by LAGRANGE (Ree & Smith 2008). It is tempting to suggest that the different models favored by the Equinae and Canidae datasets mean that different cladogenesis processes dominate. It seems plausible that horses in the Miocene and Pliocene could be interpreted as experiencing a large amount of sympatry, due to wide ranges and specialization on different feeding niches (MacFadden 1999). However, fossil Canidae also often exhibited widespread ranges, a large amount of

sympatry, and niche differentiation based on body size and feeding specialization (Wang, Tedford & Taylor 1999; Tedford, Wang & Taylor 2009), so this explanation is not entirely satisfactory. Inspection of Figures 4 and 5, which show the estimates of ancestral ranges for Equinae and Canidae under their respective best models, give somewhat more guidance. The Canidae phylogeny (Figure 5) has a large number of nodes where direct ancestors ("hooks") had widespread ranges with occurences in all areas (a geographic range of ABCD), and later descendants had a mixture of widespread and narrow descendants. This is a pattern expected from simulations under the DEC model, where sympatric-subset speciation is an important process (Matzke 2013d; Matzke 2013e). The horses (Figure 4), on the other hand, rarely had such widespread ancestors, often occupying only 2 or 3 areas. This may indicate that horses were more restricted by area-specific habitat (e.g. grasslands versus forest), where as the carnivorous canids were effectively more generalist in habitat preferences, finding prey or scavenging opportunities in all habitats.

In conclusion, this study has demonstrated several points. First and most importantly, we have shown that fossil geographic range data may be incorporated into likelihood inferences of geographic range data in much the same way that living species are currently used, as long as the computational implementation includes methods to distinguish terminal OTUs and direct ancestor OTUs. Inclusion of fossils could already be done in LAGRANGE, but only if each fossil OTU was a terminal taxon and not a direct ancestor, and only if the analysis was not time-stratified. BioGeoBEARS accounts for both contingencies. Second, we have demonstrated how

a simple model of imperfect detection can be used in likelihood analyses of fossil data, and its influence on inference of parameters and models assessed. In the Equinae and Canidae, the effect of an imperfect detection model was minimal in terms of model choice and ancestral range estimation. These groups were chosen primarily because of their extremely good fossil record and taxonomic status, out of the concern that weaker groups would lack sufficient data to reliably infer parameters and true presence and absence. Ironically, this choice may have meant that the data were "too good", resulting in little difference between inference under the assumption that a face-value reading of occurrences gave the true range, and inference under a model of imperfect detection. If the fossil data are good enough, imperfect detection models are not needed! An obvious future study would apply this method to groups with weaker fossil records, or alternatively, take the Equinae and Canidae and progressively down-sample their occurrences to assess when a detection model is useful, and where reliable inference breaks down completely.

Acknowledgements

Table 1. Parameter inferences and likelihoods for cladogenesis models with and without the model for imperfect detection.

Clade	Model	Det. Model	# tips	LnL	# params	*d*	*e*	*j*	*f*
Equinae	DEC	M0	52	-146.9	2	0.13	0.10	0	NA
Equinae	DEC+J	M0	52	-146.4	3	0.12	0.09	0.03	NA
Equinae	DIVA	M0	52	-153.9	2	0.18	0.14	0	NA
Equinae	DIVA+J	M0	52	-153.7	3	0.17	0.13	0.03	NA
Equinae	BAYAREA	M0	52	-142.2	2	0.12	0.23	0	NA
Equinae	BAYAREA+J	M0	52	-139.1	3	0.09	0.15	0.04	NA
Equinae	DEC+f	D1	52	-139.9	3	0.13	0.10	0	0.07
Equinae	DEC+J+f	D1	52	-139.6	4	0.13	0.09	0.02	0.01
Equinae	DIVA+f	D1	52	-146.4	3	0.18	0.12	0	0.10
Equinae	DIVA+J+f	D1	52	-146.4	4	0.19	0.15	0.46	0.01
Equinae	BAYAREA+f	D1	52	-135.3	3	0.12	0.22	0	0.01
Equinae	BAYAREA+J+f	D1	52	-131.9	4	0.09	0.13	0.04	0.01
Canidae	DEC	M0	93	-252.6	2	0.12	0.06	0	NA
Canidae	DEC+J	M0	93	-251.7	3	0.11	0.05	0.05	NA
Canidae	DIVA	M0	93	-261.4	2	0.14	0.06	0	NA
Canidae	DIVA+J	M0	93	-260.2	3	0.13	0.05	0.06	NA
Canidae	BAYAREA	M0	93	-265.2	2	0.13	0.19	0	NA
Canidae	BAYAREA+J	M0	93	-259.7	3	0.11	0.14	0.09	NA
Canidae	DEC+f	D1	93	-155.4	3	0.19	0.04	0	0.32
Canidae	DEC+J+f	D1	93	-154.2	4	0.16	0.01	0.10	0.01
Canidae	DIVA+f	D1	93	-161.4	3	0.23	0.05	0	0.24
Canidae	DIVA+J+f	D1	93	-160.9	4	0.24	0.07	1	0.01
Canidae	BAYAREA+f	D1	93	-160.5	3	0.20	0.18	0	0.11
Canidae	BAYAREA+J+f	D1	93	-156.2	4	0.14	0.11	0.17	0.01

Table 2. AICc values for each model.

Clade: Detection:	Equinae M0	Equinae D1	Canidae M0	Canidae D1
DEC	298.1	286.4	509.4	317.1
DEC+J	299.3	288.1	509.7	316.9
DIVA	312.1	299.4	527.0	329.1
DIVA+J	313.9	301.7	526.7	330.3
BAYAREA	288.7	277.2	534.6	327.3
BAYAREA+J	284.7	272.7	525.7	320.9

Table 3. Relative model probabilities, based on AICc.

Clade: Dispersal:	Equinae M0	Equinae D1	Canidae M0	Canidae D1
DEC	0%	0%	54%	43%
DEC+J	0%	0%	46%	50%
DIVA	0%	0%	0%	0%
DIVA+J	0%	0%	0%	0%
BAYAREA	12%	9%	0%	0%
BAYAREA+J	88%	91%	0%	7%

CAPTIONS

Figure 1A. Traditional inference procedure in likelihood analyses of historical biogeography on a phylogeny.

Figure 1B. DAG (Directed Acyclic Graph; Link and Barker, 2009) for the model of imperfect detection implemented in BioGeoBEARS. DAGs provide simple graphical representations probabilistic models, e.g. for hierarchical likelihood or Bayesian modeling. In a DAG, fixed data (grey) is hypothesized to be a function (arrows) of a model with unknown parameters (white).

Figure 2. The discrete geographic regions used in this analysis.

Figure 3. The ML estimates of parameters under the M0 model (x-axis) and D1 model of imperfect detection (y-axis).

Figure 4. BioGeoBEARS graphical output, showing the Equinae phylogeny and most-probable ancestral ranges (4A) or pie charts of all ancestral states (4B). These are the most-probable ancestral ranges under the global ML model parameters, under the best model for Equinae, BAYAREA+J, using the D1 detection model. Regions are as follows A (blue): Western United States; B (green): Rocky Mountains; C (yellow) Great Plains; D (red) Gulf Coast. Combinations of areas are represented by mixing colors, and white represents a range of all areas (ABCD). The x-axis actually depicts millions of years before the mid-point of the Blancan NALMA, where the latest-occurring OTU in the analysis is found.

Figure 5. BioGeoBEARS graphical output, showing the Canidae phylogeny and most-probable ancestral ranges (5A) or pie charts of all ancestral states (5B). These are the most-probable ancestral ranges under the global ML model parameters, under the best model for Equinae, BAYAREA+J, using the D1 detection model. Regions are as follows A (blue): Western United States; B (green): Rocky Mountains; C (yellow) Great Plains; D (red) Gulf Coast. Combinations of areas are represented by mixing colors, and white represents a range of all areas (ABCD).
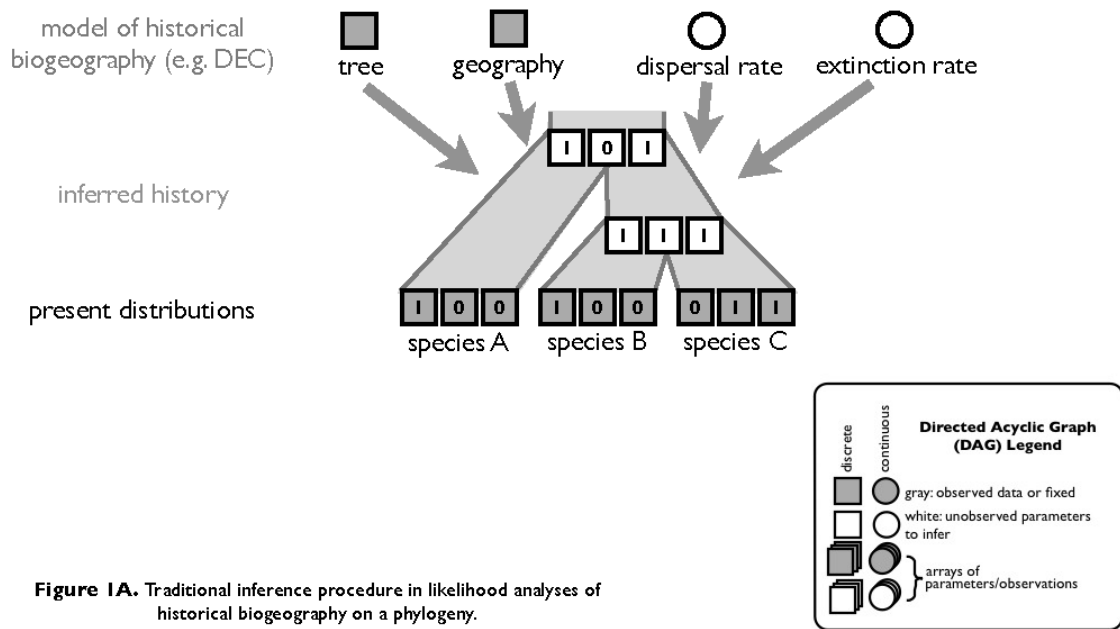
model of historical
biogeography (e.g. DEC)

