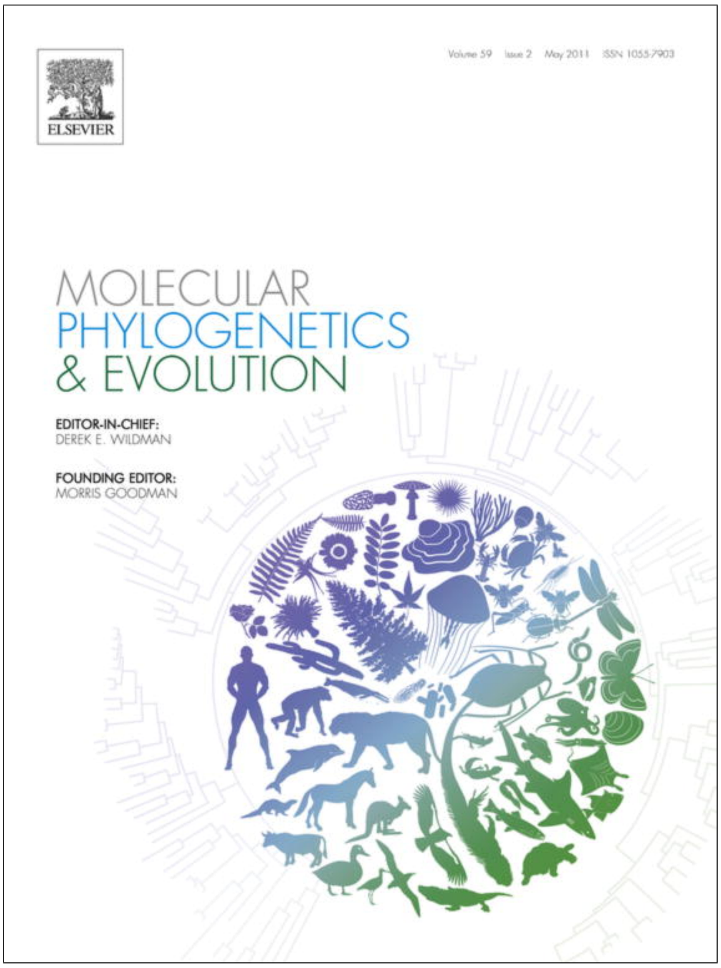


Volume 59 Issue 2 May 2011 ISSN 1055-7903
ELSEVIER
MOLECULAR
PHYLOGENETICS
& EVOLUTION
EDITOR-IN-CHIEF:
DEREK E. WILDMAN
FOUNDING EDITOR:
MORRIS GOODMAN

Molecular Phylogenetics and Evolution 59 (2011) 523–537

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Review

# Integrating statistical genetic and geospatial methods brings new power to phylogeography

Lauren M. Chan *, Jason L. Brown, Anne D. Yoder

*Department of Biology, Box 90338, Duke University, Durham, NC 27708, United States*

ARTICLE INFO

*Article history:*
Received 6 October 2010
Revised 31 January 2011
Accepted 31 January 2011
Available online 23 February 2011

*Keywords:*
Biogeography
Coalescence
Divergence
Geographic Information Systems
Spatially explicit
Statistical phylogeography
Population genetics
Gene flow

ABSTRACT

The field of phylogeography continues to grow in terms of power and accessibility. Initially uniting population genetics and phylogenetics, it now spans disciplines as diverse as geology, statistics, climatology, ecology, physiology, and bioinformatics to name a few. One major and recent integration driving the field forward is between "statistical phylogeography" and Geographic Information Systems (GIS) (Knowles, 2009). Merging genetic and geospatial data, and their associated methodological toolkits, is helping to bring explicit hypothesis testing to the field of phylogeography. Hypotheses derived from one approach can be reciprocally tested with data derived from the other field and the synthesis of these data can help place demographic events in an historical and spatial context, guide genetic sampling, and point to areas for further investigation. Here, we present three practical examples of empirical analysis that integrate statistical genetic and GIS tools to construct and test phylogeographic hypotheses. Insights into the evolutionary mechanisms underlying recent divergences can benefit from simultaneously considering diverse types of information to iteratively test and reformulate hypotheses. Our goal is to provide the reader with an introduction to the variety of available tools and their potential application to typical questions in phylogeography with the hope that integrative methods will be more broadly and commonly applied to other biological systems and data sets.

## 1. Introduction

Phylogeography continues to grow as a discipline, making rapid advances that have been fueled by new methodologies in statistical and population genetics (e.g., Buckley, 2009; Carstens and Richards, 2007; Hickerson et al., 2010; Knowles, 2009; Kozak et al., 2008; Riddle et al., 2008). Originally conceived as a means for bridging the gap between phylogenetics and population genetics, phylogeography continues to explore the processes underlying the geographic distribution of genetic diversity within and among species (Avise et al., 1987; Avise, 2000, 2009).

The field has moved considerably beyond the use of bifurcating 'species' trees as the sole source of primary data. Coalescent theory (Kingman, 1982) and the development of statistically rigorous methods for inferring historical demographic processes and testing among alternative hypotheses of population differentiation have revolutionized the field (Hickerson et al., 2010; Knowles, 2004, 2009; Nielsen and Beaumont, 2009). Methods capitalizing on known properties of the coalescent have been used to address a diversity of questions in evolutionary biology (see review in Knowles (2009)) including estimating species trees from gene trees (e.g., Carstens and Knowles, 2007; Heled and Drummond, 2010; Yang and Rannala, 2010), reconstructing changes in population size through time from ancient DNA (e.g., Chan et al., 2006; Shapiro et al., 2004; see Ramakrishnan and Hadly, 2009), and characterizing the demographic signatures associated with colonization events (e.g., Rosenblum et al., 2007). Recent reviews of coalescent-based methods underscore advances in the field and highlight some of the software programs that implement these approaches (Hickerson et al., 2010; Knowles, 2009; Kuhner, 2008; Nielsen and Beaumont, 2009; Riddle et al., 2008).

The availability of geospatial data (e.g., vegetation, climate, paleoclimate, geology) and the development of predictive modeling approaches (e.g., species distribution models, Phillips et al., 2006; mechanistic models, Buckley et al., 2010) have progressed in parallel with these innovations in population genetics and we are now on the verge of the next generation of phylogeographic analyses. An example of a geospatial technique with tremendous potential for use in phylogeographic studies is species distribution models (SDMs; also known as ecological niche models or ENMs); these have already been applied widely to evolutionarily and ecological studies. SDMs predict the distribution of a species using various climatic and geographic variables (e.g., temperature, rainfall, aspect; Phillips et al., 2006). The resulting model generates a map indicating areas of high and low habitat suitability based on

---

* Corresponding author.
*E-mail address:* lauren.chan@duke.edu (L.M. Chan).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.01.020

a species' ecological tolerance. SDMs have been used in conjunction with genetic methods to estimate ancestral distributions, the ecological interchangeability/divergence of sister taxa (and subsequently, the identification and delineation of cryptic species) and as proxies for a species dispersal potential (Graham et al., 2004; Knowles et al., 2007; Rissler et al., 2006; Stockman and Bond, 2007). While researchers must be mindful of the assumptions underlying SDMs (e.g., niche conservatism, habitat saturation, Gleasonian biotic communities), the associated uncertainties (see Elith and Leathwick, 2009; Pearson et al., 2006; Wiens et al., 2009), and the strengths and weakness of particular methodologies (e.g., Elith and Graham, 2009; Hernandez et al., 2006), these advances create considerable opportunity for merging genetic and geospatial data for the purpose of constructing and testing among temporally and spatially explicit phylogeographic hypotheses. Geographic Information Systems (GIS) provide a variety of integrative approaches that have proven useful for illuminating phylogeographic patterns and processes (see Kidd and Ritchie, 2006; Kozak et al., 2008; Richards et al., 2007). Recent empirical examples have illustrated the power of merging these data (e.g., Buckley et al., 2009; Carnaval et al., 2009; Carstens et al., 2005; Graham et al., 2004; Hugall et al., 2002; Knowles and Alvarado-Serrano, 2010; Knowles et al., 2007; Rodríguez-Robles et al., 2010; Shepard and Burbrink, 2009), but in general, relatively few phylogeographic studies have explicitly incorporated geospatial information.

Without rigorously incorporating the "geographic" component of phylogeography, there is a tendency to rely on anecdotal biogeographic inferences or simplistic classifications of biogeographic barriers. This can undervalue the influence of geography and climate on organismal distribution, and oversimplify the varying impacts geographic barriers may have (Crawford et al., 2007). Moreover, genetic patterns analyzed without consideration of spatial complexity can underestimate the effects of environmental history on organismal dispersal through time (Kozak et al., 2008).

Despite GIS technology becoming more broadly available and user friendly within the last decade, it remains underutilized in the field of phylogeography. This stems, in part, from the fact that only a handful of programs were created explicitly for phylogeographic studies. However, in reality, the abundance of currently available geospatial tools offers a rich resource for incorporating GIS into phylogeography (Table 1). More powerful and insightful phylogeographic inferences are attainable with available GIS data and tools; inventive and creative approaches to problems in phylogeography can emerge by drawing from existing methods and incorporating approaches from fields such as landscape ecology, population genetics, phylogenetics, and GIS. For example, like

**Table 1**
Summary of some spatially explicit methods of use to biogeographers.

| Software (method) | Summary | Genetic data required | Spatial data required | Key references |
|---|---|---|---|---|
| *Data exploration/pattern visualization* | | | | |
| Mesquite Cartographer, GeoPhyloBuilder, GenGIS | Projects a phylogeny onto a 2D or 3D map | Phylogenetic Tree | GPS data | Maddison and Maddison (2008), Kidd and Liu (2008), Parks et al. (2009) |
| *Matrix corrections* | | | | |
| GenAlex, Alleles in Space, R-packages such as ade4 (Mantel Test/Partial Mantel Test) | Regression analysis of pairwise matrices. Often used to test for significant isolation-by-distance | Genetic distance matrix | Geographic distance matrix | Peakall and Smouse (2006), Miller (2005), Dray and Dufour (2007), Thorpe (1996) |
| *Delineating biogeographic units/identifying barriers* | | | | |
| Barrier (Monmonier's algorithm) | Elucidates geographic positions of biogeographical boundaries | Genetic distance | Species composition data at each localities | Manni et al. (2004) |
| Alleles in Space | Implements multiple methods including Monmonier's algorithm, landscape shape interpolation, and spatial autocorrelation. | Single locus genetic data | GPS data | Miller (2005) |
| TESS, Geneland | Spatially explicit assignment method that estimates the number of populations (can be used to identify barriers) | Multilocus genetic data (microsatellites, SNPs, AFLP, sequence data) | GPS data | Chen et al. (2007), François and Durand (2010), Guillot et al. (2005) |
| Landscape Genetics GIS Toolbox | Calculates genetic landscapes by interpolating and summarizing measures of genetic diversity and divergence for multiple species | Genetic distance | GPS data | Vandergast et al. (2010) |
| *Source-sink dynamics* | | | | |
| RAMAS GIS | Estimates source-sink dynamics of populations | Population information | Landscape data, GPS data | Akçakaya (2002) |
| *Dispersal route analyses* | | | | |
| Path Matrix (Least-cost paths) | Computes LCP distances among samples given habitat heterogeneity | Species, subspecies, or populations | GPS data, relevant friction layer | Singleton et al. (2002), Ray (2005) |
| GrassGIS, CorridorDesigner, ESRI ArcMAP "corridor" tool (corridor analysis) | Computes corridors between localities given habitat heterogeneity | Species, subspecies, or populations | GPS data, relevant friction layer | GRASS Development Team (2010), Majka et al. (2007), ArcMap v9.3 ESRI 2010, |
| Circuitscape | Predicts geographic connectivity based on habitat heterogeneity and circuit theory | User defined taxonomic groups | GPS data | McRae and Beier (2007) |
| SPLATCHE 2 | Incorporates the influence of environment in the simulation of migration of a given species | Subspecies or populations, Phylogenetic data | Ancestral distribution and relevant friction layer | also see Ray et al. (2010), Knowles and Alvarado-Serrano (2010) |
| *Ancestral distribution* | | | | |
| Phylomapper (beta release) | Uses a maximum likelihood framework to estimate ancestral distributions | Phylogenetic Tree | GPS data | Lemmon and Lemmon (2008) |
| DIVA (dispersal-vicariance analysis) | Uses parsimony to estimate ancestral distributions | Phylogenetic Tree | User specified geographic units; does not explicitly integrate geography | Ronquist (1996) |

phylogeography, landscape genetics is primarily concerned with spatial patterns of genetic diversity with respect to habitat features but at smaller temporal and spatial scales. Within this field there is a rich set of methodologies for examining the correspondence between contemporary patterns of diversity and divergence among georeferenced genetic data, and quantitative information about landscape features (see Gaggiotti, 2010 and associated papers; Manel et al., 2003). Commonly used concepts and approaches in landscape genetics integrating GIS data with spatially explicit measures of genetic diversity, differentiation, and effective population size can be applied to biogeographic questions. This can help to identify barriers to gene flow and better understand how particular characteristics constrain or facilitate population connectivity in an evolutionary context.

