



# Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses

Corinne L. Richards\*, Bryan C. Carstens<sup>1</sup> and L. Lacey Knowles

Department of Ecology and Evolutionary Biology, 1109 Geddes Ave, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA. New address <sup>1</sup>Department of Biological Sciences, 202 Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA

\*Correspondence: Corinne L. Richards, Department of Ecology and Evolutionary Biology, 1109 Geddes Ave, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA.  
E-mail: clrichar@umich.edu

## ABSTRACT

Statistical phylogeographic studies contribute to our understanding of the factors that influence population divergence and speciation, and that ultimately generate biogeographical patterns. The use of coalescent modelling for analyses of genetic data provides a framework for statistically testing alternative hypotheses about the timing and pattern of divergence. However, the extent to which such approaches contribute to our understanding of biogeography depends on how well the alternative hypotheses chosen capture relevant aspects of species histories. New modelling techniques, which explicitly incorporate spatio-geographic data external to the gene trees themselves, provide a means for generating realistic phylogeographic hypotheses, even for taxa without a detailed fossil record. Here we illustrate how two such techniques – species distribution modelling and its historical extension, palaeodistribution modelling – in conjunction with coalescent simulations can be used to generate and test alternative hypotheses. In doing so, we highlight a few key studies that have creatively integrated both historical geographic and genetic data and argue for the wider incorporation of such explicit integrations in biogeographical studies.

## Keywords

Coalescent modelling, hypothesis testing, palaeoclimate, palaeodistribution modelling, species distribution modelling, statistical phylogeography.

## INTRODUCTION

Biogeographical research seeks to identify the processes structuring organismal diversity at a variety of geographic and taxonomic scales, from community patterns of species richness to higher-order taxonomic study. Molecular data are featured prominently in contemporary biogeographical studies because patterns of genetic variation, when interpreted in the context of geography, can provide insights into the historical demographic and biogeographical history of species (Avise *et al.*, 1987; Avise, 2000; Knowles & Maddison, 2002). However, whereas the relationship between geographic distribution and genetic variation is central to biogeography, as Kidd & Ritchie (2006) recently noted, phylogeographic research has to date placed most of its emphasis on the ‘phylo’ component, and much less on ‘geography’, despite the inherent information that the spatial-geographic component contains about the evolutionary past. These authors illustrate the potential of new GIS-based techniques to bring phylogeography back into balance, not only allowing a more powerful investigation of the geographic components of genetic varia-

tion, but also facilitating the formation of historical biogeographical hypotheses. We argue that GIS-based approaches to generating such alternative hypotheses, when coupled with genetic approaches to testing them, have the potential to increase profoundly the rigour of phylogeographic research. Herein we aim to provide readers with the necessary tools and conceptual background to take advantage of this powerful combination of distribution and coalescent-based modelling techniques in generating and testing biogeographical hypotheses. This approach has broad utility given that the required data can be readily generated for many taxa.

## Improving phylogeographic studies through hypothesis testing

In phylogeography, intraspecific genetic data are interpreted in a geographic context to infer historical and contemporary population structure and demography (Avise *et al.*, 1987; Avise, 1989, 2000). The processes generating such genetic structure will differ among species, and may include demographic events such as population bottlenecks and expansions, as well as various

types of population divergence, ranging from vicariant events to differentiation with migration (reviewed in Knowles, in press). Whereas traditional phylogeographic studies have been applied in many contexts, they have been particularly informative about the biogeographical consequences of climate change. For example, a number of studies have detected population bottlenecks coincident with the restriction of species distributions to disjunct refugia during the Earth's most recent glacial cycles (Cook *et al.*, 2001; McCracken *et al.*, 2001; Fedorov & Stenseth, 2002; Carstens *et al.*, 2004; Knowles & Richards, 2005; Steele & Storfer, 2006). Other applications of phylogeographic analyses include inferring post-glacial colonization routes (Bernatchez & Wilson, 1998; Taberlet *et al.*, 1998; Hewitt, 2000), defining species boundaries (da Silva & Patton, 1998), and assigning and assessing conservation priorities (Avice, 1992; Moritz & Faith, 1998; Richards & Knowles, 2007). Phylogeographic comparisons across codistributed taxa can also be informative about changes in the community structure of biogeographical regions over time (e.g. Schneider *et al.*, 1998; Riddle *et al.*, 2000; Sullivan *et al.*, 2000; Carstens *et al.*, 2005a; Riginos, 2005). To date, most descriptions of genetic variation and the underlying processes generating it have focused on the contemporary geographic distribution of the focal taxon (but see Hugall *et al.*, 2002).