tree   geography   dispersal rate   extinction rate

| 1 | 0 | 1 |

inferred history

| 1 | 1 | 1 |

present distributions

| 1 | 0 | 0 | | 1 | 0 | 0 | | 0 | 1 | 1 |

species A   species B   species C

**Directed Acyclic Graph
(DAG) Legend**

discrete   continuous

gray: observed data or fixed

white: unobserved parameters
to infer

} arrays of
parameters/observations

Figure 1A. Traditional inference procedure in likelihood analyses of
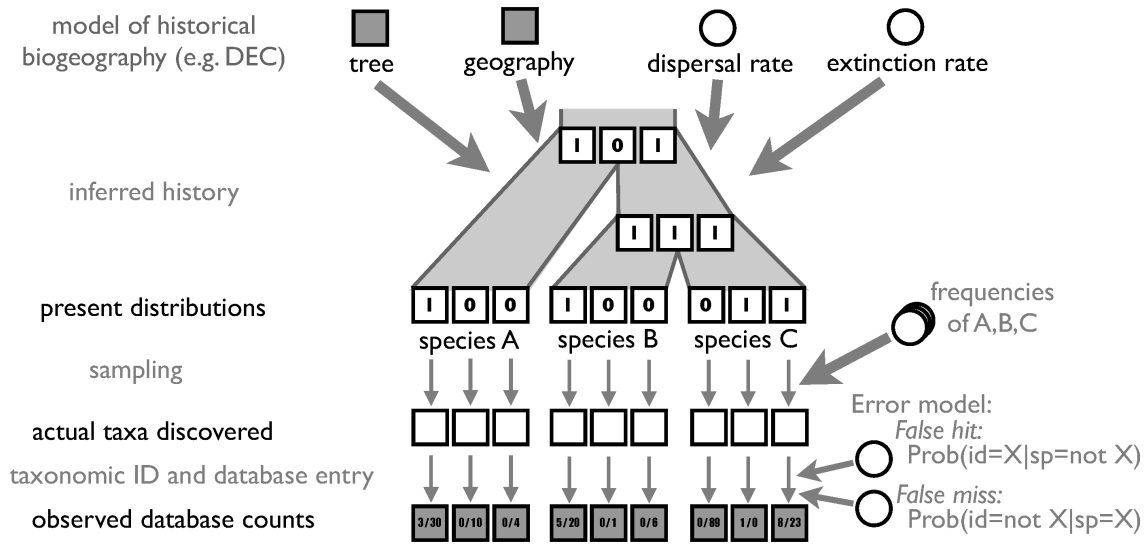historical biogeography on a phylogeny.

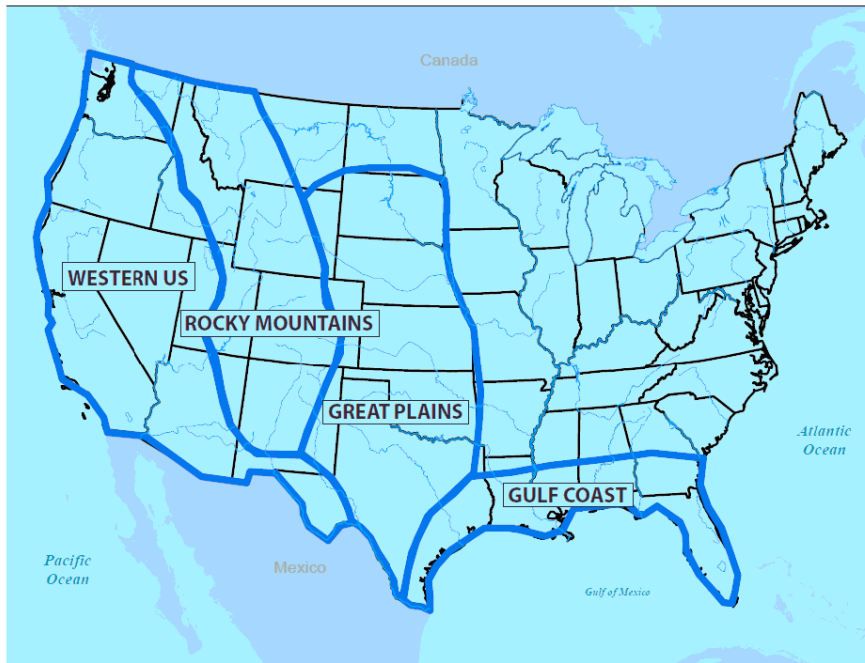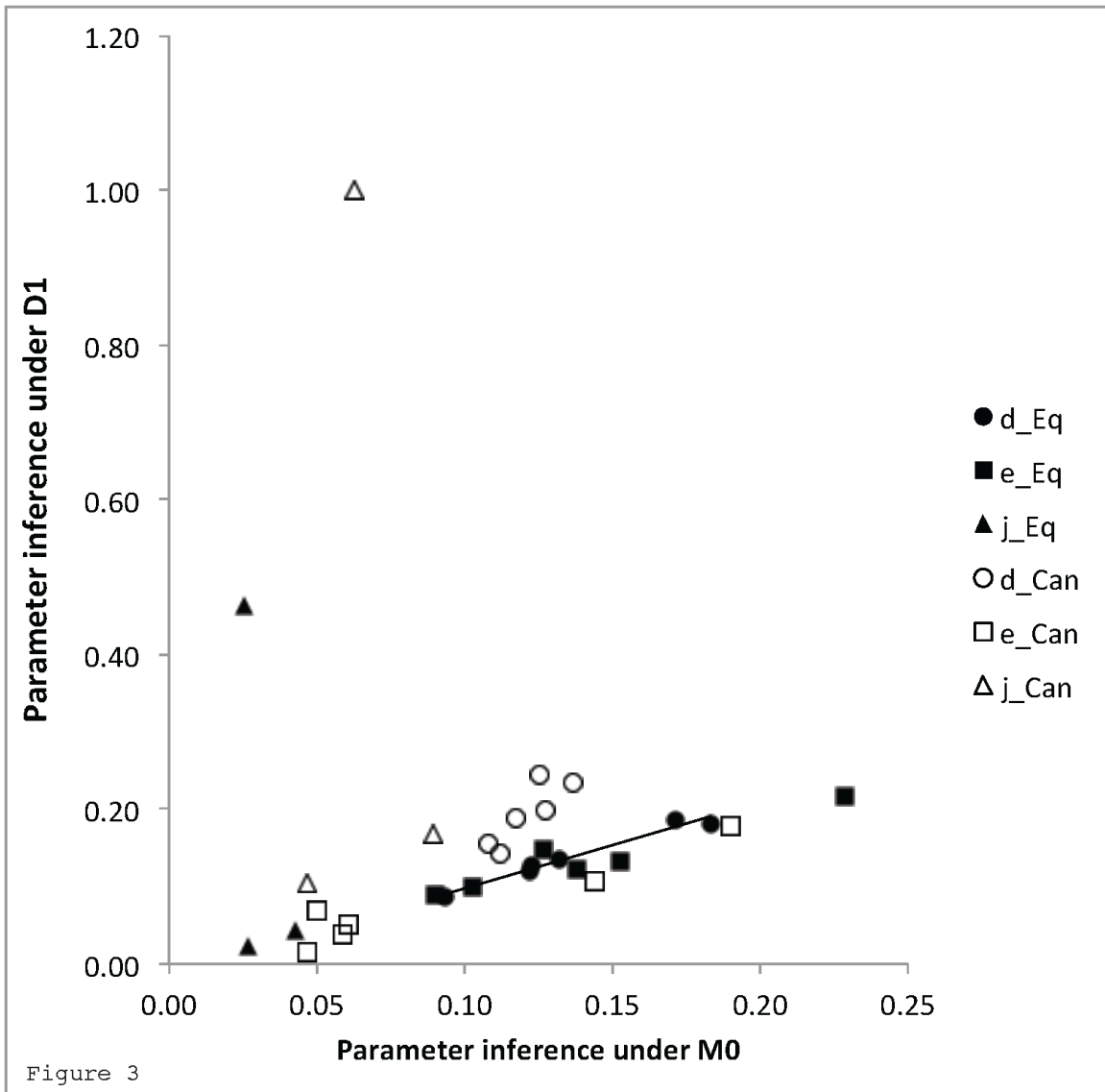**Figure 1B.** DAG (Directed Acyclic Graph) for the model of imperfect detection implemented in BioGeoBEARS.