Integrative approaches will ultimately allow us to more thoroughly consider and examine the range of potential histories underlying divergence patterns within and among species. Hypotheses generated under the exploration of one type of data are testable by the other, and jointly considering both types of information will aid in the refinement of hypotheses and the recognition of potential mechanisms previously not considered (Buckley, 2009; Knowles, 2009). While geospatial data must be used with caution (see Box 1 for discussion of species distribution modeling), they are a practical and informative tool that can place inferred demographic events in an historical and spatial context, guide genetic sampling, and point to areas for further investigation. Identifying how demographic events coincide with changes in landscape and environmental histories, such as climatic variables and the distribution of suitable habitat over time, can reveal the ecological and evolutionary mechanisms that may underlie population differentiation.

The use of ecological information and historical climatic and environmental data to guide the construction of appropriate phylogenetic and demographic models has added to our understanding of the role of particular geological barriers and climatic changes in intraspecific divergence. Thus, approaching phylogeographic studies from multiple independent perspectives can help to highlight some of the potential mechanisms underlying diversification so that we more thoroughly consider relevant and testable alternative hypotheses that might not otherwise be apparent.

## 2. Practical applications

Phylogeography aims to understand how patterns of divergence within species and species complexes coincide with current and historical geologic, geographic, and landscape features. By evaluating phylogeographic hypotheses within a statistical framework that unites phylogenetic and population genetic perspectives, we can infer the processes underlying differentiation and select among alternative evolutionary histories (Knowles and Maddison, 2002). Such integrative approaches will benefit even commonly addressed problems in phylogeography.

For instance, one goal of phylogeography is to identify and characterize the processes underlying patterns of divergence, both within and among species. Identifying dispersal corridors and estimating the degree of current and historical migration in natural populations can help us understand the importance of population connectivity to divergence (Example 1). Coalescent models for estimating migration are commonly used in population genetics, however, phylogeography is only beginning to implement integrative solutions that combine historical demographic and landscape perspectives. The fields of landscape ecology and landscape genetics, in comparison, utilize a number of methods to examine spatial genetic structure, habitat connectivity and genetic divergence given local landscape features (Table 1; also see Sork and Waits, 2010 and associated papers). Although landscape genetics often focuses

**Box 1**
Best practices guide to species distribution modeling.

| | Best practices | Things to avoid |
|---|---|---|
| *Locality data:*<br>The quality, distribution and number of points are directly related to the accuracy of the model | – Use GPS data collected with confident taxonomic identification<br>– Try to collect locality data that is evenly sampled throughout the species range and avoid biases in the sampling method (e.g., sampling only from road transects)<br>– All coordinates should be in the same datum<br>– Use as many high quality locality points as possible. It is better to have only a limited number of points that satisfy the above conditions than many points of vague credence | – Poorly documented data, particular those from databases which the taxonomic ID cannot be clarified (this is particularly important for studies on cryptic taxa or species that are easily misidentified)<br>– Georeferenced data lacking the original locality or, alternatively, data that lack a measure of error distances for the coordinates. If locality is listed, check the clarity of the description. Minimize the use of data that have been carelessly georeferenced or are based on vague localities |
| *Model parameters:*<br>The program settings and spatial scale of the area being modeled are directly related to the quality of the model | - For the primary model (this only applies to methods that use pseudoabsence points): when determining the geographic area that you are going to model, use an area slightly larger than the species known range. For organisms with low vagility, include adjacent habitat for which it could conceivably dispersed into but likely does not occur within. This is very important in the selection of the pseudoabsence points<br>*Note*: most models can easily be projected into a larger area after the initial model is created | - Blindly using the default settings. It is important to understand the assumptions and consequences of the parameters selected. For example in *MaxEnt*, the regularization multiplier is very important in determining model complexity. Recent studies using independent AIC tests suggest the default settings can often lead to under parameterized models (Warren, 2010, for overview see Elith et al. (2011)) |
| *Environmental layers:*<br>The model is based on the values from these layers, hence the highest quality data at relevant resolutions should be used | - Include variables that are likely to be directly relevant to the species being modeled<br>- For species with limited distributions or that have specific microhabitat preferences it is important to use the highest resolution layers available | - Adding all available climate data without regard to the redundancy of the data. Many environmental variables are tightly correlated making some redundant. Though usually not a problem in the creation of species distribution models, using all the climate variables might violate the assumptions of many statistical tests and could lead to spurious results |
| *Evaluating the models*: | - If not included in your model; consider the effects of fire history, glaciations, contagious diseases, anthropogenic factors, recent geological change, the species' movement potential through the landscape or biotic interactions can have on the present distribution of your species | |

at smaller geographic scales and does not incorporate a deep evolutionary history, drawing from this field can provide a unique perspective for phylogeography, and more importantly, a spatially and ecologically relevant framework for future research and analyses. Genetic interpolation (Rodríguez-Robles et al., 2010; Vandergast et al., 2008, 2010) and habitat resistance models coupled with least cost path (LCP) calculations (Spear et al., 2005; Vignieri, 2005) are common approaches to understanding genetic connectivity given landscape features and new analytical approaches in landscape genetics are rapidly being developed (Gaggiotti, 2010). When applied in an historical framework these approaches can provide new insights into evolutionary and phylogeographic patterns and processes.

The influence of biogeographic barriers such as rivers, mountains, savannahs, and climate on the distribution of species and populations is a time-honored theme in phylogeography. Given that organismal biologists will always be interested in determining the ecological and geological factors that have most strongly influenced the distribution of their focal group, phylogeographic analysis often centers on a single taxon sampled from restricted geographic ranges (Example 2). General biogeographic models are commonly used to generate phylogeographic expectations, but the set of possible histories for any given population is large, and in most instances, we must choose a subset of hypotheses to test among (Beaumont et al., 2010). Diverse types of geospatial data provide a resource for guiding the construction of alternative hypotheses relevant to the focal taxon.

Not surprisingly, the spatio-temporal pattern of divergence in a taxonomic group might be similar to those of multiple co-distributed species, particularly if a widespread geological or climatic mechanism plays a large role in differentiation (Arbogast and Kenagy, 2001; Hickerson et al., 2006). Congruence in phylogeographic structure and divergence time is expected in such cases, even when taxa differ in ecology, demography, and life history (Example 3). Some phylogeographic barriers are rather conspicuous; the closure of the Central American isthmus (Bermingham and Martin, 1998; Collins et al., 1996) and the formation of particular archipelagos in Pacific Islands (Neall and Trewick, 2008), for example, are relatively discrete events well-delineated in space and time. In these instances, deciding which taxon pairs are expected to have diverged at the same time is unambiguous. In contrast, for some cases it may be difficult to identify the predominant features or factors that may have been associated with divergence, potentially biasing interpretation towards mere plausibility rather than testability (Knowles, 2009). Exploring GIS data with simulations (Knowles and Alvarado-Serrano, 2010) or SDM (Carnaval et al., 2009) can aid the delineation of population pairs hypothesized to have been influenced by a similar event and thus, have a shared pattern of divergence.

For this review, we have purposefully chosen tractable data sets that are typical of those generated by single-investigator studies. The integrative approaches outlined here can certainly be applied to questions on broader geographic scales and to larger, more complex datasets. However, for the purposes of this paper, we focus on ways in which integrative phylogeographic methods can be both approachable and intuitive for investigators who are in the early stages of data exploration. Each of the three examples presented here illustrates the unique perspectives that can emerge from the simultaneous consideration of genetic and geographic data. For each example we introduce the study system and primary question, outline the integrative approach we take, and briefly discuss the results as well as analytical and interpretive considerations and alternative approaches. Thinking broadly about how to incorporate disparate types of information in phylogeographic studies will help move towards a truly integrative and more iterative framework for understanding those factors important in population diversification (Buckley, 2009).

## 3. Example 1 – population connectivity: visualizing putative dispersal corridors

For this first example, we use concepts from landscape genetics to explore patterns of genetic connectivity among populations, developing hypotheses of directionality and strength of gene flow among populations and across the landscape. This approach is of particular interest to conservation biologists who wish to identify those regions of the landscape that are crucial for maintaining gene flow among populations of interest. We illustrate the approach with an iguanid lizard (*Oplurus cuvieri*) endemic to Madagascar with a broad distribution across island northwest region. We integrate SDMs, GIS, and haplotype networks to estimate realistic dispersal networks, focusing on the varying connectivity of habitat between sample localities of *O. cuvieri* in recent history. Previous genetic work with *O. cuvieri* identified three divergent mitochondrial clades with low genetic diversity and structure at two nuclear loci (Chan et al., submitted for publication). Based on shared mitochondrial and nuclear haplotypes across regions, gene flow may have been substantial. By using a landscape genetic perspective, we can explore how the distribution of suitable habitat may have promoted or influenced connectivity, and inversely, differentiation.

Landscape genetic approaches have most often been used to examine how current spatial features influence contemporary levels of connectivity and divergence among populations. Applying similar methods in an historical framework can lend insight into population level processes that might underlie phylogeographic divergence and how these processes might have changed over time. Habitat resistance models and approximations of the least cost path are one commonly used set of geospatial tools for examining population connectivity in a spatially explicit framework (Clark et al., 2008; Spear et al., 2005; Storfer et al., 2010; Vignieri, 2005). We combine haplotype networks and SDM to construct a geographically explicit hypothesis of population connectivity for *O. cuvieri*. As with many other problems in phylogeography, the method we outline here is just one of many ways to integrate genetic data and GIS to examine migration. We demonstrate a simplistic yet intuitive use of genetic and GIS data to generate hypotheses regarding habitat corridors and then apply a model-based coalescent approach to explore how genetic estimates of migration fit the geographic scenario.