Because biogeography and phylogeography are concerned with historical events that cannot be directly observed or experimentally replicated, our understanding of these fields is necessarily shaped by the identification of positive evidence. That is, where one of several competing historical hypotheses is identified as more probable than the others (Cleland, 2001). In this situation, tests of competing hypotheses that represent a range of possible explanations for a given phenomenon (Chamberlin, 1890) provide a framework for exploring alternative historical scenarios. Whereas phylogeographers have traditionally formulated hypotheses about the events (e.g. vicariance or migration) leading to an observed population genetic structure by comparing the shape of the genealogy with the geographic distribution of the species (e.g. Avice, 2000), this descriptive approach is prone to over-interpretation (Edwards & Beerli, 2000; Knowles & Maddison, 2002; Hudson & Turelli, 2003; Wakeley, 2003; Knowles, 2004). Because of the stochasticity of gene-lineage coalescence (Kingman, 1982; Hudson, 1992), the geographic distribution of genetic variation may not accurately reflect the population history (Pamilo & Nei, 1988; Takahata, 1989; Hudson & Coyne, 2002).

To avoid the potential problems that arise when the genealogical history of a locus is implicitly equated with the population history (i.e. interpretations concerning the biogeographical and demographic past are based on a visual inspection of a gene tree), the analysis of genetic data can proceed by means of statistical phylogeographic approaches (Knowles & Maddison, 2002), whereby the stochasticity of genetic processes is explicitly considered (Hudson, 1990; Wakeley, 2007). However, statistical phylogeographic inferences rely on explicit models of historical scenarios (e.g.

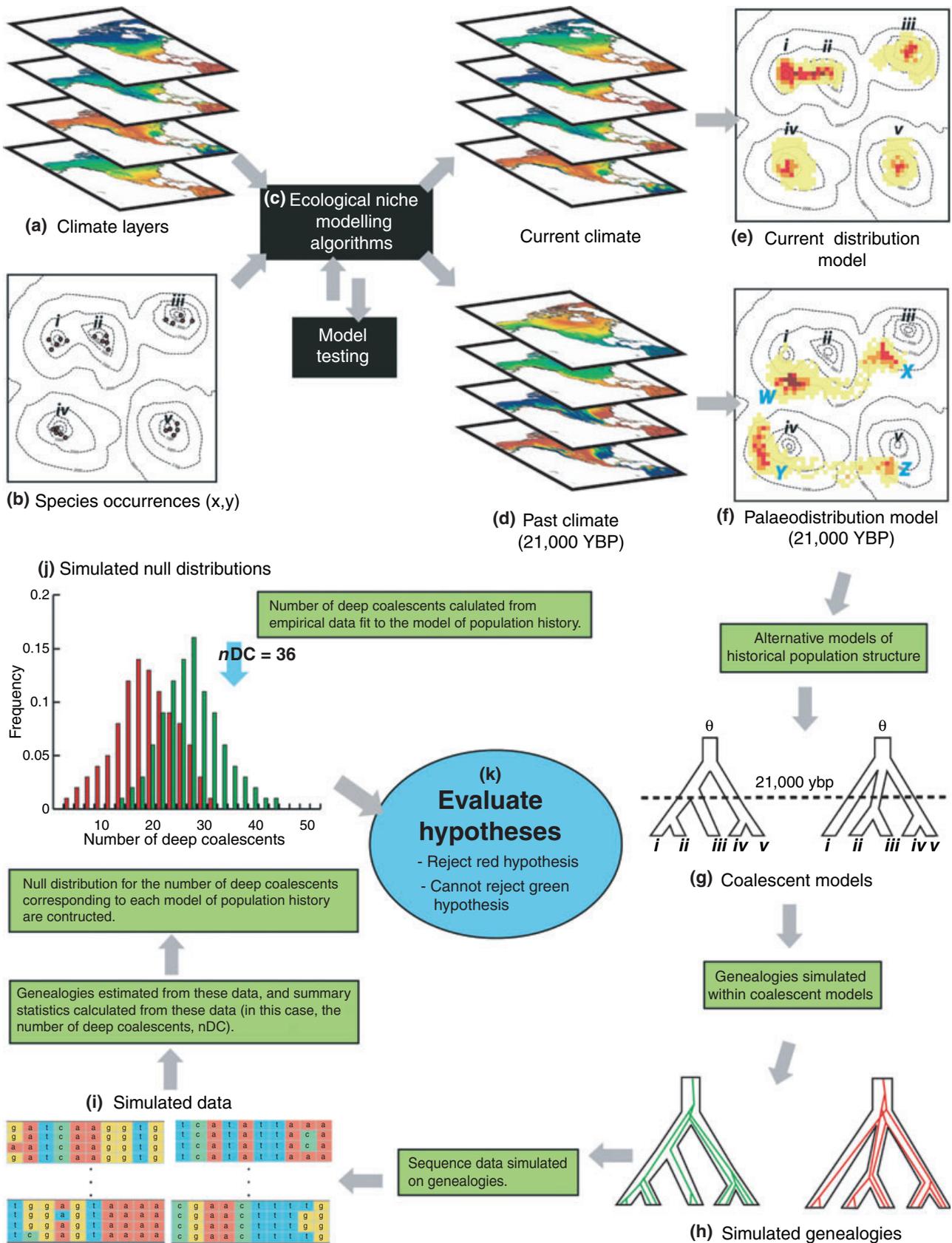
divergence with gene flow, isolation by distance, or population expansion). The choice of a model may be guided by a variety of factors. For example, decisions regarding the potential geographic configuration and temporal sequence of population divergence could be based on fossil data (e.g. Brunhoff *et al.*, 2003), packrat middens (Cognato *et al.*, 2003), palaeoenvironmental data (Tribusch & Schonswetter, 2003), or possibly be estimated from multi-locus data sets (Knowles & Carstens, 2007). However, such data are not available for all species. Herein we provide a step-by-step demonstration of how species distribution modelling techniques, coupled with palaeoclimate estimates, can provide the information necessary for generating alternative models (e.g. hypotheses about past population structure and likely corridors for migration) in cases for which no external information on past distributions has previously been available. We then walk through the steps involved in using empirical genetic data to test such hypotheses in a coalescent framework.