Figure 2.

Figure 3

BioGeoBEARS BAYAREA on Equinae D1_detectionModel
anstates: global optim, 4 areas max. d=0.119; e=0.216; mf=0.005; j=0; LnL=−135.3

Millions of years ago
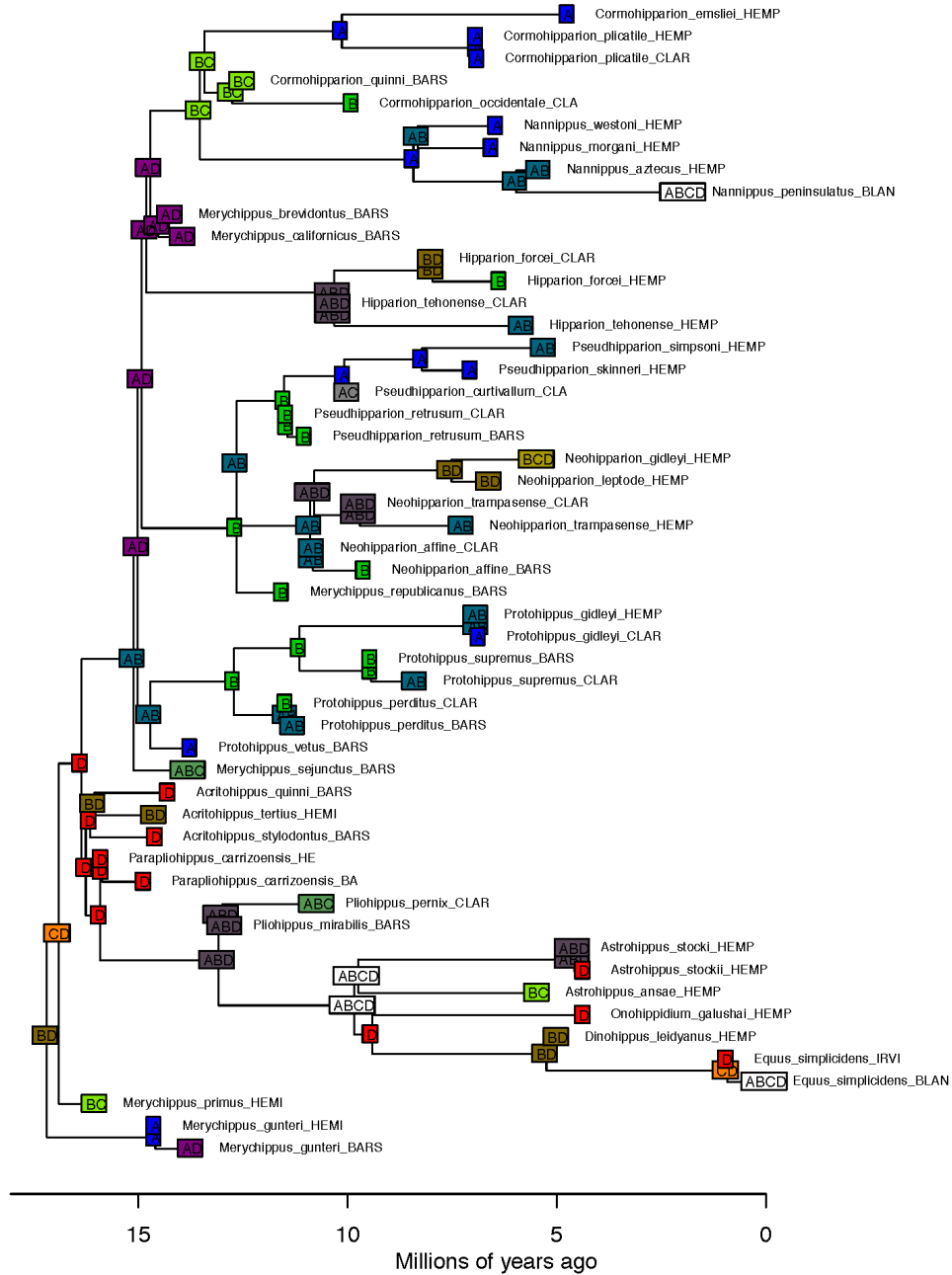
Figure 4A

210

BioGeoBEARS BAYAREA on Equinae D1_detectionModel
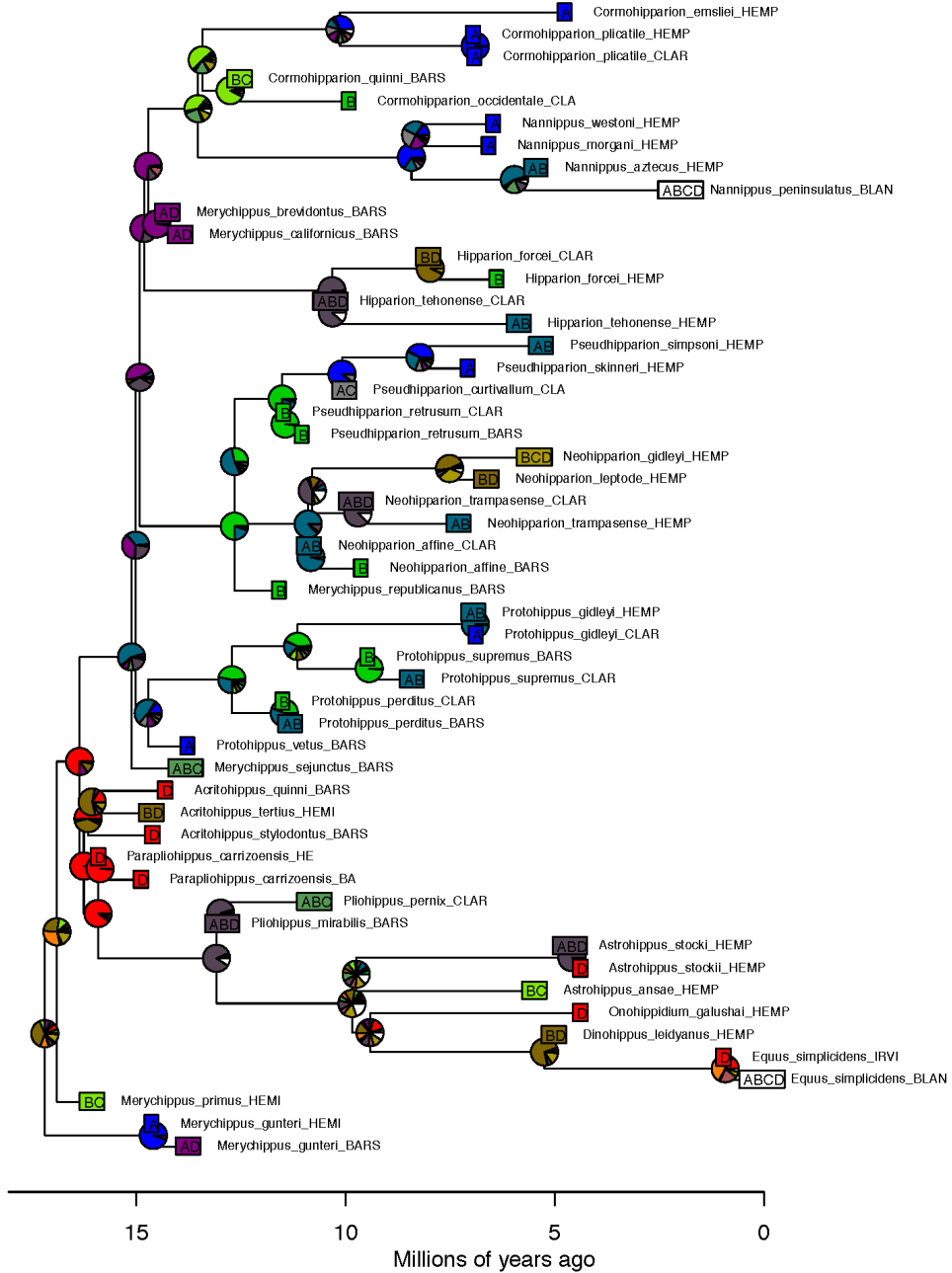anstates: global optim, 4 areas max. d=0.119; e=0.216; mf=0.005; j=0; LnL=−135.3