### 3.1. Approach

#### 3.1.1. Visualizing dispersal corridors

Haplotype networks for the mitochondrial ND1 locus and for one nuclear exon (PRLR) in *O. cuvieri* were taken from Chan et al. (submitted for publication). An SDM was generated in *MaxEnt 3.3.2* using the 19 current WorldClim climate layers (Hijmans et al., 2005) and the collection localities of *O. cuvieri* included in the genetic data set (Fig. 1A–C). We converted the model to a "dispersal cost" layer by inverting the SDM (Supplementary materials). In other words, high probability of occurrence in the SDM has a low cost to dispersal through that region, whereas regions with low or no probability of occurrence have high dispersal costs. For the nuclear haplotype network (Fig. 1D), we generated a population connectivity map by summing the least-cost paths (LCPs) among all shared haplotypes from different localities in *ArcGIS* using the dispersal cost as the friction layer. Because a single LCP oversimplifies landscape processes, we chose to use categories of LCPs to better depict habitat heterogeneity

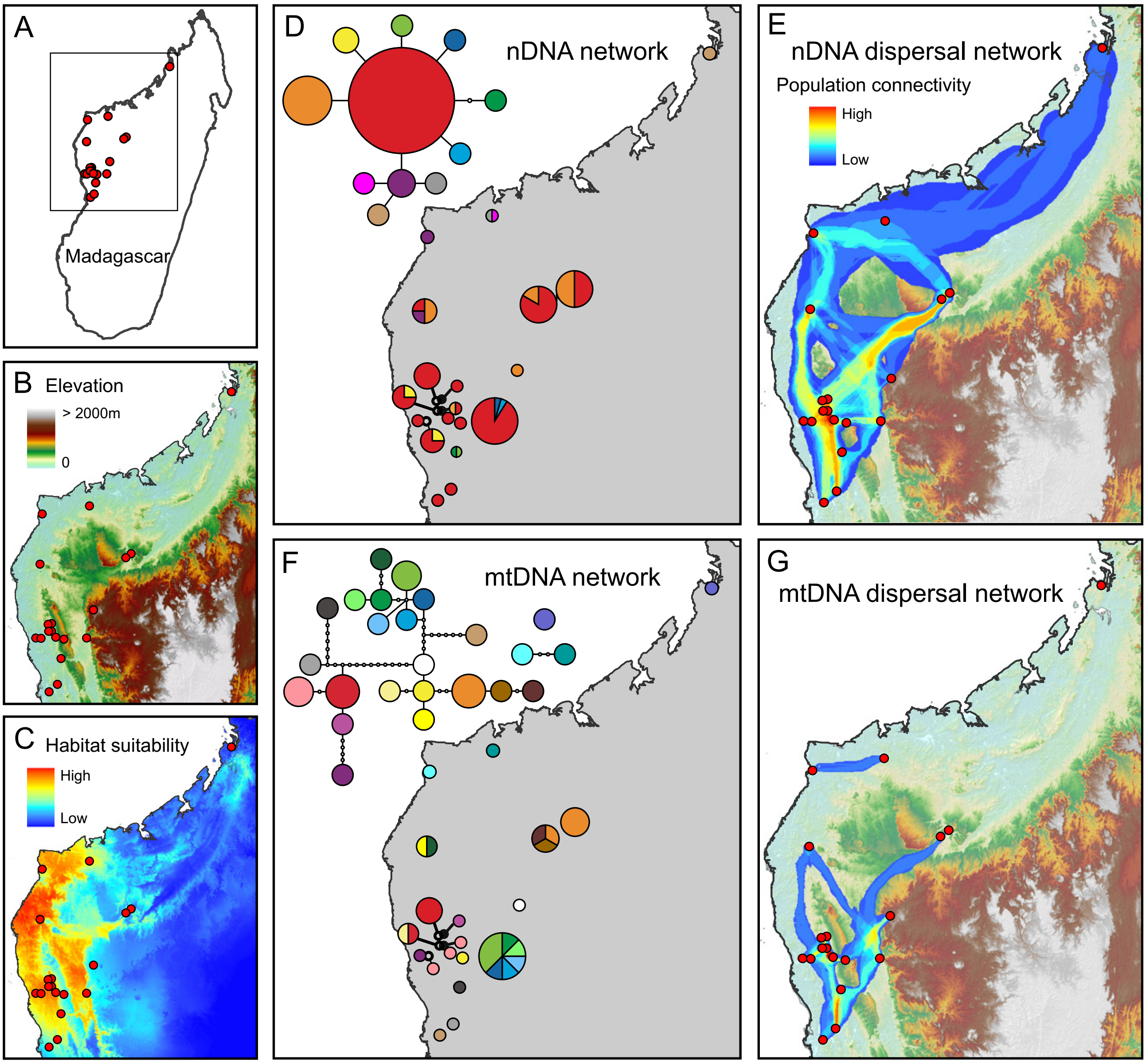


**Fig. 1.** Construction of dispersal networks for *Oplurus cuvieri*. (A) Distribution of species locality data in northwestern Madagascar. (B) Digital elevation model (DEM) of NW Madgascar. (C) "Current" species distribution model of *O. cuvieri*; warmer colors represent areas of higher habitat suitability. (D) Nuclear DNA haplotype network and georeferenced pie diagrams depicting number of samples (diameter) and haplotype frequencies (color corresponds to haplotype network) for each locality. (E) Nuclear DNA dispersal network overlaid on a DEM; warmer colors depict higher population connectivity. (F) mtDNA haplotype network (G) mtDNA dispersal network.

and its varying roles in dispersal. For each comparison we classified the LCPs into three categories: the lowest 1% LCPs, lowest 2% LCPs, lowest 5% LCPs and then subsequently summed each pairwise comparison. Areas with the hottest colors are those hypothesized to offer the greatest ease of dispersal (Fig. 1E). An analogous approach was used to calculate the population connectivity map for the mitochondrial dataset, but given the higher haplotype diversity we used sister haplotypes from unique localities rather than shared haplotypes to generate the dispersal networks (Fig. 1F and G). The power of these visual representations of habitat corridors is that they can then be used to identify populations for more intensive genetic sampling, formulate hypotheses regarding population connectivity, and identify putative barriers to dispersal. While this dispersal network is based on current climatic conditions and estimated niche space, historical connectivity maps can also be estimated to investigate migration corridors during particular time periods.

### 3.1.2. Estimating migration

In the *Oplurus* example, we explored how coalescent estimates of mitochondrial and nuclear haplotype migration correspond to putative dispersal corridors. We excluded three samples from the north due to the large geographic distance between them and the more densely sampled regions to the south. Among the latter, we categorized the samples into four regions: northwest (NW), northeast (NE), southwest (SW), and southeast (SE) populations (Fig. 2A). We estimated migration for each of six population pairs under a multilocus, coalescent-based framework implemented in (Hey and Nielsen, 2007) and conducted multiple independent runs to ensure mixing within runs and convergence among runs.

Non-zero migration was inferred for all population pairs, although estimates were asymmetrical for most pairs, with evidence of migration in one direction, but little to no migration in the opposite direction (Fig. 2A). Inferred migration rates were particularly low overall for some population pairs (e.g., NW–SW,

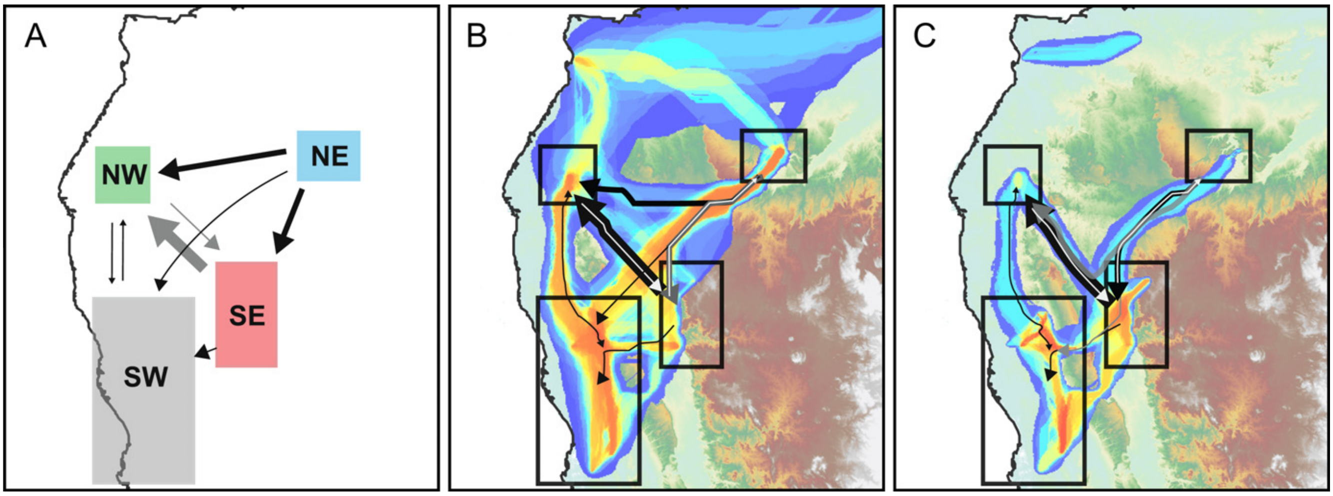


**Fig. 2.** Migration estimates for *Oplurus cuvieri*. Arrow width depicts the amount of migration between populations (note: if an arrow is absent, migration was ⩽0.001). (A) We estimated migration between each of the four populations under a multilocus, coalescent-based framework implemented in *IMa*. Migration results overlaid onto the (B) nDNA dispersal network or (C) mtDNA dispersal network using LCPs to infer the migration path between the four populations.

SE–SW, and NE–SW), but greater among the remaining population pairs.

To better illustrate the migration rates in the context of the dispersal networks, we generated the lowest LCPs among populations with inferred migration. The widths of the LCPs were weighted according to the inferred migration rate (with larger estimates corresponding to wider paths). The length of the corridors connecting populations with high migration rates differs between the nuclear (Fig. 2B) and mitochondrial data (Fig. 2C). For instance, migration from NE to NW follows a much longer corridor for the LCP based on the mitochondrial dispersal map than it does for shared nuclear map.

### 3.2. Take homes

Overlaying population migration on the LCP maps serves to highlight areas and corridors that may have served as source populations or historical refugia. Many integrative genetic and geospatial methods are emerging from landscape genetics (see Gaggiotti, 2010 and associated papers) and their application to problems in phylogeography is sure to lend insight into processes underlying evolutionary divergence. Future research could test alternate models of when migration may have occurred among neighboring populations to tease apart historical colonization events from recent gene flow. In addition, hypotheses about population expansion and contraction through time can yield insight into the intraspecific history of *O. cuvieri* populations. By exploring our data with this integrative approach, hypotheses of putative refugia, historical fluctuations in population demography, and past and present gene flow are revealed. Furthermore, these data can serve as the foundation for quantitative approaches testing the correlations between dispersal paths and the observed genetic data (see Storfer et al. (2010) for overview of statistical methods).

## 4. Example 2 – the influence of biogeographic barriers: constructing and testing among alternative hypotheses

In the second example we focus on a single species and demonstrate how two differing types of spatially explicit information can be used as a foundation for constructing alternative phylogeographic hypotheses. We use genetic data, the coalescent, species distribution models, and reconstructed ancestral distributions to examine the phylogeography of an amphibian within the Central American isthmus. The Hourglass Treefrog (*Dendropsophis ebraccatus*) occurs in the lowlands of Costa Rica and Panama surrounding the Cordillera de Talamanca (Robertson et al., 2009; Savage, 2002; Fig. 3A). Mitochondrial clades of *D. ebraccatus* coincide with putative geographic barriers that are based on the geologic history of the region, and on ecological and environmental changes associated with the formation of the Cordillera del Talamanca (Robertson et al., 2009). Weak statistical support for relationships among these genetic groups makes it difficult to infer the processes underlying divergence. The incorporation of geographically explicit data makes it possible to explore alternative perspectives, potentially providing insight into other possible mechanisms.