### A brief methodological outline

There are two major components to the coupled distribution and genetic-modelling approach: (1) generating alternative phylogeographic hypotheses for the empirical data, and (2) statistically testing these hypotheses. Each of these components involves a series of steps (see Fig. 1) and one or more modelling techniques, which will be described in detail in the following sections.

In terms of generating alternative biogeographical hypotheses (Component I below), the necessary data consist of a set of GIS layers containing information about the pertinent aspects of the current environment for the geographic area and species of interest (Fig. 1a), a set of georeferenced localities that describe where the species has been documented to occur (Fig. 1b), and, for the case of palaeodistributions, a second set of GIS layers describing an estimate of the environment at a particular time period of interest in the past (Fig. 1d). Using these inputs and any of several species distribution modelling algorithms (Fig. 1c), both the current (Fig. 1e) and past (Fig. 1f) distributions of the focal species can be estimated. These estimates of a species' past distributions, or palaeodistribution models, can then guide the generation of alternative biogeographical hypotheses (Fig. 1g).

The testing of alternative biogeographical hypotheses requires two inputs: a set of data simulated under the respective population models that represent the biogeographical hypotheses (Fig. 1h,i), and an empirical genetic data set. Each replicate of the simulated data can be characterized using a summary statistic (see Knowles, in press), generating an expectation for the pattern of genetic variation under a specific biogeographical hypothesis (Fig. 1j). The same summary statistic can then be computed for the empirical genetic data and compared with that of the simulated data for a statistical evaluation of the biogeographical hypotheses (Fig. 1k). These steps are explained in detail below in the subsection Component II.



**Figure 1** Schematic describing the process of generating alternative biogeographical hypotheses using palaeodistribution models and of testing them using coalescent simulations and empirical genetic data. In (e) and (f) the gradient from red to white differentiates areas with predicted high to low suitability, respectively, for the species in question.

## COMPONENT I: GENERATING ALTERNATIVE BIOGEOGRAPHICAL HYPOTHESES

Generating a set of alternative hypotheses about the biogeographical history of a taxon of interest should be the one of the first steps in any phylogeographic study. However, this task has historically been difficult as information about past distributions, other than what might be inferred from the empirical genetic data (e.g. Avise, 2000), is sparse to non-existent for many taxa. In this section we describe how species distribution modelling techniques can be used to generate models of species past distributions. First we provide a brief introduction to species distribution modelling, including empirical applications that illustrate how the integration of phylogeographic and species distribution modelling techniques can improve our understanding of the processes influencing contemporary patterns of biodiversity. The available algorithms and data sources for distribution modelling, as well as those relevant to generating palaeodistribution models, are then discussed along with the potential sources of error and limitations of these approaches. Finally, we describe how the resulting palaeodistribution estimate can be translated into a set of alternative biogeographical hypotheses, which can then be statistically tested using coalescent simulations.