Millions of years ago

Figure 4B

BioGeoBEARS DEC on Canidae D1_detectionModel
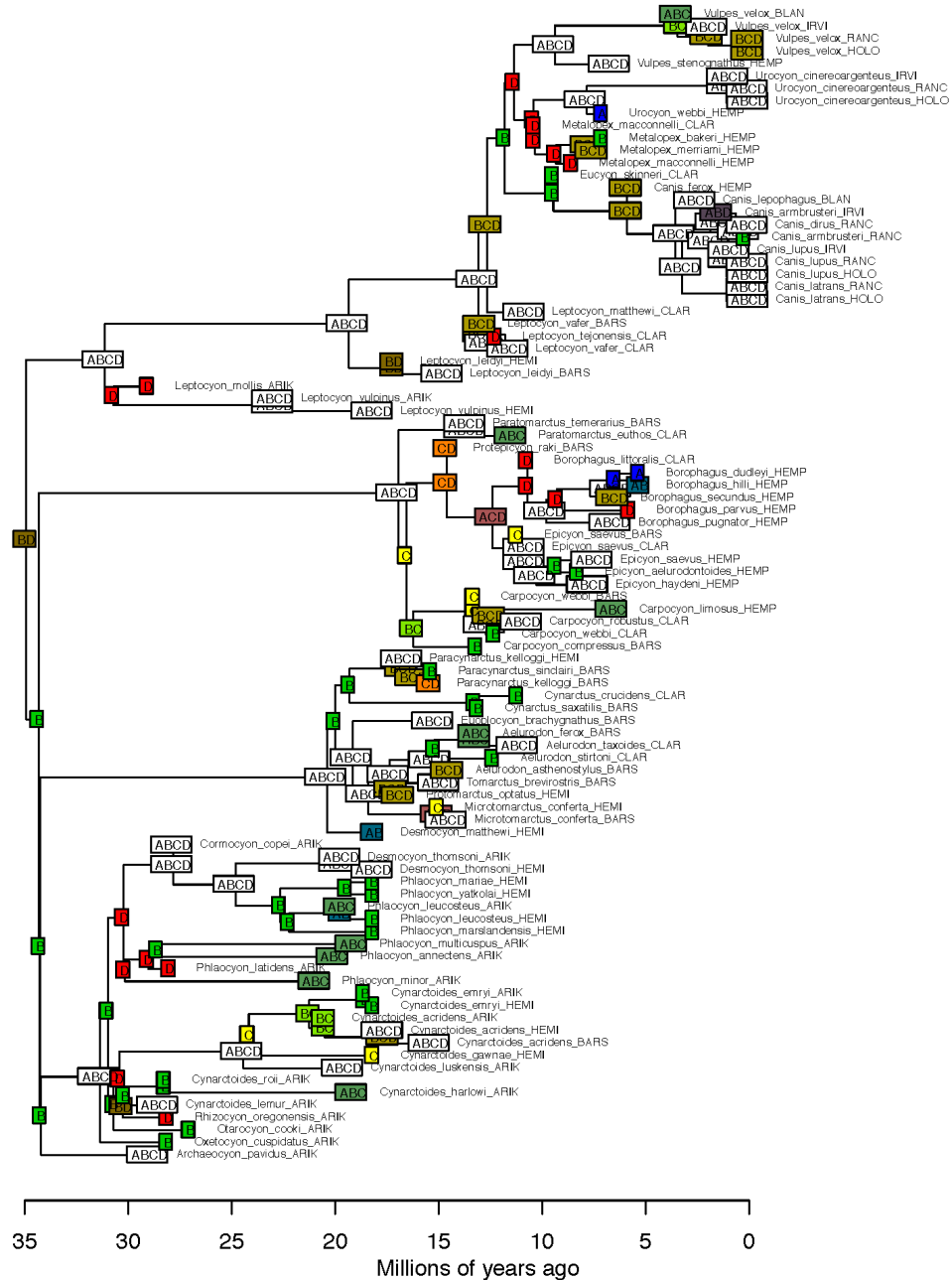anstates: global optim, 4 areas max. d=0.188; e=0.039; mf=0.322; j=0; LnL=−155.4

Millions of years ago

Figure 5A

BioGeoBEARS DEC on Canidae D1_detectionModel
anstates: global optim, 4 areas max. d=0.188; e=0.039; mf=0.322; j=0; LnL=−155.4
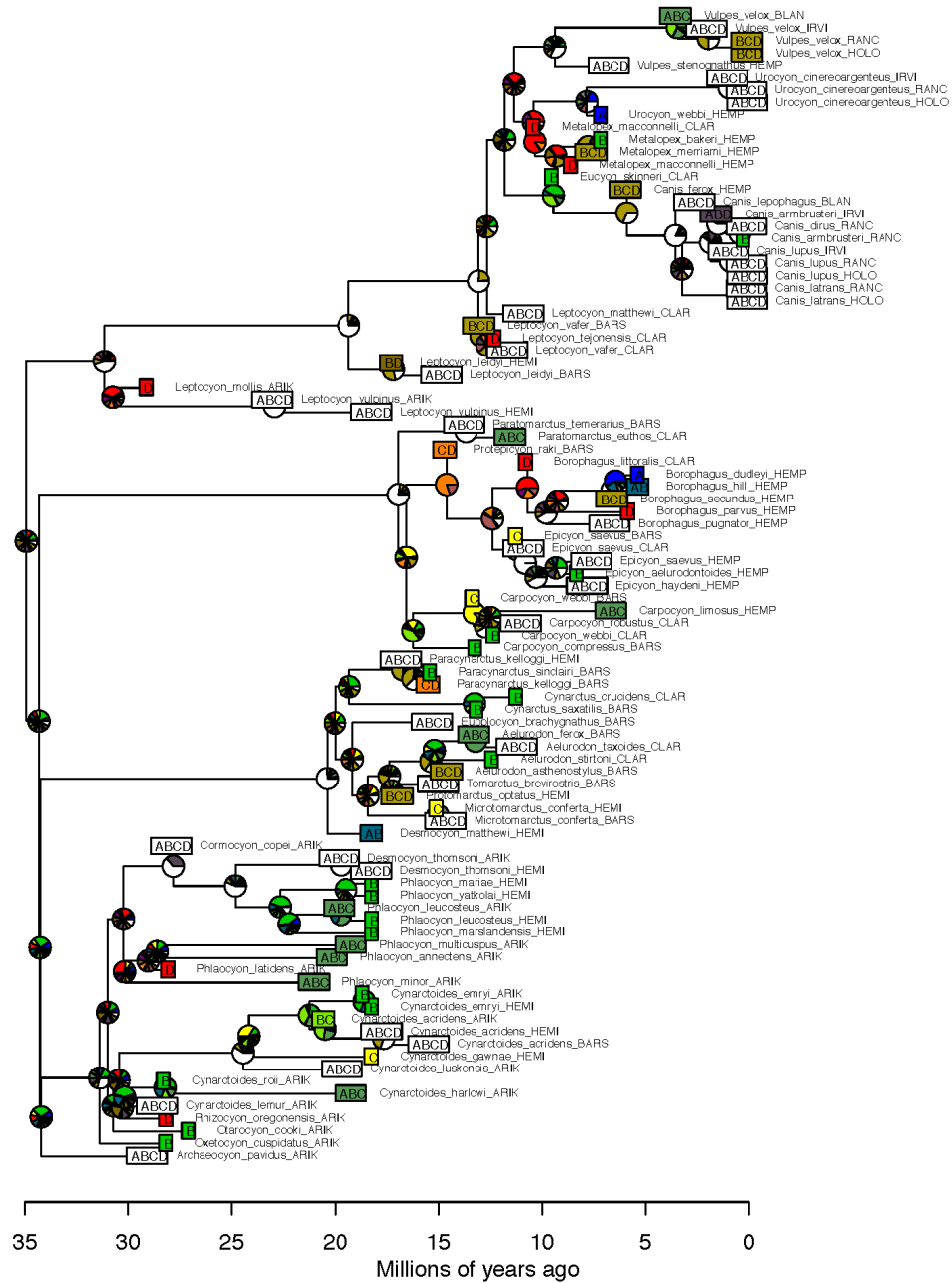
Millions of years ago

Figure 5B

213

References

Alexandrou, M.A., Swartz, B.A., Matzke, N.J. & Oakley, T.H. (2013) Genome
duplication and the convergent evolutionary origins of complex migratory
behavior in salmonid fishes. *Systematic Biology,* **Submitted**.

Ali, S.S., Yu, Y., Pfosser, M. & Wetschnig, W. (2012) Inferences of biogeographical
histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary
MCMC analysis implemented in RASP (Reconstruct Ancestral State in
Phylogenies). *Annals of Botany,* **109,** 95-107.