### 4.1. Approach

#### 4.1.1. Constructing geospatially explicit hypotheses

We use two different modes of geospatial information to guide the construction of alternative hypotheses to a strictly geological hypothesis. The first approach utilizes SDMs to identify areas of ecological stability for *D. ebraccatus* (Fig. 3). Regions that remained suitable for *D. ebraccatus* throughout the late Pleistocene, despite climatic fluctuations associated with glacial cycles, may have served as refugia. By identifying these regions, we have a means for hypothesizing how population divergence and colonization may have proceeded. We generated SDMs for *D. ebraccatus* based on current climate and paleoclimatic data obtained from the WorldClim database (Carnaval and Moritz, 2008; Hijmans et al., 2005) in *MaxEnt 3.3.2* (Phillips et al., 2006). Verified locality data (in this case, 15 collection sites from lowland regions encircling the Cordillera de Talamanca; Robertson et al., 2009; Fig. 3A) were used to generate species distribution models for four time periods - present day, mid-Holocene (6 kybp), last glacial maximum (LGM, 21 kybp), and last interglacial (LIG, ∼120–140 kybp) (Fig. 3B). These four layers were summed in *ArcGIS* and the resulting layer was reclassified to show only the top 20% of values indicating putative areas of high ecological stability throughout the Pleistocene for *D. ebraccatus*. Predicted areas of ecological stability pointed to several potential refugia with the largest being located on the northeastern side of the Cordillera de Talamanca and with a somewhat smaller refugium in the southwest portion of the distribution (Fig. 3C).

For the second approach, we estimated the ancestral distribution of the Hourglass Treefrog using the phylogenetic relationships among individuals and their corresponding localities (Fig. 4). Phylomapper (Lemmon and Lemmon, 2008) uses a maximum

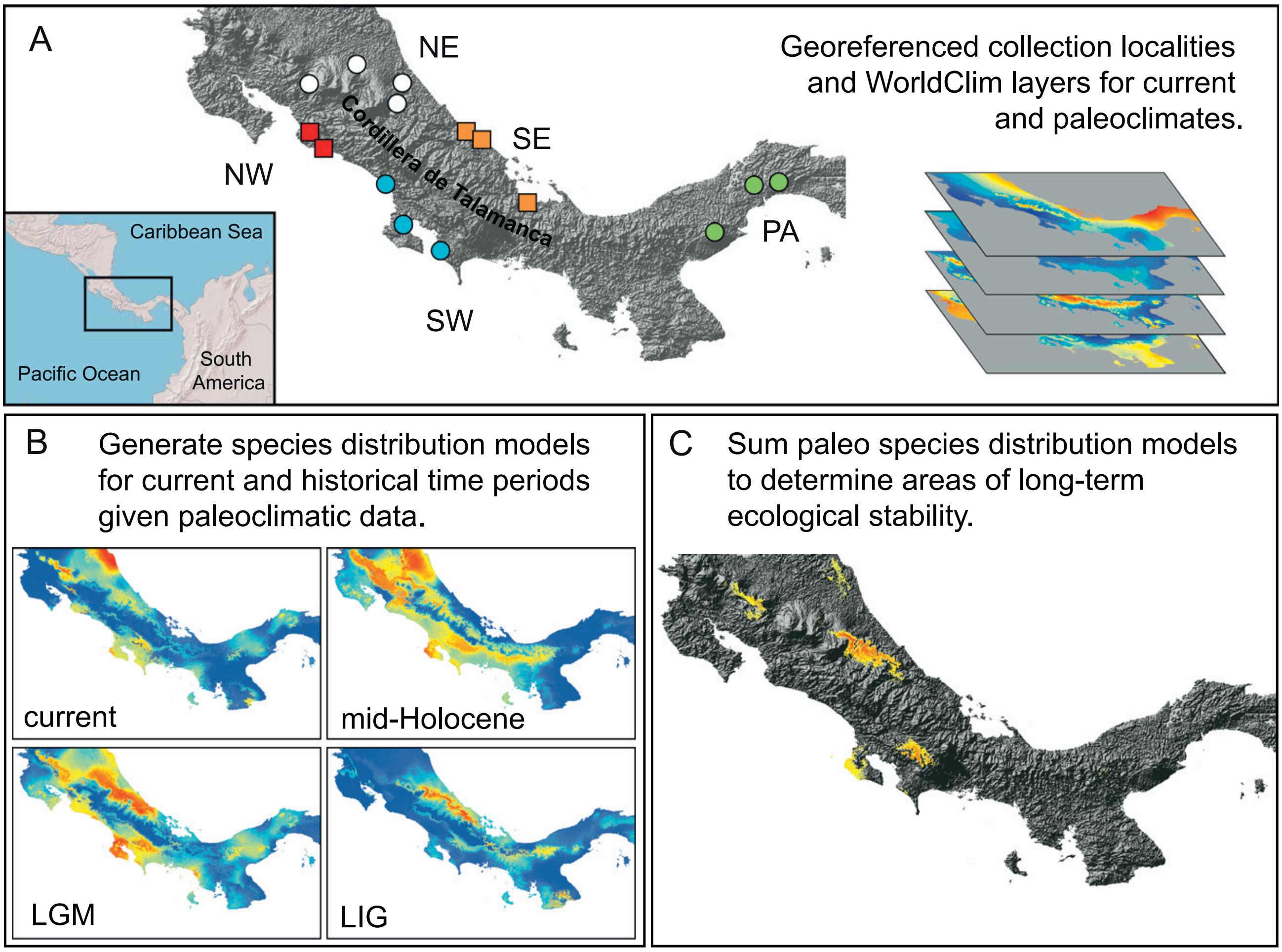


**Fig. 3.** Defining regions of ecological stability for *D. ebraccatus* (A–C) and estimating ancestral distribution (D). To define regions of ecological stability, verified point localities and climate layers from WorldClim database (A) are used in *MaxEnt* to generate species distribution models for current and historical periods (B). Paleoniche models are summed in *ArcGIS* to highlight regions of stability (C).

likelihood approach to reconstruct the localities of ancestral nodes in an ultrametric phylogeny of georeferenced haplotypes (Note: Phylomapper is in a pre-release stage of development; while it should not be used for final analyses of ancestral reconstruction, we use it here to generate novel hypotheses which we test by other means). We used non-parametric rate-smoothing to obtain an ultrametric tree of arbitrary units for the strict consensus phylogeny from a Bayesian phylogenetic analysis of 78 mitochondrial haplotypes from 15 collection localities (Robertson et al., 2009). The ancestral distribution for each node in the phylogeny was estimated in Phylomapper using default settings. To facilitate visualization, we imported the ancestral locations into *ArcMap 9.3* as a point shapefile and interpolated the localities into a continuous map using a tensioned spline based on the estimated point age (see Supplementary materials). The reconstruction of ancestral distributions based on this single estimate of the mitochondrial gene tree suggests that the oldest populations of *D. ebraccatus* were located in the central and northeastern regions with intermediate aged clades centered in the northwest and the youngest clades centered in the southern part of the isthmus (Fig. 4).

Based on the mitochondrial topology and biogeographic predictions from geological information, Robertson et al. (2009) hypothesized a southern ancestral distribution for *D. ebraccatus* with subsequent divergence, either through vicariance or migration, between Caribbean and Pacific populations on either side of the Cordillera de Talamanca (Fig. 5A). Here, we test this hypothesis in addition to three hypotheses based on the predicted regions of stability and on the ancestral distribution estimates. First, if the northeastern edge of the Cordillera de Talamanca and the southern Pacific coast served as two refugia, as suggested by current and paleo SDMs we might expect divergence across the Cordillera de Talamanca with subsequent migration along the Caribbean and Pacific coasts (Fig. 5B). Alternatively, based on the ancestral distributions of successively older clades, we can hypothesize that

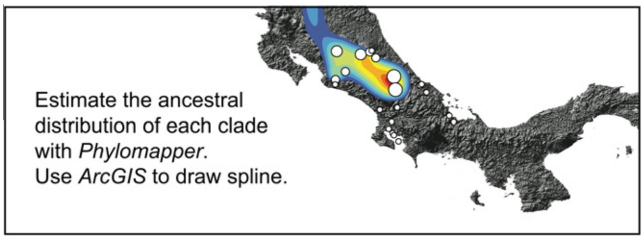


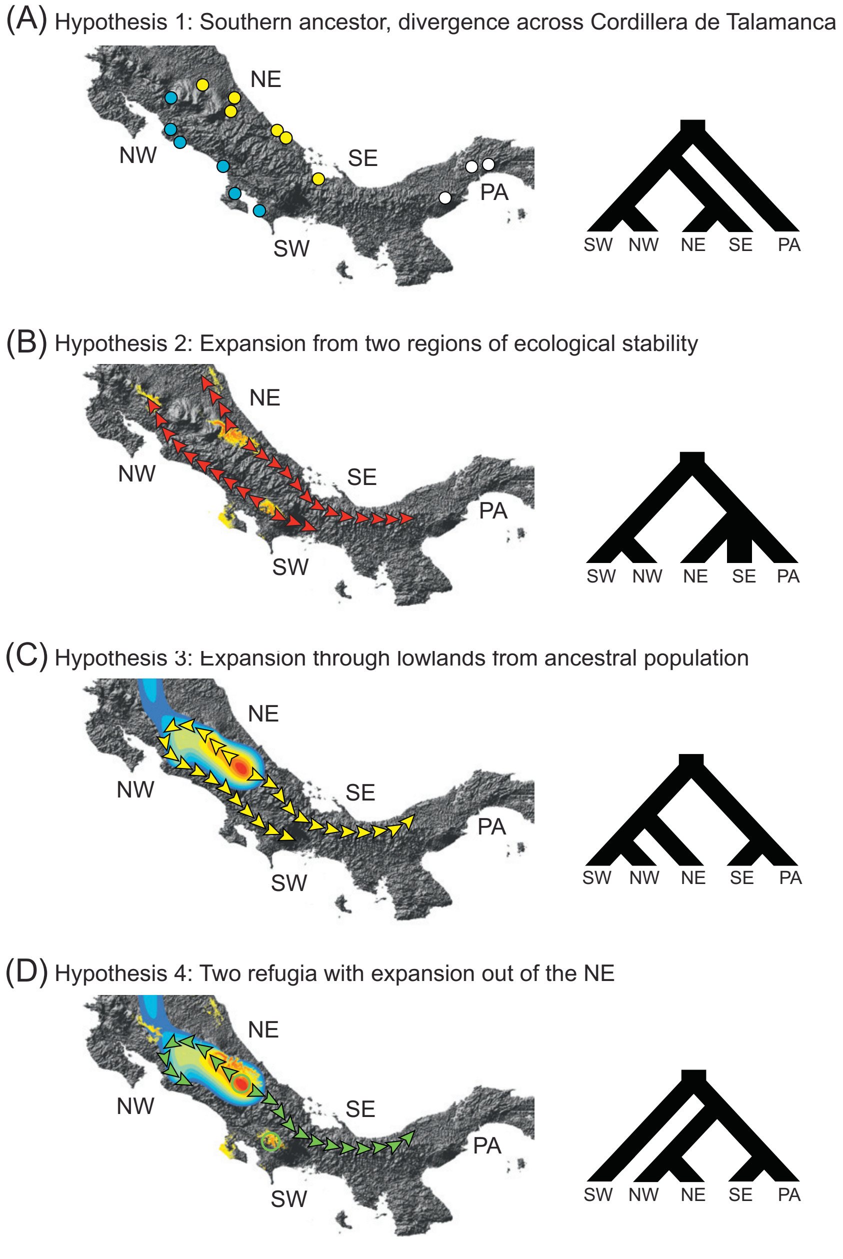
**Fig. 4.** Estimates of ancestral distribution. An ultrametric phylogeny for all samples and corresponding point localities are used in *Phylomapper* to determine the ancestral locality of nodes in the phylogeny. Circles represent the centroid of the estimated distribution at each ancestral node; the diameter of the circle depicts relative age (larger circles represent older nodes). The colored overlay represents the ancestral distribution over time (cooler colors for more recent distributions) and was generated from a tensioned spline of the estimated point ages from *Phylomapper* (see Supplementary materials).

movement occurred out of the northeast in two directions: south along the Caribbean coast and north and northwest around the tip of the Cordillera de Talamanca and south along the Pacific coast (Fig. 5C). Combining these two geospatial perspectives, we are also able to develop a third hypothesis with two refugia as in the first model, but with movement out of the northeastern refugium to the north and south along the Caribbean coast and around the northern point of the Cordillera de Talamanca (Fig. 5D).