### Applications of species distribution modelling to phylogeography

Species distribution models have been applied to a variety of research questions, including explorations of hybridization (Swenson, 2006), speciation (Losos & Glor, 2003; Graham *et al.*, 2004a), diversity gradients (Graham *et al.*, 2005, 2006; Weins *et al.*, 2006), and extinction (Martínez-Meyer *et al.*, 2004; Bond *et al.*, 2006). Because phylogeography and species distribution modelling both seek to understand biogeographical patterns and the processes generating them through studies of spatial-geographic variation, they each provide independent, but complementary, information. For this reason, studies that integrate these two sources of information are particularly powerful at detecting biogeographical patterns and inferring their causes. For example, Rissler *et al.* (2006) found concordant phylogeographic patterns among Californian reptiles and amphibians, suggesting that geographic features such as the Central Valley and the San Francisco Bay represent important barriers to dispersal. Maps of the predicted distributions of these species and lineages, generated using a species distribution modelling algorithm, were then used to identify areas of endemism and their geographic relationships to these barriers. As is the case for most phylogeographic studies, Rissler *et al.* (2006) generated hypotheses about the effects of specific geographic features on gene flow using patterns of genetic variation alone. However, their use of species distribution models, which draw upon a different set of data, supported these hypotheses from an ecological standpoint as well, revealing similar discontinuities in species distributions, and, conversely, routes of

interconnectedness. Another example illustrating how phylogeography and distribution modelling can be integrated is the study by Bond *et al.* (2006), which investigated the role that population extinction has played in defining the current distribution of *Apomastus* spiders in the Los Angeles basin. Phylogeographic data were used to detect genetic structure and signatures of population extinction, and species distribution models were used to identify regions where the spiders would probably have been found had the area's habitat not been altered by urban development.

### Distribution modelling techniques and available data sources

To generate a species distribution model, the set of conditions that offer the best prediction of the geographic distribution of a species are identified using environmental data from sites of known species occurrence (Austin, 1985; Peterson, 2001; Pearson & Dawson, 2003; Elith *et al.*, 2006). Models can be based on a variety of climatic or other environmental variables, for example measures of temperature, precipitation, elevation, ground cover, or soil type. The spatial distributions of these variables (usually captured in a set of GIS data layers, see Fig. 1a), along with a set of georeferenced sites of known species occurrence (see Fig. 1b), are then evaluated by one of several possible modelling algorithms (Fig. 1c). Each algorithm is designed to extract the relationship between environmental variation and species occurrence, although they differ in methodology and input formats (see Table 1; see also Elith *et al.*, 2006, for a recent review and comparison among techniques). This relationship is then used to predict the species' distribution given the environmental conditions of the area and time period of interest. These could be current climate measurements (Fig. 1e) or estimated climatic conditions at some time in the past (Fig. 1f) or future.

Many GIS-based environmental layers are publicly available, and an appropriate data set can often be assembled from these sources (see Table 2 for a list of data sets commonly used in distribution modelling). Species distribution data may be collected in the field or, for many taxa, gleaned from one of a number of searchable Internet data bases (see Table 3 for examples). Some data bases provide georeferenced data (i.e. X, Y coordinates corresponding to a geographic coordinate system, such as decimal degrees or UTM), but in most cases only verbal descriptions of localities are provided and georeferencing is left to the user. A set of georeferencing guidelines for the MANIS/HerpNet/ORNIS distributed natural history networks can be found at <http://manisnet.org/Georef-Guide.html>. See Graham *et al.* (2004b) for a review of the various promises and challenges of using specimen data from natural history collections for distribution modelling.

Methods for modelling species distributions differ in a number of ways, including in how they select relevant predictor variables, weight the individual variables' contributions, and predict patterns of occurrence (see Guisan & Zimmerman, 2000; Elith *et al.*, 2006). Whereas some