Bailey, L.L., Hines, J.E., Nichols, J.D. & MacKenzie, D.I. (2007) Sampling design trade-
offs in occupancy studies with imperfect detection: examples and software.
Ecological Applications, **17,** 281-290.

Baldwin, B.G. & Sanderson, M.J. (1998) Age and rate of diversification of the
Hawaiian silversword alliance (Compositae). *Proceedings of the National
Academy of Sciences,* **95,** 9402-9406.

Bapst, D.W. (2013) A stochastic rate-calibrated method for time-scaling phylogenies
of fossil taxa. *Methods in Ecology and Evolution,* **4,** 724-733.

Benavides, E., Baum, R., Snell, H.M., Snell, H.L. & Sites, J.W., Jr. (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). *Journal of Biogeography,* **40,** 1512-1523.

Bennett, G.M. & O'Grady, P.M. (2011) Review of the native Hawaiian leafhopper

genus Nesophrosyne (Hemiptera: Cicadellidae: Deltocephalinae) with

description of eight new species associated with Broussaisia arguta

(Hydrangeaceae). *Zootaxa,* **2805,** 1-25.

Benton, M.J. (2009) The Fossil Record: Biological or Geological Signal? *The*

paleobiological revolution : essays on the growth of modern paleontology (eds

D. Sepkoski & M. Ruse), pp. 43-59. University of Chicago Press, Chicago.

Bernardi, G., Bucciarelli, G., Costagliola, D., Robertson, D.R. & Heiser, J.B. (2004)

Evolution of coral reef fish Thalassoma spp. (Labridae). 1. Molecular

phylogeny and biogeography. *Marine Biology,* **144,** 369-375.

Boettiger, C. & Temple Lang, D. (2012) Treebase: an R package for discovery, access

and manipulation of online phylogenies. *Methods in Ecology and Evolution,* **3,**

1060-1066.

Bottjer, D.J., Droser, M.L. & Jablonski, D. (1988) Palaeoenvironmental trends in the

history of trace fossils. *Nature,* **333,** 252-255.

Bottjer, D.J. & Jablonski, D. (1988) Paleoenvironmental Patterns in the Evolution of

Post-Paleozoic Benthic Marine Invertebrates. *PALAIOS,* **3,** 540-560.

Box, G.E.P. & Draper, N.R. (1987) *Empirical model-building and response surfaces.*

Wiley, New York.

Bremer, K. (1992) Ancestral Areas: A Cladistic Reinterpretation of the Center of

Origin Concept. *Systematic Biology,* **41,** 436-445.

Brown, G.K., Nelson, G. & Ladiges, P.Y. (2006) Historical biogeography of

    Rhododendron section Vireya and the Malesian Archipelago. *Journal of*

    Biogeography, **33,** 1929-1944.

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. *Journal of Biogeography,* **38,** 531-550.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference : a*

    practical information-theoretic approach, 2nd edn. Springer, New York.

Byrd, R.H., Lu, P., Nocedal, J. & Zhu, C. (1995) A limited memory algorithm for bound

    constrained optimization. *SIAM J. Scientific Computing,* **16,** 1190-1208.

Carlquist, S.J. (1965) *Island life; a natural history of the islands of the world,* 1st edn.

    Published for the American Museum of Natural History by the Natural

    History Press, Garden City, N.Y.

Carlquist, S.J. (1974) *Island biology*. Columbia University Press, New York,.

Carrasco, M.A., Kraatz, B.P., Davis, E.B. & Barnosky, A.D. (2005) Miocene Mammal

    Mapping Project (MIOMAP). University of California Museum of

    Paleontology.

Carson, H.L. (1974) Patterns of speciation in Hawaiian Drosophila inferred from

    ancient chromosomal polymorphism. *Genetic Mechanisms of Speciation in*

    Insects (ed. M.J.D. White), pp. 81-93. Springer Netherlands.

Carson, H.L. & Templeton, A.R. (1984) Genetic Revolutions in Relation to Speciation

     Phenomena: The Founding of New Populations. *Annual Review of Ecology and*

     Systematics, **15,** 97-131.

Ceccarelli, F.S. & Zaldívar-Riverón, A. (2013) Broad polyphyly and historical

     biogeography of the neotropical wasp genus Notiospathius (Braconidae:

     Doryctinae). *Molecular Phylogenetics and Evolution,* **69,** 142-152.

Clague, D.A. (1996) The growth and subsidence of the Hawaiian-Emperor volcanic

     chain. *The Origin and Evolution of Pacific Island Biotas, New Guinea to Eastern*

     Polynesia: Patterns and Processes (eds A. Keast & S.E. Miller), pp. 35-50. SPB

     Academic Publishing, Amsterdam.

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. *Systematic Biology,* **57,** 693-707.

Clark, J.R., Wagner, W.L. & Roalson, E.H. (2009) Patterns of diversification and

     ancestral range reconstruction in the southeast Asian–Pacific angiosperm

     lineage Cyrtandra (Gesneriaceae). *Molecular Phylogenetics and Evolution,* **53,**

     982-994.

Cowie, R.H. & Holland, B.S. (2006) Dispersal is fundamental to biogeography and the

     evolution of biodiversity on oceanic islands. *Journal of Biogeography,* **33,**

     193-198.

Coyne, J.A. & Orr, H.A. (2004) *Speciation*. Sinauer Associates, Sunderland, Mass.

Cracraft, J. & Shipp, R.L. (1988) Vicariance Biogeography: Theory, Methods, and Applications: Introduction to the Symposium. *Systematic Zoology,* **37,** 219-220.

Crisci, J.V. (2001) The voice of historical biogeography. *Journal of Biogeography,* **28,** 157-168.

Cunningham, C.W., Omland, K.E. & Oakley, T.H. (1998) Reconstructing ancestral character states: a critical reappraisal. *Trends in Ecology & Evolution,* **13,** 361-366.

Darwin, C. (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.,* 1st edn. John Murray, London.

de Queiroz, A. (2005) The resurrection of oceanic dispersal in historical biogeography. *Trends Ecol Evol,* **20,** 68-73.

DeSalle, R. & Templeton, A.R. (1988) Founder Effects and the Rate of Mitochondrial DNA Evolution in Hawaiian Drosophila. *Evolution,* **42,** 1076-1084.

Diaz Gomez, J.M. (2011) Estimating Ancestral Ranges: Testing Methods with a Clade of Neotropical Lizards (Iguania: Liolaemidae). *PLoS ONE,* **6,** e26412.

Donoghue, M.J. & Moore, B.R. (2003) Toward an Integrative Historical Biogeography. *Integrative and Comparative Biology,* **43,** 261-270.

Donovan, S.K. & Paul, C.R.C. (1998) The adequacy of the fossil record. pp. x, 312 p. John Wiley, Chichester ; New York.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology,* **7,** 214.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian Phylogenetics
with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution,* **29,** 1969-
1973.

Dunbar-Co, S. (2008) Island Evolution: Phylogeny, Adaptive Radiation, and
Biogeography of *Plantago* (Plantaginaceae) in the Hawaiian Islands. Ph.D.,
University of Hawai'i.

Ebach, M.C. & Tangney, R.S. (2007) *Biogeography in a changing world*. CRC Press,
Boca Raton, FL.

Eddelbuettel, D. & Francois, R. (2013) Rcpp: Seamless R and C++ Integration. CRAN:
The Comprehensive R Archive Network.

Eddelbuettel, D., Francois, R. & Bates, D. (2013) RcppArmadillo: Rcpp integration for
Armadillo templated linear algebra library. CRAN: The Comprehensive R
Archive Network.

Edgecombe, G.D. (2010) Palaeomorphology: fossils and the inference of cladistic
relationships. *Acta Zoologica,* **91,** 72-80.

Escoto, B. (2011) Marginal Likelihood Computation via Arrogance Sampling.
*arXiv.org***,** 1-12.

Ezard, T.H.G., Thomas, G.H. & Purvis, A. (2013) Inclusion of a near-complete fossil
record reveals speciation-related molecular evolution. *Methods in Ecology
and Evolution,* **4,** 745-753.

FAUNMAP Working Group (1994) FAUNMAP: a database documenting late
Quaternary distributions of mammal species in the United States. *Illinois
State Museum Scientific Papers,* **25,** 1-690.

Felsenstein, J. (1981) Evolutionary Trees From Gene Frequencies and Quantitative Characters: Finding Maximum Likelihood Estimates. *Evolution,* **35,** 1229-1242.

Felsenstein, J. (2004) *Inferring Phylogenies.* Sinauer Associates, Sunderland, Massachusetts.

Foote, M. (1996) On the Probability of Ancestors in the Fossil Record. *Paleobiology, 22,* 141-151.

Fortey, R.A. (2009) Biogeography and Evolution in the Early Paleozoic. *The paleobiological revolution : essays on the growth of modern paleontology* (eds D. Sepkoski & M. Ruse), pp. 60-72. University of Chicago Press, Chicago.