### 4.1.2. Statistical phylogeography

In this example, we are most interested in evaluating the relative probability of these alternative population divergence models given uncertainty regarding historical population parameters such as ancestral effective population size and the timing of population divergence. Thus, we use an approximate Bayesian computation (ABC) approach to estimate the relative support for these four population divergence models. ABC is a flexible approach to population genetics and phylogeography capable of simultaneously estimating model parameters and model support (e.g., Beaumont et al., 2002; Hickerson et al., 2006; Wegmann et al., 2010).

General ABC methodology and available software has been covered recently (e.g., Lopes and Beaumont, 2009; Csilléry et al., 2010) so we do not go into detail here. There are many possibly approaches for conducting ABC analyses and the specific set of software programs used depends in part on the summary statistics desired, the genetic data included (e.g., single locus, multilocus sequence data, mixed data), the complexity of the population model (e.g., including recombination, selection, etc.), and the computer



(A) Hypothesis 1: Southern ancestor, divergence across Cordillera de Talamanca

(B) Hypothesis 2: Expansion from two regions of ecological stability

(C) Hypothesis 3: Expansion through lowlands from ancestral population

(D) Hypothesis 4: Two refugia with expansion out of the NE

**Fig. 5.** Alternative hypotheses for the phylogeographic history of *D. ebraccatus*. (A) The mitochondrial topology and geology of the isthmus suggests a southern ancestral distribution (white circles) with subsequent divergence across the Cordillera de Talamanca into Caribbean (yellow circles) and Pacific (green circles) clades. (B) Areas of putative ecological stability indicate a northeastern refugium and a southwestern refugium with migration from these centers (red arrows). (C) Estimates of the ancestral distribution of clades suggest the ancestral distribution of *D. ebraccatus* was centered in the northeast, subsequent colonization of other regions may have occurred along two paths (yellow arrows) following lower elevation habitat surrounding the Cordillera de Talamanca. (D) Combined ecological stability and ancestral distribution point to two potential refugia as in B (blue circles) with migration through low lying areas (as in C, green arrows).

platforms available to the user. For this example, we were interested in simulating data for a single locus without recombination or selection within a five population model that includes population size change, but not exponential growth or migration (Fig. 6). We also assumed that population history might be reflected at within population metrics of genetic diversity (e.g., average pairwise distance within populations) in addition to global summary statistics. Thus, our ABC pipeline used the program *BayeSSC* (Anderson et al., 2005; Excoffier et al., 2000) to simulate single locus datasets under each five population model and calculate three global summary statistics. The simulated data was then inputed to the program *arlsumstat* (Excoffier and Lischer, 2010) to calculate additional global and within population summary statistics. While many programs can conduct multi-population, single locus simulations, the format of the simulated data from *BayeSSC* can be read by *arlsumstat* without manipulation.

To parameterize coalescent simulations, sequence statistics such as base pair frequencies and the transition–transversion ratio were estimated for all *D. ebraccatus* mitochondrial sequences in *MrModeltest* (Nylander, 2004). For each of the five populations, we estimated $\theta$ within a Bayesian framework using *Lamarc* (Kuhner, 2006) and converted this to effective population size assuming a mutation rate of 2% per million generations.

For the ABC simulations, we used an R-script in *R 2.11.1* (R Development Core Team, 2010) to draw from uniform prior distributions for ancestral population size and divergence time parameters under the population model (Fig. 6). A shell script was used to tie *BayeSSC* and *arlsumstat* together for the simulation and calculation of summary statistics. Initial simulations were conducted to ensure that the parameters for the simulation yielded datasets similar to the empirical dataset. One million simulations were conducted under each general population model. Simulated parameters and associated summary statistics from all four population models were combined and we used *msReject*, part of the *msBayes* package (Hickerson et al., 2007), to perform the acceptance rejection step given a set of observed summary statistics. We used a simple validation procedure to ensure that the set of summary statistics used were good predictors of the underlying population model. The final rejection/acceptance procedures used three global summary statistics and the average and standard deviation for six within population summary statistics.

The posterior probability for two refugia model (Fig. 5B) was 0.996 indicating negligible support for any of the other three models. Given this set of alternative population divergence models, we have strong evidence for the Cordillera de Talamanca is an effective biogeographic barrier, but no support for Panama as necessarily being sister to caribbean and Pacific populations (Fig. 5A).

### 4.2. Take homes

In this example, we tested simple population models and the support we find for Hypothesis 2 is only within the context of these four models. Integrative methods in phylogeography are iterative and the ideal next step is to reconsider possible phylogeographic histories for this species given these results. For instance, a logical set of future analyses might include more complex models such as those incorporating migration and exponential growth to examine whether the demographic history of these populations are in line with expectations for expansion from isolated refugia.

Species distribution models and reconstructed ancestral distributions are just two means of exploring alternative phylogeographic scenarios. The ecology of the study system, the creativity of the investigator, and the types of spatially explicit information available will ultimately determine the range of hypotheses that may be the most insightful or revealing as well as the most appropriate means for choosing among them. That said it is important to be cognizant of the assumptions being made when constructing and testing hypotheses to avoid circularity and avoid over- and misinterpretation of the results. For instance, in this example genealogies underlie both the estimation of ancestral distribution in *Phylomapper* and hypothesis testing with ABC, but we only use the geographic locality of ancestral nodes within this phylogeny, and not the genealogy itself, to formulate a population divergence model. The test of this hypothesis using ABC does not rely on an accurate estimate of the gene genealogy, but instead uses an approach that incorporates both phylogenetic and population genetic uncertainty.

We have illustrated how ABC methods can be used to choose among general population divergence hypotheses, but geospatial data may also lead to the construction of explicit demographic predictions (e.g., population expansion or population stability), connectivity models (e.g., asymmetrical migration), or temporal hypotheses (e.g., early divergence, late divergence). In these cases, other statistical methods such as full likelihood models (e.g., Beerli and Felsenstein, 2001; Hey and Nielsen, 2007; Yang and Rannala, 2010), and Bayesian methods (e.g., Drummond et al., 2005; Drummond and Rambaut, 2007) may be more appropriate (see Nielsen and Beaumont (2009) for a recent review).

The ways in which geospatial data and statistical approaches are combined to explore, identify, and investigate biogeographic barriers within a focal taxon are numerous. We hope that the single simplified approach illustrated here is convincing evidence to show the power of using geospatially explicit data to derive alternative hypothesis testable with statistical phylogeography.

## 5. Example 3 – comparative phylogeography: detecting underlying mechanisms of diversification

In the third example we focus on comparisons of phylogeographic divergence across distantly related taxa. The field that is loosely defined as "comparative phylogeography" is of particular interest to investigators that are most interested in underlying climatic and/or geological mechanisms than in the ecological or evolutionary history of a single taxon. Here, we illustrate the approach by investigating montane animals in southwestern North America. We demonstrate the use of GIS and SDMs to guide the construction of comparative phylogeographic hypotheses in the absence of clear geospatial criteria for the delineation of populations.

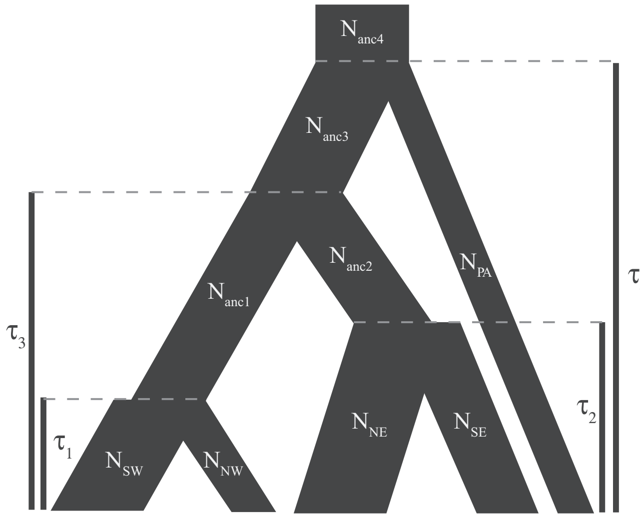


**Fig. 6.** Schematic of an example population model. Ancestral population sizes ($N_{anc}$) and divergence times ($\tau$) were sampled from uniform priors.

The Madrean Sky Islands are a series of pinyon-juniper and oak-woodland forested mountains separated from each other by lowland desert grasslands and desert scrub. Many species that occur in the high elevations forests do not occur in the warmer, more arid

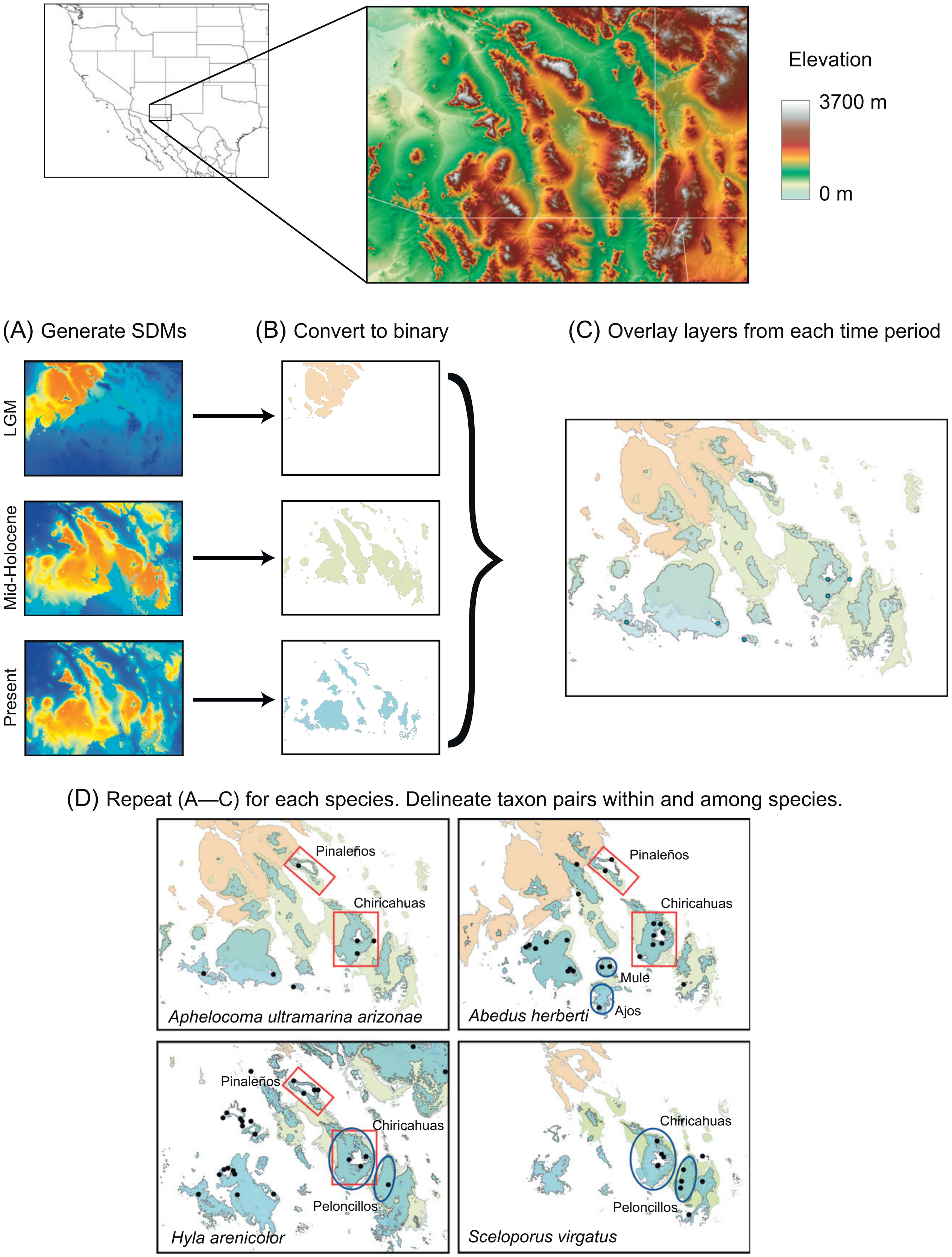


**Fig. 7.** Using current and paleoclimatic species distribution models to identify population pairs expected to have simultaneous divergence. Steps A–C are completed for each taxon individual. The resulting distributions through time are then compared (D) to delineate the taxon pairs expected to exhibit simultaneous divergence.