**Table 1** Examples of species distribution modelling algorithms available on the Internet.

Algorithm	Description	(X,Y) Input*	Software	URL	Reference
BIOCLIM	Envelope model	P	DIVA-GIS	<a href="http://www.diva-gis.org/">http://www.diva-gis.org/</a>	Nix (1986), Busby (1991)
Domain	Gower distances	P	DIVA-GIS	<a href="http://www.diva-gis.org/">http://www.diva-gis.org/</a>	Carpenter <i>et al.</i> (1993)
GARP	Genetic algorithm	P	DESKTOPGARP	<a href="http://www.nhm.ku.edu/desktopgarp/index.html">http://www.nhm.ku.edu/desktopgarp/index.html</a>	Stockwell & Peters (1999)
Generalized additive model (GAM)	Regression	PA	GRASP	<a href="http://www.unine.ch/cscf/grasp/">http://www.unine.ch/cscf/grasp/</a>	Lehmann <i>et al.</i> (2002)
Generalized linear model (GLM)	Regression	PA	GRASP	<a href="http://www.unine.ch/cscf/grasp/">http://www.unine.ch/cscf/grasp/</a>	Lehmann <i>et al.</i> (2002)
MAXENT	Maximum entropy	P	MAXENT	<a href="http://www.cs.princeton.edu/~schapire/maxent/">http://www.cs.princeton.edu/~schapire/maxent/</a>	Phillips <i>et al.</i> (2006)

\*P, presence only; PA, presence and absence.

**Table 2** Examples of commonly used environmental data sets.

Data set	Description	Source	URL
WORLDCLIM	Interpolated climate layers for global land areas	Hijmans <i>et al.</i> (2005)	<a href="http://www.worldclim.org/">http://www.worldclim.org/</a>
SRTM 90m DEMs	90-m-resolution digital elevation data for global land areas	The Consultative Group for International Agriculture Research's - Consortium for Spatial Information (CGIAR-CSI)	<a href="http://srtm.csi.cgiar.org/">srtm.csi.cgiar.org/</a>
Several available	Global current climate, environmental variables, and future climate scenarios	Intergovernmental Panel on Climate Change (IPCC)	<a href="http://www.ipcc.ch/">www.ipcc.ch/</a>
HYDRO1k	Global topographically derived data (e.g. streams, drainage basins, etc.)	United States Geological Service (USGS)*	<a href="http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html">edc.usgs.gov/products/elevation/gtopo30/hydro/index.html</a>

\*Several other useful data sets, including some with global coverage, are available from the USGS ([edc.usgs.gov/](http://edc.usgs.gov/)).

**Table 3** Examples of species distribution data bases available on the Internet.

Name	Taxon specific?	Geographic coverage	URL
Global Biodiversity Information Facility (GBIF)	No	Global	<a href="http://www.gbif.org/">www.gbif.org/</a>
World Information Network on Biodiversity (REMIB)	No	146 countries	<a href="http://www.conabio.gob.mx/remib_ingles/doctos/remib_ing.html">www.conabio.gob.mx/remib_ingles/doctos/remib_ing.html</a>
European Natural History Specimen Information Network (ENHSIN)	No	Europe	<a href="http://www.nhm.ac.uk/research-curation/projects/ENHSIN/">www.nhm.ac.uk/research-curation/projects/ENHSIN/</a>
Australian Biodiversity Information Facility (ABIF)	No	Australia	<a href="http://www.abif.org/">www.abif.org/</a>
The Biota of Canada Information Network (CBIF)	No	Canada	<a href="http://www.cbif.gc.ca/">www.cbif.gc.ca/</a>
Distributed Information for Biological Collections (SpeciesLink)	No	Brazil	<a href="http://splink.cria.org.br/index?&amp;setlang=en">splink.cria.org.br/index?&amp;setlang=en</a>
Instituto Nacional de Biodiversidad (INBio)	No	Costa Rica	<a href="http://www.inbio.ac.cr/en/default.html">www.inbio.ac.cr/en/default.html</a>
HerpNet	Yes – reptiles and amphibians	Global	<a href="http://www.herpNet.org/">www.herpNet.org/</a>
Ornithological Information System (ORNIS)	Yes – birds	Global	<a href="http://olla.berkeley.edu/ornisnet/">olla.berkeley.edu/ornisnet/</a>
Mammal Networked Information System (MANIS)	Yes – mammals	Global	<a href="http://manisnet.org/">manisnet.org/</a>
System-wide Information Network for Genetic Resources (SINGER)	Yes – crop, forage and tree species	Global	<a href="http://singer.grinfo.net/">singer.grinfo.net/</a>
Ocean Biogeographic Information System (OBIS)	Yes – marine taxa	Global	<a href="http://www.iobis.org/">www.iobis.org/</a>
Missouri Botanical Garden (Tropicos)	Yes – plants	Global	<a href="http://mobot.mobot.org/W3T/Search/vast.html">mobot.mobot.org/W3T/Search/vast.html</a>

algorithms require only records of species presence, others require both presence and absence data (see Table 1 for examples of each). Ultimately, the choice of modelling algorithm should be based on both the resulting distribution estimate's intended use and the available data (Fielding & Bell, 1997; Loiselle *et al.*, 2003; Graham *et al.*, 2004b; Elith *et al.*, 2006). However, newer algorithms, such as boosted regression trees and maximum entropy methods (e.g. MAXENT), appear to outperform several of the more established methods (e.g. GARP, BIOCLIM) in comparisons across a number of species and geographic regions (Elith *et al.*, 2006).