Gillespie, R.G. & Baldwin, B.G. (2010) Island biogeography of remote archipelagos: Interplay between ecological and evolutionary processes. *The theory of island biogeography revisited* (eds J.B. Losos, R.E. Ricklefs & R.H. MacArthur), pp. 358-387. Princeton University Press, Princeton.

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 in Ecology & Evolution,* **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. *Systematic Biology,* **60,** 451-465.

Goloboff, P.A., Farris, J.S. & Nixon, K.C. (2008) TNT, a free program for phylogenetic analysis. *Cladistics,* **24,** 774-786.

Gong, Z., Matzke, N.J., Ermentrout, B., Song, D., Vendetti, J.E., Slatkin, M. & Oster, G. (2012) Evolution of patterns on Conus shells. *Proceedings of the National Academy of Sciences,* **109,** E234-E241.

Gould, S.J. & Eldredge, N. (1972) Punctuated equilibria: an alternative to phyletic gradualism. *Models in paleobiology***,** 82-115.

Graham, R.W. & Lundelius, J., E. L. (2010) FAUNMAP II: New data for North America with a temporal extension for the Blancan, Irvingtonian and early Rancholabrean.

Green, P.J. (1995) Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika,* **82,** 711-732.

Grehan, J.R. & Schwartz, J.H. (2009) Evolution of the second orangutan: phylogeny and biogeography of hominid origins. *Journal of Biogeography,* **36,** 1823-1844.

Hampton, L.C. & Kaneshiro, K.Y. (1976) *Drosophila* of Hawaii: Systematics and Ecological Genetics. *Annual Review of Ecology and Systematics,* **7,** 311-345.

Hardy, C.R. & Linder, H.P. (2005) Intraspecific Variability and Timing in Ancestral Ecology Reconstruction: A Test Case from the Cape Flora. *Systematic Biology,* **54,** 299-316.

Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno Roark, F. & Near, T.J. (2010)

Early bursts of body size and shape evolution are rare in comparative data. *Evolution,* **64,** 2385-2396.

Harris, A.J., Wen, J. & Xiang, Q.-Y. (2013) Inferring the biogeographic origins of inter-continental disjunct endemics using a Bayes-DIVA approach. *Journal of Systematics and Evolution,* **51,** 117-133.

Harte, J. (2011) *Maximum entropy and ecology : a theory of abundance, distribution, and energetics*. Oxford University Press, Oxford ; New York.

Havill, N.P., Campbell, C.S., Vining, T.F., LePage, B., Bayer, R.J. & Donoghue, M.J. (2008) Phylogeny and biogeography of Tsuga (Pinaceae) inferred from nuclear ribosomal ITS and chloroplast DNA sequence data. *Systematic botany,* **33,** 478-789.

Heads, M.J. (2012) *Molecular panbiogeography of the tropics*. University of California Press, Berkeley.

Hennig, W. (1966) *Phylogenetic systematics*. University of Illinois Press, Urbana,.

Herrera, N.D. (2013) Molecular Phylogenetics and Historical Biogeography of Cockles and Giant Clams (Bivalvia: Cardiidae). Master of Science, The Florida State University.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T. & Antonelli, A. (2010) Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science,* **330,** 927-931.

Hormiga, G. (2002) *Orsonwelles*, a new genus of giant linyphiid spiders (Araneae) from the Hawaiian Islands. *Invertebrate Systematics,* **16,** 369-448.

Hormiga, G., Arnedo, M. & Gillespie, R.G. (2003) Speciation on a Conveyor Belt: Sequential Colonization of the Hawaiian Islands by *Orsonwelles* Spiders (Araneae, Linyphiidae). *Systematic Biology,* **52,** 70-88.

Huelsenbeck, J.P. (1995) Performance of Phylogenetic Methods in Simulation. *Systematic Biology,* **44,** 17-48.

Huelsenbeck, J.P., Alfaro, M.E. & Suchard, M.A. (2011) Biologically Inspired Phylogenetic Models Strongly Outperform the No Common Mechanism Model. *Systematic Biology,* **60,** 225-232.

Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. (2003) Stochastic Mapping of Morphological Characters. *Systematic Biology,* **52,** 131-158.

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.

Hunt, G. (2013) Testing the link between phenotypic evolution and speciation: an integrated palaeontological and phylogenetic analysis. *Methods in Ecology and Evolution,* **4,** 714-723.

Jablonski, D., Flessa, K.W. & Valentine, J.W. (1985) Biogeography and Paleobiology. *Paleobiology,* **11,** 75-90.

Jeffreys, H. (1961) *The theory of probability*. Oxford University Press, Oxford.

Jordan, S., Simon, C. & Polhemus, D. (2003) Molecular Systematics and Adaptive Radiation of Hawaii's Endemic Damselfly Genus Megalagrion (Odonata: Coenagrionidae). *Systematic Biology,* **52,** 89-109.

Jukes, T.H. & Cantor, C.R. (1969) *Evolution of protein molecules.* Academic Press, New York.

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. *Current Biology,* **5,** 1129-1139.

Keitt, T.H., Bivand, R., Pebesma, E. & Rowlingson, B. (2012) rgdal: Bindings for the Geospatial Data Abstraction Library. CRAN: The Comprehensive R Archive Network.

Kelly, C.M.R., Barker, N.P., Villet, M.H. & Broadley, D.G. (2009) Phylogeny, biogeography and classification of the snake superfamily Elapoidea: a rapid radiation in the late Eocene. *Cladistics,* **25,** 38-63.

Kéry, M. & Schmidt, B.R. (2008) Imperfect detection and its consequences for monitoring for conservation. *Community Ecology,* **9,** 207-216.

Kidwell, S.M. (2001) Preservation of Species Abundance in Marine Death Assemblages. *Science,* **294,** 1091-1094.

Kidwell, S.M. (2002) Time-averaged molluscan death assemblages: Palimpsests of richness, snapshots of abundance. *Geology,* **30,** 803-806.

Kidwell, S.M. & Holland, S.M. (2002) The Quality of the Fossil Record: Implications for Evolutionary Analyses. *Annual Review of Ecology and Systematics,* **33,** 561-588.

Klein, R.G. & Cruz-Uribe, K. (1984) *The analysis of animal bones from archeological sites*. University of Chicago Press, Chicago.

Kodandaramaiah, U. (2010) Use of dispersal–vicariance analysis in biogeography – a critique. *Journal of Biogeography,* **37,** 3-11.

Köhler, G. & Vesely, M. (2010) A Revision of the *Anolis sericeus* Complex with the Resurrection of *A. wellbornae* and the Description of a New Species (Squamata: Polychrotidae). *Herpetologica,* **66,** 207-228.

Laliberté, E. & Legendre, P. (2010) A distance-based framework for measuring functional diversity from multiple traits. *Ecology,* **91,** 299-305.

Laliberté, E. & Shipley, B. (2011) FD: measuring functional diversity from multiple traits, and other tools for functional ecology. CRAN: The Comprehensive R Archive Network.

Landis, M., Matzke, N.J., Moore, B.R. & Huelsenbeck, J.P. (2013) Bayesian Analysis of Biogeography when the Number of Areas is Large. *Systematic Biology*.

Lapoint, R.T., O'Grady, P.M. & Whiteman, N.K. (2013) Diversification and dispersal of the Hawaiian Drosophilidae: The evolution of Scaptomyza. *Molecular Phylogenetics and Evolution,* **69,** 95-108.

Lerner, Heather R.L., Meyer, M., James, Helen F., Hofreiter, M. & Fleischer, Robert C. (2011) Multilocus Resolution of Phylogeny and Timescale in the Extant Adaptive Radiation of Hawaiian Honeycreepers. *Current Biology,* **21,** 1838-1844.

Lieberman, B. (2004) Range Expansion, Extinction, and Biogeographic Congruence: A Deep Time Perspective. *Frontiers of biogeography : new directions in the*

geography of nature (eds M.V. Lomolino & L.R. Heaney), pp. 111-124. Sinauer Associates, Sunderland, Mass.

Lieberman, B.S. (2000) *Paleobiogeography*. Kluwer Academic/Plenum Publishers, New York.

Lieberman, B.S. (2002) Phylogenetic biogeography with and without the fossil record: gauging the effects of extinction and paleontological incompleteness. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **178,** 39-52.

Lieberman, B.S. (2003) Paleobiogeography: The Relevance of Fossils to Biogeography. *Annual Review of Ecology, Evolution, and Systematics,* **34,** 51-69.

Lieberman, B.S. (2005) Geobiology and paleobiogeography: tracking the coevolution of the Earth and its biota. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **219,** 23-33.

Link, W.A. & Barker, R.J. (2009) *Bayesian Inference: with ecological applications*. Academic Press.

Lomolino, M.V., Lomolino, M.V. & Lomolino, M.V. (2010) *Biogeography,* 4th edn. Sinauer Associates, Sunderland, Mass.

MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton.