lowland habitats and thus, populations are currently isolated on the tops of mountains, or "sky islands." Vegetation records from packrat middens in the southwestern deserts confirm that in cooler, wetter periods throughout the Pleistocene, pinyon-juniper forest extended into lower elevations potentially connecting some montane populations as recently as 8 kybp (Van Devender, 1977, 1990a,b). Habitat and environmental changes may have restricted some species to refugia, whereas other species may have experienced only limited range contraction. If species differ substantially with respect to ecological breadth, expectations for patterns of divergence are not necessarily straightforward; historical patterns of habitat connectivity will vary depending on taxon specific ecologies. Deciding which taxa are likely to have similar histories of divergence can be guided by information about species ecology, current distribution, and estimates of historical connectivity and fragmentation.

Single taxon phylogeographic studies of the Madrean Sky Island archipelago have found recent differentiation among mountains (e.g., Barber, 1999a; Finn et al., 2007; Masta, 2000; McCormack et al., 2008; Sullivan, 1994; Tennessen and Zamudio, 2008). However, it is unclear how temporally coincident these patterns are or across species among mountain islands. Divergence times among multiple population pairs of the Mexican Jay (*Aphelocoma ultramarina arizonae*) suggests that divergence among at least some populations occurred simultaneously in line with the hypothesis that elevational shifts in habitat were a geographically widespread isolating mechanism (McCormack et al., 2008), but it is unknown whether these patterns are common across species.

### *5.1. Approach*

#### *5.1.1. Delineating population pairs*

In our analysis of simultaneous divergence, we include four taxa that are broadly separated both phylogenetically and ecologically, but which have largely overlapping distributions within the pinyon-juniper habitat of the Madrean Sky Islands. Phylogeographic structure at mitochondrial loci for populations within this island archipelago has been previously reported for the taxa we chose to include: the Mexican Jay (*Aphelocoma ultramarina arizonae*; McCormack et al., 2008), the Striped Plateau lizard (*Sceloporus virgatus*; Tennessen and Zamudio, 2008), a giant water bug (*Abedus herberti*; Finn et al., 2007) and the Canyon Treefrog (*Hyla arenicolor*; Barber, 1999a).

We obtained locality information for each species from the original phylogeographic studies, from museum databases, and the Global Biodiversity Inventory Foundation (GBIF). For each species we used the current distribution records to generate present-day distribution models as well as paleodistribution models for the LGM and mid-Holocene with *MaxEnt 3.3.2* (Fig. 7A). Based on the current SDM and the occurrence points, we decided between an occurrence probability threshold of 50% or 75% for each species, choosing the value where the model contained the majority of the genetic sampling points. Using *ArcGIS* we calculated presence/absence layers at the corresponding threshold for all three SDM layers (Fig. 7B). The layers for each species were overlaid with one another to pinpoint regions of putative range contraction during the last 21,000 yr (Fig. 7C and D). Exploring the overlap of distribution maps across species allows us to identify which population pairs we expect to diverge at the same time; population pairs from different species that diverge across the same region are particularly clear using this approach (Fig. 7D). Additionally, widespread climatic changes are hypothesized to have had similar affects across the region; if this is true, we expect to find congruence in divergence times across regions as found by McCormack et al. (2008) for Mexican Jays.

The Chiricahua Mountains and Peloncillo Mountains show similar patterns of connectivity for two species between the mid-Holocene and the present SDMs. Similarly, three sampled species for the Pinaleños and Chiricahua Mountains show less distinct but also congruent patterns of connectivity that has been disrupted since the mid-Holocene (Fig. 7D). Thus, for the purpose of this example, we chose to test for simultaneous divergence across six taxon pairs: two population pairs were from the Chiricahua Mountains and Peloncillo Mountains (*Hyla versicolor*, and *Sceloporus virgatus*), three were from the Chiricahuas and Pinaleños (*Aphelocoma ultramarina arizonae*, *Abedus herberti*, and *Hyla arenicolor*) and one pair was from Sierra los Ajos and Mule Mountain (*Abedus herberti*). In addition to the expectation of simultaneous divergence within regions, we also expect similar divergence between different population pairs of the same species when the mechanism is geographically widespread.

#### *5.1.2. Testing for simultaneous divergence*

Although comparisons of independent estimates of divergence time have been insightful in a number of examples (e.g., Hurt et al., 2009; McCormack et al., 2008), variation in effective population size among species will contribute to variance in coalescent times (Arbogast et al., 2002; Hickerson et al., 2003). We used *msBayes* (Hickerson et al., 2007) to test for simultaneous divergence across these six taxon pairs using a hierarchical approximate Bayesian computational (HABC) approach. Like standard ABC approaches, this method is very flexible and explicitly incorporates this coalescent variance. Notably, it can be used to estimate the number of divergence times among multiple taxon pairs and to test specific hypotheses about the model of speciation or divergence (see Carnaval et al., 2009; Hickerson and Meyer, 2008). For

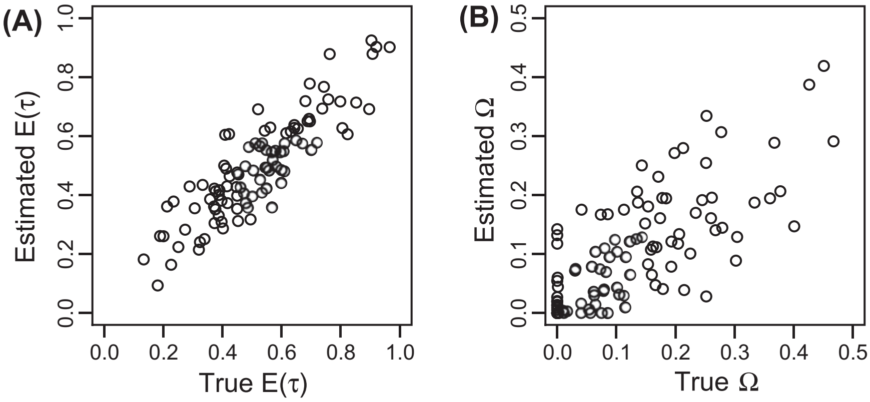


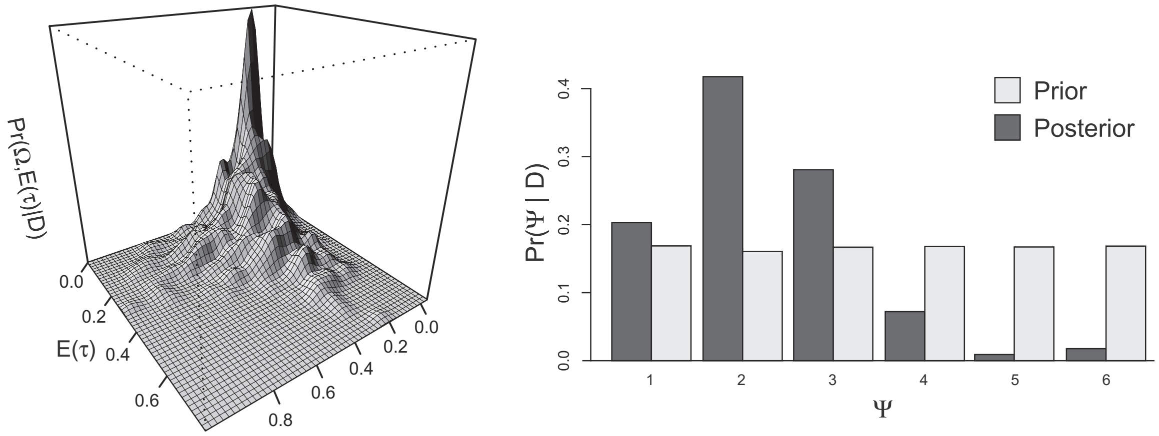
**Fig. 8.** Results from validation procedure for six taxon pair model. (A) True $E(\tau)$ versus estimated $E(\tau)$ and (B) true $\Omega$ versus estimated $\Omega$.

the sake of simplicity, we only test a simple model of vicariance without migration or population size change, though clearly, more complex models incorporating migration, colonization, and population expansion, for example, may add additional insight into this phylogeographic question (e.g., see Carnaval et al., 2009).

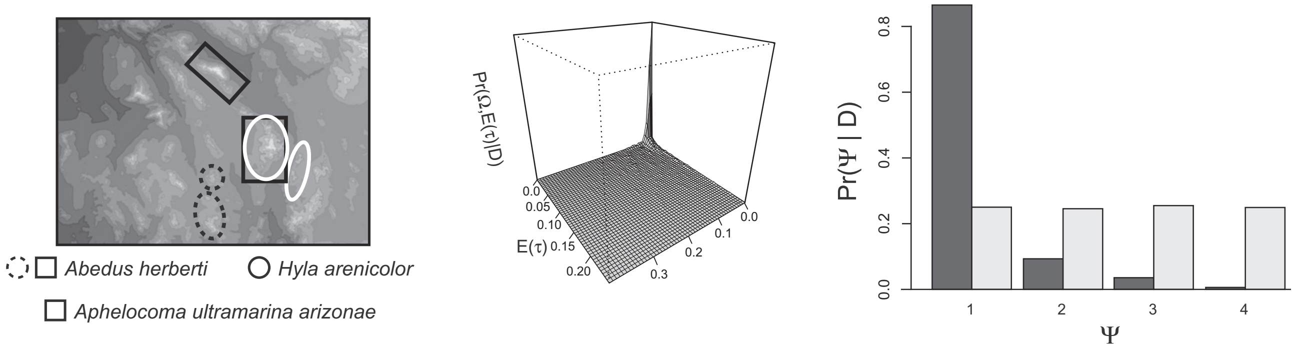
We obtained the mitochondrial datasets for these population pairs from GenBank and the authors (Barber, 1999a,b; Finn et al., 2007; McCormack et al., 2008; Tennessen and Zamudio, 2008). We used *MrModeltest* (Nylander, 2004) to estimate the transition-transversion ratio for each population pair and specified this in the parameter file for *msBayes*. Twenty million simulated datasets were generated under the specified multi-taxon model including all six population pairs and a vector of summary statistics was calculated for each. We used the *msBayes* pipeline to conduct the rejection/acceptance step and local linear regression to generate the posterior distribution for model parameters given the empirical data, including $\Psi$ (the number of divergence times in our data), $\tau$ (mean divergence time), and $\Omega$ (variance among divergence times). Model validation was performed by simulating 250 datasets under the model and examining the relationship between the "true" simulated $E(\tau)$ and $\Omega$ and the posterior estimates for these parameters (Fig. 8).