MacFadden, B.J. (1999) *Fossil Horses: Systematics, Paleobiology, and Evolution of the Family Equidae*. Cambridge University Press, Cambridge.

Mackenzie, D.I. (2005) Was it there? Dealing with imperfect detection for species presence/absence data. *Australian & New Zealand Journal of Statistics,* **47,** 65-74.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.B. (2003) Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology,* **84,** 2200-2207.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology,* **83,** 2248-2255.

MacKenzie, D.I., Nichols, J.D., Seamans, M.E. & Gutiérrez, R.J. (2009) Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology,* **90,** 823-835.

Maguire, K.C. & Matzke, N.J. (2009) Maximum Likelihood estimation of biogeographic histories using the fossil record. *Journal of Vertebrate Paleontology,* **29,** 7A.

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. *Palaeogeography Palaeoclimatology Palaeoecology,* **267,** 175-184.

Mahler, D.L., Ingram, T., Revell, L.J. & Losos, J.B. (2013) Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. *Science,* **341,** 292-295.

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. *Molecular Phylogenetics and Evolution,* **66,** 54-68.

Matzke, N.J. (2013a) BioGeoBEARS: an R package for model testing and ancestral state estimation in historical biogeography. *Methods in Ecology and Evolution, submitted*.

Matzke, N.J. (2013b) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. CRAN: The Comprehensive R Archive Network, Berkeley, CA.

Matzke, N.J. (2013c) cladoRcpp: C++ implementations of phylogenetic calculations. CRAN: The Comprehensive R Archive Network, Berkeley, CA.

Matzke, N.J. (2013d) Formal Model Testing of the Dispersal-Extinction-Cladogenesis (DEC) Model Reveals that Founder-event Speciation is a Dominant Process Structuring the Biogeography of Island Clades. *Systematic Biology,* **submitted**.

Matzke, N.J. (2013e) Model selection reveals differences in cladogenesis processes operating in island versus continental clades. *Journal of Biogeography,* **submitted**.

Matzke, N.J. (2013f) TreeRogue: R code for digitizing trees.

Matzke, N.J. & Maguire, K.C. (2011) Effect of fossil range and extinction data in DEC analyses: a case study with Canidae. *Journal of Vertebrate Paleontology,* **31,**

Program and Abstracts, Society of Vertebrate Paleontology Annual Meeting, November 2011, Las Vegas.

Matzke, N.J. & Sidje, R.B. (2013) rexpokit: R wrappers for EXPOKIT; other matrix functions. CRAN: The Comprehensive R Archive Network, Berkeley, CA.

Mayr, E. (1954) Change of genetic environment and evolution. *Evolution as a process* (eds J. Huxley, A.C. Hardy & E.B. Ford), pp. 157-180. Allen & Unwin, London.

Mickevich, M.F. & Mitter, C. (1981) Treating polymorphic characters in systematics: A phylogenetic treatment of electrophoretic data. *Advances in cladistics: Proceedings of the first meeting*

of the Willi Hennig Society (eds V.A. Funk & D.R. Brooks), pp. 45-58. New York Botanical Garden Press, New York.

Moler, C. & Van Loan, C. (2003) Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review,* **45,** 3-49.

Mooers, A.Ø. (2004) Effects of Tree Shape on the Accuracy of Maximum Likelihood–Based Ancestor Reconstructions. *Systematic Biology,* **53,** 809-814.

Mooers, A.Ø. & Schluter, D. (1999) Reconstructing Ancestor States with Maximum Likelihood: Support for One-and Two-Rate Models. *Systematic Biology,* **48,** 623-633.

Moore, B.R. & Donoghue, M.J. (2007) Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. *American Naturalist,* **170 Suppl 2,** S28-55.

Moore, B.R., Smith, S.A., Ree, R.H. & Donoghue, M.J. (2006) Incorporating Fossil Data in Biogeographic Inference: A Likelihood Approach.

Moore, J.R. (2012) Do terrestrial vertebrate fossil assemblages show consistent taphonomic patterns? *PALAIOS,* **27,** 220-234.

Moore, J.R., Norman, D.B. & Upchurch, P. (2007) Assessing relative abundances in fossil assemblages. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **253,** 317-322.

Morrone, J.J. (2006) Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. *Annu. Rev. Entomol.,* **51,** 467-494.

Nash, J.C. & Varadhan, R. (2011) Unifying Optimization Algorithms to Aid Software System Users: optimx for R. *Journal of Statistical Software,* **43,** 1-14.

Nash, J.C. & Varadhan, R. (2012) optimx: A Replacement and Extension of the optim() Function. CRAN: The Comprehensive R Archive Network.

Nasrallah, C.A., Mathews, D.H. & Huelsenbeck, J.P. (2011) Quantifying the Impact of Dependent Evolution among Sites in Phylogenetic Inference. *Systematic Biology,* **60,** 60-73.

Nee, S. & May, R.M. (1997) Extinction and the Loss of Evolutionary History. *Science,* **278,** 692-694.

Nelson, G.J. & Rosen, D.E. (1981) Vicariance biogeography : a critique : symposium of the Systematics Discussion Group of the American Museum of Natural History, May 2-4, 1979. pp. xvi, 593 p. Columbia University Press, New York.

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. *Systematic Biology,* **52,** 820-838.

Nesbitt, S.J., Smith, N.D., Irmis, R.B., Turner, A.H., Downs, A. & Norell, M.A. (2009) A Complete Skeleton of a Late Triassic Saurischian and the Early Evolution of Dinosaurs. *Science,* **326,** 1530-1533.

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.

Nielsen, R. (2002) Mapping Mutations on Phylogenies. *Systematic Biology,* **51,** 729-739.

Nürk, N.M. (2011) Phylogenetic analyses in St. John's wort (Hypericum): inferring character evolution and historical biogeography.

Nylander, J.A., Olsson, U., Alstrom, P. & Sanmartin, I. (2008) Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: Turdus). *Syst Biol,* **57,** 257-268.

O'Leary, M.A., Bloch, J.I., Flynn, J.J., Gaudin, T.J., Giallombardo, A., Giannini, N.P., Goldberg, S.L., Kraatz, B.P., Luo, Z.-X., Meng, J., Ni, X., Novacek, M.J., Perini, F.A., Randall, Z.S., Rougier, G.W., Sargis, E.J., Silcox, M.T., Simmons, N.B., Spaulding, M., Velazco, P.M., Weksler, M., Wible, J.R. & Cirranello, A.L. (2013) The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. *Science,* **339,** 662-667.

Page, R.D.M. (1988) Quantitative Cladistic Biogeography: Constructing and Comparing Area Cladograms. *Systematic Zoology,* **37,** 254-270.

Pagel, M. (1999a) Inferring the historical patterns of biological evolution. *Nature,* **401,** 877-884.

Pagel, M. (1999b) The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Systematic Biology,* **48,** 612-622.

Paradis, E. (2012) *Analysis of phylogenetics and evolution with R,* 2nd edn. Springer, New York.

Parenti, L.R. & Ebach, M.C. (2009) *Comparative biogeography : discovering and classifying biogeographical patterns of a dynamic Earth*. University of California Press, Berkeley.

Parham, J.F., Donoghue, P.C.J., Bell, C.J., Calway, T.D., Head, J.J., Holroyd, P.A., Inoue, J.G., Irmis, R.B., Joyce, W.G., Ksepka, D.T., Patané, J.S.L., Smith, N.D., Tarver, J.E., van Tuinen, M., Yang, Z., Angielczyk, K.D., Greenwood, J.M., Hipsley, C.A., Jacobs, L., Makovicky, P.J., Müller, J., Smith, K.T., Theodor, J.M., Warnock, R.C.M. & Benton, M.J. (2012) Best Practices for Justifying Fossil Calibrations. *Systematic Biology,* **61,** 346-359.

Paul, C.R.C. (1998) Adequacy, Completeness and the Fossil Record. *The adequacy of the fossil record* (eds S.K. Donovan & C.R.C. Paul), pp. 1-22. John Wiley, Chichester ; New York.

Paulay, G. & Meyer, C. (2002) Diversification in the Tropical Pacific: Comparisons Between Marine and Terrestrial Systems and the Importance of Founder Speciation. *Integrative and Comparative Biology,* **42,** 922-934.

Peterson, A.T. & Lieberman, B. (2012) Species' Geographic Distributions Through Time: Playing Catch-up with Changing Climates. *Evolution: Education and Outreach,* **5,** 569-581.

Posada, D. & Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics,* **14,** 817-818.

Posada, D. & Crandall, K.A. (2001) Selecting the Best-Fit Model of Nucleotide Substitution. *Systematic Biology,* **50,** 580-601.

Pyron, R.A. (2011) Divergence Time Estimation Using Fossils as Terminal Taxa and the Origins of Lissamphibia. *Systematic Biology,* **60,** 466-481.

R Core Team (2013) R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria.