In our HABC analyses including all six population pairs, we found low support for single population (posterior probability $\Psi = 1$ is 0.2029) and highest support for two or three divergence times ($PP_{\Psi=2} = 0.4175834$ (Fig. 9A). Subsequently, we tested subsections of the data set for simultaneous divergence among four taxon pairs and found strong support for single divergence event

(A) Six taxon pairs



(B) Four taxon pairs



(C) Two taxon pairs

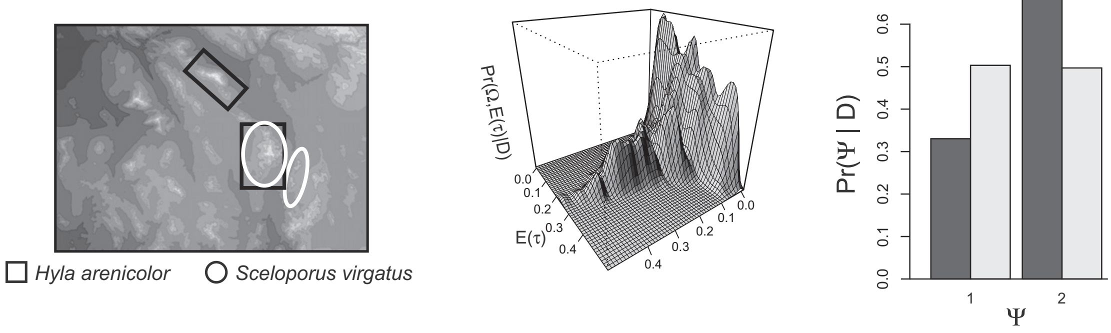


**Fig. 9.** Joint posterior probability for $E(\tau)$ (average divergence time) and $\Omega$ (variance in divergence times/average divergence time) and histogram of the prior and posterior distributions for $\Psi$ (the number of divergence times) for (A) all six taxon pairs, (B) four taxon pairs, and (C) two taxon pairs.

among the two *Abedus* taxon pairs, *Hyla* from the Peloncillos and the Chiricahua Mountains, and *Aphelocoma* from the Pinaleños and the Chiricahua Mountains (Fig. 9B). There was little support for a single divergence event between the remaining two population pairs suggesting a total of three divergence events across these taxa (Fig. 9C). We converted estimates of $\tau$ to divergence times assuming a range of average mutation rates including 1%, 2%, and 4% per million years (my). For the four taxon pairs, the mean divergence times were approximately 39.2, 19.6, and 9.8 kybp, respectively. We did not estimate separate divergence times for the remaining two taxon pairs, but the average $E(\tau)$ for these two combined was approximately 749, 375, and 187 kybp. The divergence time based the most rapid mutation rates for the four taxa is roughly congruent with a Holocene divergence coincident with hypothesized climatic changes (Van Devender, 1977, 1990a,b; Holmgren et al., 2003), but other estimates based on 1% and 2% per myr were predictably older.

### 5.2. Take homes

Surprisingly, simultaneous divergence occurred for multiple taxa from multiple regions: taxa for a single pair of mountain islands did not share a single divergence time, and likewise, not all pairs for a single taxon (in the case of *Hyla*) shared a divergence time. Despite relatively recent connectivity in the four taxon pairs (Fig. 9B), the two other taxon pairs (Fig. 9C) have much older divergence times indicating that there has not been gene flow in contrast to predictions by paleoSDMs.

We tested only a single demographic model in this example and additional models that include migration may be used to test hypotheses about colonization and "soft vicariance" (Hickerson and Meyer, 2008). As is true for single taxon studies, comparative analyses that utilize information from the nuclear and mitochondrial genomes will add resolution and statistical support to our phylogeographic hypotheses. Using geospatial information such as paleoniche models to guide our expectations for congruence across taxa will contribute to a fuller understanding of how widespread phenomena influence disparate taxa.

This example illustrates an important issue in utilizing climatic data in phylogeography: because climate layers are derived from limited and scattered current and paleoclimatic reference points, the error in derived models for certain regions may be high, but not apparent. This does not make the climate data unuseable. Rather, we must understand the limits of the data we are working with and take care not to over interpret the results given such limits. The paleoclimatic layers used in this example provide a convenient reference for delineating putative population pairs that may have separated as a result of changing climate. Given the spatial and temporal error potentially inherent in the model (e.g., they show connectivity at 6 kybp, but paleoecological records from nearby regions suggest loss of connectivity at 10–8 kybp, Van Devender, 1990a,b; Holmgren et al., 2003), it is important that we use these layers to construct hypotheses that are then explored using independent data (i.e. genetic data).

## 6. Conclusion

There are many elegant ways to integrate geospatial information into phylogeographic studies to elucidate both patterns of divergence and the associated processes. Phylogeography is iterative in nature, looking constantly among phylogenetic, population genetic, and geospatial patterns of differentiation (Buckley, 2009). Merging geospatial and genetic data is important from the early stages of data exploration all the way to complex analyses including multiple taxa, loci, and heterogeneous landscapes through time. It can help to direct future field work and sampling strategies, refine hypotheses, and can ultimately lead us to a better understanding of phylogeographic differentiation. Geospatial data can be readily integrated into most existing datasets and studies and the examples here show just a few of many possible approaches. Thinking broadly and creatively, drawing from fields such as ecology, geospatial modeling, and landscape genetics, and explicitly integrating methods within quantitative and statistical frameworks will help us better explore evolutionary histories and hence, the processes responsible for spatial patterns of genetic differentiation.

The ways in which spatially explicit data and approaches can be united with genetic data are endless. Keeping in mind the strengths and weaknesses of different types of data and various analytical methods when we integrate approaches to explore and test phylogeographic questions will strengthen our inferences and help elucidate underlying processes. As geospatial and other types of data are more routinely incorporated into phylogeographic studies, novel approaches, insightful perspectives, and new directions will emerge. Additionally, advances in other fields, such as landscape genetics and population ecology, will contribute to more integrative and collaborative research. As we write this, new resources in landscape genetics (see Gaggiotti, 2010) and novel approaches in phylogeography (e.g., Knowles and Alvarado-Serrano, 2010) are being developed that will strengthen and further the field; the potential is exceptional.

## Acknowledgments

We thank P. Barber, A. Carnaval, D. Lytle, J. McCormack, and J. Robertson for sharing data. T. Milledge and J. Pormann of the Duke Shared Cluster Resource, M. Hickerson, and D. Wegmann assisted with compiling, installing, and troubleshooting software. We thank members of the Yoder Lab, D. Swofford, D. Beamer, J. Robertson, and the instructors and participants of the 2009 Statistical Phylogeography course for fruitful discussions. This manuscript benefited from comments by members of the Yoder Lab, J. Robertson, and an anonymous reviewer.

## Appendix A. Supplementary material

Supplementary methods outlining methods used in this article can be found, in the online version, at doi:10.1016/j.ympev.2011.01.020.

## References

Akçakaya, H.R., 2002. RAMAS Metapop: viability analysis for stage-structured metapopulations (version 5.0). Applied Biomathematics. Setauket, New York.

Anderson, C.N.K., Ramakrishnan, U., Chan, Y.L., Hadly, E.A., 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. Bioinformatics 21, 1733–1734.

Arbogast, B.S., Kenagy, G.J., 2001. Comparative phylogeography as an integrative approach to historical biogeography. Journal of Biogeography 28, 819–825.

Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., Slowinski, J.B., 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annual Review of Ecology and Systematics 33, 707–740.

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA.

Avise, J.C., 2009. Phylogeography: retrospect and prospect. Journal of Biogeography 36, 3–15.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18, 489–522.

Barber, P., 1999a. Phylogeography of the canyon treefrog, *Hyla arenicolor* (Cope) based on mitochondrial DNA sequence data. Molecular Ecology 8, 547–562.

Barber, P., 1999b. Patterns of gene flow and population genetic structure in the canyon treefrog, *Hyla arenicolor* (Cope). Molecular Ecology 8, 563–576.

Beaumont, M.A., Zhang, W., Balding, D., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Beaumont, M.A., Nielsen, R., Robert, C., Hey, J., Gaggiotti, O., Knowles, L., Estoup, A., Panchal, M., Corander, J., Hickerson, M., Sisson, S.A., Fagundes, N., Chikhi, L., Beerli, P., Vitalis, R., Cornuet, J.-M., Huelsenbeck, J., Foll, M., Yang, Z., Rousset, F., Balding, D., Excoffier, L., 2010. In defence of model-based inference in phylogeography REPLY. Molecular Ecology 19, 436–446.

Beerli, P., Felsenstein, J., 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences 98, 4563–4568.

Bermingham, E., Martin, A.P., 1998. Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. Molecular Ecology 7, 499–517.

Buckley, D., 2009. Toward an organismal, integrative, and iterative phylogeography. Bioessays 31, 784–793.

Buckley, T.R., Marske, K.A., Attanayake, D., 2009. Identifying glacial refugia in a geographic parthenogen using palaeoclimate modelling and phylogeography: the New Zealand stick insect *Argosarchus horridus* (White). Molecular Ecology 18, 4650–4663.

Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J., Sears, M.W., 2010. Can mechanism inform species' distribution models? Ecology Letters 13, 1041–1054.

Carnaval, A., Moritz, C., 2008. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. Journal of Biogeography 35, 1187–1201.

Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T., Moritz, C., 2009. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. Science 323, 785–789.

Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Systematic Biology 56, 400–411.

Carstens, B.C., Richards, C., 2007. Integrating coalescent and ecological; niche modeling in comparative phylogeography. Evolution 61, 1439–1454.

Carstens, B.C., Brunsfeld, S.J., Demboski, J.R., Good, J.M., Sullivan, J., 2005. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. Evolution 59, 1639–1652.

Chan, Y.L., Anderson, C.N.K., Hadly, E.A., 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. PLoS Biology 2, 451–460.

Chan, LM, D Choi, A. Raselimanana, HA Rakotondravony, and AD Yoder. Defining spatial and temporal patterns of phylogeographic structure in Madagascar's iguanid lizards (Genus Oplurus). Submitted.

Chen, C., Durand, E., Forbes, F., François, O., 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes 7, 747–756.

Clark, R.W., Brown, W.S., Stechert, R., Zamudio, K.R., 2008. Integrating individual behaviour and landscape genetics: the population structure of timber rattlesnake hibernacula. Molecular Ecology 17, 719–730.

Collins, L.S., Budd, A.F., Coates, A.G., 1996. Earliest evolution associated with closure of the Tropical American Seaway. Proceedings of the National Academy of Science 93, 6069–6072.

Crawford, A.J., Bermingham, E., Polanía, C., 2007. The role of tropical dry forest as a long-term barrier to dispersal: a comparative phylogeographic analysis of dry forest tolerant and intolerant frogs. Molecular Ecology 16, 4789–4807.

Csilléry, K., Blum, M., Gaggiotti, O., François, O., 2010. Approximate Bayesian Computation (ABC) in practice. Trends in Ecology and Evolution 25, 410–418.

Dray, S., Dufour, A., 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22, 1–20.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular Biology and Evolution 22, 1185–1192.

Elith, J., Graham, C.H., 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. Ecography 32, 66–77.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40, 677–697.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 11, 43–57.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10, 564–567.

Excoffier, L., Novembre, J., Schneider, S., 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. Journal of Heredity 91, 506–510.

Finn, D.S., Blouin, M.S., Lytle, D.A., 2007. Population genetic structure reveals terrestrial affinities for a headwater stream insect. Freshwater Biology 52, 1881–1897.

François, O., Durand, E., 2010. Spatially explicit Bayesian clustering models in population genetics. Molecular Ecology Resources 10, 773–784.

Gaggiotti, O.E., 2010. Preface to the special issue: advances in the analysis of spatial genetic data. Molecular Ecology Resources 10, 757–759.

Graham, C., Ron, S., Santos, J., Schneider, C., Moritz, C., 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. Evolution 58, 1781–1793.

GRASS Development Team, 2010. Geographic Resources Analysis Support System (GRASS) Software, Version 6.4.0. Open Source Geospatial Foundation. http://www.grass.osgeo.org.

Guillot, G., Mortier, F., Estoup, A., 2005. Geneland: a computer package for landscape genetics. Molecular Ecology Notes 5, 712–715.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27, 570–580.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29, 773–785.

Hey, J., Nielsen, R., 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings of the National Academy of Sciences 104, 2785–2790.

Hickerson, M.J., Meyer, C.P., 2008. Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. BMC Evolutionary Biology 8, 322.

Hickerson, M.J., Gilchrist, M.A., Takebayashi, N., 2003. Calibrating a molecular clock from phylogeographic data: moments and likelihood estimators. Evolution 57, 2216–2225.

Hickerson, M.J., Stahl, E.A., Lessios, H.A., 2006. Test for simultaneous divergence using approximate Bayesian computation. Evolution 60, 2435–2453.

Hickerson, M.J., Stahl, E.A., Takebayashi, N., 2007. msBayes: pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. BMC Bioinformatics 8, 268.

Hickerson, M.J., Carstens, B.C., Cavender-Bares, J., Crandall, K.A., Graham, C.H., Johnson, J.B., Rissler, L., Victoriano, P.F., Yoder, A.D., 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution 54, 291–301.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Holmgren, C., Peñalba, M., Rylander, K., Betancourt, J., 2003. A 16,000 14C yr B.P. Packrat midden series from the USA–Mexico Borderlands. Quaternary Research 60, 319–329.

Hugall, A., Moritz, C., Moussalli, A., Stanisic, J., 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proceedings of the National Academy of Science 99, 6112–6117.

Hurt, C., Anker, A., Knowlton, N., 2009. A multilocus test of simultaneous divergence across the isthmus of Panama using snapping shrimp in the genus *Alpheus*. Evolution 63, 514–530.

Kidd, D.M., Liu, X., 2008. Geophylobuilder 1.0: an ArcGIS extension for creating 'geophylogenies'. Molecular Ecology Resources 8, 88–91.

Kidd, D.M., Ritchie, M.G., 2006. Phylogeographic information systems: putting the geography into phylogeography. Journal of Biogeography 33, 1851–1865.

Kingman, J., 1982. The coalescent. Stochastic Processes and their Applications 13, 235–248.

Knowles, L.L., 2004. The burgeoning field of statistical phylogeography. Journal of Evolutionary Biology 17, 1–10.

Knowles, L.L., 2009. Statistical phylogeography. Annual Review of Ecology and Systematics 40, 593–612.

Knowles, L.L., Alvarado-Serrano, D.F., 2010. Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. Molecular Ecology 19, 3727–3745.

Knowles, L.L., Maddison, W.P., 2002. Statistical phylogeography. Molecular Ecology 11, 2623–2635.

Knowles, L.L., Carstens, B., Keat, M., 2007. Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. Current Biology 17, 940–946.

Kozak, K.H., Graham, C.H., Wiens, J.J., 2008. Integrating GIS-based environmental data into evolutionary biology. Trends Ecology and Evolution 23, 141–148.

Kuhner, M.K., 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22, 768–770.

Kuhner, M.K., 2008. Coalescent genealogy samplers: windows into population history. Trends in Ecology and Evolution 24, 87–93.

Lemmon, A., Lemmon, E., 2008. A likelihood framework for estimating phylogeographic history on a continuous landscape. Systematic Biology 57, 544–561.

Lopes, J.S., Beaumont, M.A., 2009. ABC: a useful Bayesian tool for the analysis of population data. Infection, Genetics and Evolution 10, 826–833.

Maddison, D.R., Maddison, W.P., 2008. Cartographer, a Mesquite package for plotting geographic data. Version 1.3. <http://www.mesquiteproject.org/packages/cartographer>.

Majka, D., Jenness, J., Beier, P., 2007. CorridorDesigner: ArcGIS tools for designing and evaluating corridors. <http://www.corridordesign.org>.

Manel, S., Schwartz, M., Luikart, G., Taberlet, P., 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology and Evolution 18, 189–197.

Manni, F., Guérard, E., Heyer, E., 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". Human Biology 76, 173–190.

Masta, S., 2000. Phylogeography of the jumping spider *Habronattus pugillis* (Araneae: Salticidae): recent vicariance of sky island populations? Evolution 54, 1699–1711.

McCormack, J.E., Bowen, B.S., Smith, T.B., 2008. Integrating paleoecology and genetics of bird populations in two sky island archipelagos. BMC Biology 6, 28.

McRae, B., Beier, P., 2007. Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences 104, 19885.

Miller, M.P., 2005. Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. Journal of Heredity 96, 722–724.

Neall, V.E., Trewick, S.A., 2008. The age and origin of the Pacific islands: a geological overview. Philosophical Transactions B 363, 3293–3308.

Nielsen, R., Beaumont, M.A., 2009. Statistical inferences in phylogeography. Molecular Ecology 18, 1034–1047.

Nylander, J., 2004. MrModeltest v2. Program distributed by the author.

Parks, D.H., Porter, M., Churcher, S., Wang, S., Blouin, C., Whalley, J., Brooks, S., Beiko, R.G., 2009. GenGIS: a geospatial information system for genomic data. Genome Research 19, 1896–1904.

Peakall, R., Smouse, P., 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6, 288–295.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P., Lees, D.C., 2006. Model-based uncertainty in species range prediction. Journal of Biogeography 33, 1704–1711.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecological Modeling 190, 231–259.

R Development Core Team, 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.

Ramakrishnan, U., Hadly, E.A., 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. Molecular Ecology 18, 1310–1330.

Ray, N., 2005. Pathmatrix: a geographical information system tool to compute effective distances among samples. Molecular Ecology Resources 5, 177–180.

Ray, N., Currat, M., Foll, M., Excoffier, L., 2010. SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. Bioinformatics 26, 2993–2994.

Richards, C.L., Carstens, B.C., Knowles, L.L., 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. Journal of Biogeography 34, 1833–1845.

Riddle, B.R., Dawson, M.N., Hadly, E.A., Hafner, D.J., Hickerson, M.J., Mantooth, S.J., Yoder, A.D., 2008. The role of molecular genetics in sculpting the future of integrative biogeography. Progress in Physical Geography 32, 173–202.

Rissler, L.J., Hijmans, R.J., Graham, C.H., Moritz, C., Wake, D.B., 2006. Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. American Naturalist 167, 655–666.

Robertson, J.M., Duryea, M.C., Zamudio, K.R., 2009. Discordant patterns of evolutionary differentiation in two Neotropical treefrogs. Molecular Ecology 18, 1375–1395.

Rodríguez-Robles, J.A., Jezkova, T., Leal, M., 2010. Climatic stability and genetic divergence in the tropical insular lizard *Anolis krugi*, the Puerto Rican 'Lagartijo Jardinero de la Montaña'. Molecular Ecology 19, 1860–1876.

Ronquist, F., 1996. DIVA version 1.1. Computer program and manual available by anonymous FTP from Uppsala University. <http://ftp.uu.se> or <http://ftp.systbot.uu.se>.

Rosenblum, E.B., Hickerson, M.J., Moritz, C., 2007. A multilocus perspective on colonization accompanied by selection and gene flow. Evolution 61, 2971–2985.

Savage, J.M., 2002. The Amphibians and Reptiles of Costa Rica: A Herpetofauna between Two Continents, between Two Seas. University of Chicago Press, Chicago. 934 pp.

Shapiro, B., Drummond, A., Rambaut, A., Wilson, M., Matheus, P., Sher, A., Pybus, O., Gilbert, M., Barnes, I., Binladen, J., Willerslev, E., Hansen, A., Baryshnikov, G., Burns, J., Davydov, S., Driver, J., Froese, D., Harington, C., Keddie, G., Kosintsev, P., Kunz, M., Martin, L., Stephenson, R., Storer, J., Tedford, R., Zimov, S., Cooper, A., 2004. Rise and fall of the Beringian steppe bison. Science 306, 1561–1565.

Shepard, D.B., Burbrink, F.T., 2009. Phylogeographic and demographic effects of Pleistocene climatic fluctuations in a montane salamander, *Plethodon fourchensis*. Molecular Ecology 18, 2243–2262.

Singleton, P.H., Gaines, W., Lehmkuhl, J.F., 2002. Landscape permeability for large carnivores in Washington: a Geographic Information System weighted-distance and least-cost corridor assessment. USDA Forest Service Research Paper. PNW-RP 549. Pacific Northwest Field Station, OR.

Sork, V.L., Waits, L., 2010. Contributions of landscape genetics – approaches, insights, and future potential. Molecular Ecology 19, 3489–3495.

Spear, S.F., Peterson, C.R., Matocq, M.D., Storfer, A., 2005. Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). Molecular Ecology 14, 2553–2564.

Stockman, A.K., Bond, J.E., 2007. Delimiting cohesion species: extreme population structuring and the role of ecological interchangeability. Molecular Ecology 16, 3374–3392.

Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R., Waits, L.P., 2010. Landscape genetics: where are we now? Molecular Ecology 19, 3496–3514.

Sullivan, R.M., 1994. Micro-evolutionary differentiation and biogeographic structure among coniferous forest populations of the Mexican woodrat (*Neotoma mexicana*) in the American Southwest: a test of the vicariance hypothesis. Journal of Biogeography 21, 369–389.

Tennessen, J.A., Zamudio, K.R., 2008. Genetic differentiation among mountain island populations of the Striped Plateau Lizard, *Sceloporus virgatus* (Squamata: Phrynosomatidae). Copeia, 558–564.

Thorpe, R.S., 1996. The use of DNA divergence to help determine the correlates of evolution of morphological characters. Evolution 50, 524–531.

Van Devender, T.R., 1977. Holocene woodlands in the southwestern deserts. Science 198, 189–192.

Van Devender, T.R., 1990a. Late Quaternary vegetation and climate of the Sonoran Desert, United States and Mexico. In: Betancourt, J.L., Van Devender, T.R., Martin, P.S. (Eds.), Packrat Middens: The Last 40,000 Years of Biotic Change. University of Arizona, Tucson, pp. 134–165.

Van Devender, T.R., 1990b. Late Quaternary vegetation and climate of the Chihuahuan Desert, United States and Mexico. In: Betancourt, J.L., Van Devender, T.R., Martin, P.S. (Eds.), Packrat Middens: The Last 40,000 Years of Biotic Change. University of Arizona, Tucson, pp. 104–133.

Vandergast, A.G., Bohonak, A.J., Hathaway, S.A., Boys, J., Fisher, R.N., 2008. Are hotspots of evolutionary potential adequately protected in southern California? Biological Conservation 141, 1648–1664.

Vandergast, A.G., Perry, W.M., Lugo, R.V., Hathaway, S.A., 2010. Genetic landscapes GIS toolbox: tools to map patterns of genetic divergence and diversity. Molecular Ecology 11, 158–161.

Vignieri, S.N., 2005. Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). Molecular Ecology 14, 1925–1937.

Wegmann, D., Leuenberger, C., Neuenschwander, S., Excoffier, L., 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics 11, 116.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proceedings of the National Academy of Sciences 106, 19729–19736.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences 107, 9264–9269.