R Hackathon et al. (2012) phylobase: Base package for phylogenetic structures and comparative data. CRAN: The Comprehensive R Archive Network.

Raff, R.A. (2007) Written in stone: fossils, genes and evo-devo. *Nat Rev Genet,* **8,** 911-920.

Ramirez, S.R., Roubik, D.W., Skov, C. & Pierce, N.E. (2010) Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biological Journal of the Linnean Society,* **100,** 552-572.

Ree, R.H. (2005) Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution,* **59,** 257-265.

Ree, R.H. (2013) Lagrange configurator version 20130526 (beta). Rick Ree,
http://www.reelab.net.

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.

Ree, R.H. & Sanmartín, I. (2009) Prospects and challenges for parametric models in
historical biogeographical inference. *Journal of Biogeography,* **36,** 1211-1220.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range
evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology,*
57, 4-14.

Riddle, B.R. (1996) The molecular phylogeographic bridge between deep and
shallow history in continental biotas. *Trends in Ecology & Evolution,* **11,** 207-
211.

Robinson, D.M., Jones, D.T., Kishino, H., Goldman, N. & Thorne, J.L. (2003) Protein
Evolution with Dependence Among Codons Due to Tertiary Structure.
Molecular Biology and Evolution, **20,** 1692-1704.

Ronquist, F. (1996) *DIVA version 1.1. Computer program and manual available by
anonymous FTP from Uppsala University (ftp.uu.se or ftp.systbot.uu.se)*.

Ronquist, F. (1997) Dispersal-Vicariance Analysis: A New Approach to the
Quantification of Historical Biogeography. *Systematic Biology,* **46,** 195-203.

Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference
under mixed models. *Bioinformatics,* **19,** 1572-1574.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L. & Rasnitsyn, A.P. (2012) A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. *Systematic Biology,* **61,** 973-999.

Ronquist, F. & Sanmartin, I. (2011) Phylogenetic Methods in Biogeography. *Annual Review of Ecology, Evolution, and Systematics,* **42,** 441-464.

Rosen, B.R. (1988) From fossils to earth history: applied historical biogeography. *Analytical biogeography : an integrated approach to the study of animal and plant distributions* (eds A.A. Myers & P.S. Giller), pp. 437-481. Chapman and Hall, London ; New York.

Roy, K., Jablonski, D. & Valentine, J.W. (2004) Beyond Species Richness: Biogeographical Patterns and Biodiversity Using Other Metrics of Diversity. *Frontiers of biogeography : new directions in the geography of nature* (eds M.V. Lomolino & L.R. Heaney), pp. 151-170. Sinauer Associates, Sunderland, Mass.

Royle, J.A., Nichols, J.D. & Kéry, M. (2005) Modelling occurrence and abundance of species when detection is imperfect. *Oikos,* **110,** 353-359.

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.

Sanmartin, I. (2007) Event-based biogeography: Integrating patterns, processes, and time. *Biogeography in a changing world* (eds M.C. Ebach & R.S. Tangney), pp. 135-159. CRC Press, Boca Raton, FL.

Sanmartín, I. (2012) Historical Biogeography: Evolution in Time and Space. *Evolution: Education and Outreach,* **5,** 555-568.

Sanmartin, I., Van der Mark, P. & Ronquist, F. (2008) Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography,* **35,** 428-449.

Santos, C.M.D. (2007) On basal clades and ancestral areas. *Journal of Biogeography,* **34,** 1470-1469.

Sepkoski, D. & Ruse, M. (2009) The paleobiological revolution : essays on the growth of modern paleontology. pp. 568 p. University of Chicago Press, Chicago.

Sidje, R.B. (1998) Expokit: A Software Package for Computing Matrix Exponentials. *ACM Trans. Math. Softw.,* **24,** 130-156.

Sidje, R.B. & Hansen, N.R. (2012) expoRkit: Expokit in R. CRAN: The Comprehensive R Archive Network.

Slater, G.J. & Harmon, L.J. (2013) Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods in Ecology and Evolution,* **4,** 699-702.

Smith, A.B. (1994) *Systematics and the fossil record : documenting evolutionary patterns.* Blackwell Science, Oxford, OX ; Cambridge, Mass., USA.

Smith, S.A. (2010) Lagrange C++ Manual. Google.com, Google Code.

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. Google.com, Google Code.

Springer, M.S., Meredith, R.W., Janecka, J.E. & Murphy, W.J. (2011) The historical biogeography of Mammalia. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **366,** 2478-2502.

Statisticat, L. (2013a) LaplacesDemon Tutorial.

Statisticat, L. (2013b) LaplacesDemon: Complete Environment for Bayesian Inference.

Stigall, A.L. & Lieberman, B.S. (2006) Quantitative palaeobiogeography: GIS, phylogenetic biogeographical analysis, and conservation insights. *Journal of Biogeography,* **33,** 2051-2060.

Stigall Rode, A.L. & Lieberman, B.S. (2005) Paleobiogeographic patterns in the Middle and Late Devonian emphasizing Laurentia. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **222,** 272-284.

Swofford, D.L. (2007) PAUP* FAQ. Sinauer Associates, Florida State University.

Tedford, R.H., Wang, X. & Taylor, B.E. (2009) Phylogenetic Systematics of the North American Fossil Caninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History***,** 1-218.

Templeton, A.R. (1980) The theory of speciation via the founder principle. *Genetics,* **94,** 1011-1038.

Templeton, A.R. (2008) The reality and importance of founder speciation in evolution. *BioEssays,* **30,** 470-479.

Terry, R.C. (2010) The dead do not lie: using skeletal remains for rapid assessment of historical small-mammal community baselines. *Proceedings of the Royal Society B: Biological Sciences,* **277,** 1193-1201.

Turvey, S.T. & Blackburn, T.M. (2011) Determinants of species abundance in the Quaternary vertebrate fossil record. *Paleobiology,* **37,** 537-546.

Upchurch, P. & Hunn, C.A. (2002) "Time": the neglected dimension in cladistic biogeography? *Geobios,* **35, Supplement 1,** 277-286.

Valentine, J.W., Jablonski, D., Krug, A.Z. & Berke, S.K. (2013) The sampling and estimation of marine paleodiversity patterns: implications of a Pliocene model. *Paleobiology,* **39,** 1-20.

Van Dam, M. (2013) Comparative Biogeography of Dune Restricted Insects of the Desert Southwest. Ph. D., University of California, Berkeley.

Van Dam, M. & Matzke, N.J. (2013) Testing hypothesis of geographical connectivity with multiclade analyses of historical biogeography: case study from the desert Southwest. *In preparation*.

Velazco, P.M. & Patterson, B.D. (2013) Diversification of the Yellow-shouldered bats, Genus Sturnira (Chiroptera, Phyllostomidae), in the New World tropics. *Molecular Phylogenetics and Evolution,* **68,** 683-698.

Vermeij, G.J. & Herbert, G.S. (2004) Measuring Relative Abundance in Fossil and Living Assemblages. *Paleobiology,* **30,** 1-4.

Wallace, A.R. (1855) XVIII.—On the law which has regulated the introduction of new species. *The Annals and Magazine of Natural History,* **16,** 184-196.

Wang, X., Tedford, R.H. & Taylor, B.E. (1999) Phylogenetic systematics of the

  Borophaginae (Carnivora: Canidae). *Bulletin of the American Museum of*

  Natural History, 1-391.

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. *Systematic*

  Biology, **62,** 494-498.

Webb, C.O. & Ree, R.H. (2012) Historical biogeography inference in Malesia. *Biotic*

  evolution and environmental change in Southeast Asia (eds D. Gower, K.

  Johnson, J. Richardson, B. Rosen, L. Ruber & S. Williams), pp. 191-215.

  Cambridge University Press, Cambridge.

Wiley, E.O. (1981) *Phylogenetics: the Theory and Practice of Phylogenetic Systematics*.

  Wiley, New York.

Wood, H.M., Matzke, N.J., Gillespie, R.G. & Griswold, C.E. (2013) Treating Fossils as

  Terminal Taxa in Divergence Time Estimation Reveals Ancient Vicariance

  Patterns in the Palpimanoid Spiders. *Systematic Biology,* **62,** 264-284.

Yu, Y., Harris, A.J. & He, X. (2010) S-DIVA (Statistical Dispersal-Vicariance Analysis):

  A tool for inferring biogeographic histories. *Molecular Phylogenetics and*

  Evolution, **56,** 848-850.

Yu, Y., Harris, A.J. & He, X.-J. (2013) RASP (Reconstruct Ancestral State in

  Phylogenies) 2.1 beta.

Zink, R.M., Blackwell-Rago, R.C. & Ronquist, F. (2000) The shifting roles of dispersal

  and vicariance in biogeography. *Proc Biol Sci,* **267,** 497-503.

Zuckerkandl, E. & Pauling, L. (1965) Evolutionary divergence and convergence in

proteins. *Evolving genes and proteins,* **97,** 97-166.