As with any modelling approach, the amount and type of data used can influence the accuracy of the predicted distributions. For example, generating an accurate projection of a species' distribution typically requires samples from at least 20 localities (Stockwell & Peterson, 2002; but see Pearson *et al.*, 2007). Biases in terms of where the samples are collected can affect the model's output, particularly if some areas are more accessible than others (reviewed in Graham *et al.*, 2004b), as can the choice of environmental data and modelling algorithm (Araújo & Guisan, 2006). Likewise, to the extent that recent habitat changes (e.g. ground cover) affect the presence/absence of a species, distribution models based on such rapidly changing variables run the risk of being inaccurate. For current climate layers based on multi-year averages (e.g. WorldClim: 1950–2000), however, such short-term fluctuations are less likely to unduly influence the projected distributions. Species distribution models do not take into account the potential effects of biotic exclusion, dispersal limitation, or historical contingency on species ranges. As such, it is important to recognize that these models reflect species *potential* ranges rather than their realized ranges (Araújo & Guisan, 2006). This distinction can be important for some applications, for example in conservation planning.

### Applications of palaeodistribution modelling to phylogeography

Whereas species distribution models are generally built on current environmental and species occurrence data, the inferences drawn from this approach are not limited to the present. As discussed above, distribution models can be projected onto models of the climate at some future time, for example to predict species invasions (Roura-Pascual *et al.*, 2004) or to understand how future climate change might influence species distributions (Parra-Olea *et al.*, 2005). Similarly, models of the current niche can be projected onto models of the past climate (e.g., Hugall *et al.*, 2002; Carstens & Richards, 2007; Knowles *et al.*, 2007) to reconstruct the distribution of suitable habitat at that point in the past (see Fig. 1f). For example, Hugall *et al.* (2002) used this approach to estimate the historical range of a snail in the Australian wet tropics. Comparisons between the snail's probable past distribution and its population-genetic structure, as well as the population-genetic structuring of several codistributed

vertebrates, identified a common vicariant history among the species of vertebrates. Palaeodistribution models have also been used to identify putative locations for Pleistocene refugia (Peterson *et al.*, 2004; Carstens & Richards, 2007; Knowles *et al.*, 2007), to identify historical migration pathways (Ruegg *et al.*, 2006), and to provide information about potential dispersal corridors (Carstens & Richards, 2007). In other studies, palaeodistribution models have shed light on the degree to which organismal ranges have changed over time (Lawton, 1993; Gaston, 1996).

### Generating a palaeodistribution model

Palaeodistribution models can be generated using the algorithms and data sets described above. The only additional requirement is a set of palaeoclimate estimates on which to project the species distribution (Fig. 1d; see also Cane *et al.*, 2006, for a review of recent progress in palaeoclimate modelling). Because projecting species distributions onto palaeoclimatic conditions requires the set of current and historical climate layers to be congruent, palaeodistribution studies are limited to those data for which both current measurements and palaeoclimate estimates are available. At present, we are aware of only a few publicly available palaeoclimate model outputs, and none is provided in a ready-to-use format for palaeodistribution modelling. As described below, however, these publicly available data can be re-formatted for this purpose.

The US National Oceanic & Atmospheric Administration's National Climatic Data Center (NOAA-NCDC) runs a World Data Center (WDC) for Paleoclimatology (<http://www.ncdc.noaa.gov/paleo/>) from which the outputs of several palaeoclimate models can be downloaded and viewed (<http://www.ncdc.noaa.gov/paleo/modelvis.html>). The available model runs include some from the Paleoclimate Modelling Inter-comparison Project (PMIP) as well as some from other modelling groups. These raw outputs can be downscaled and calibrated for use with a set of current climate layers in palaeodistribution modelling. An example set of palaeoclimate layers, generated using the CCM1 model (Kutzbach & Guetter, 1986; Wright *et al.*, 1993) for the last glacial maximum (21,000 yr BP), can be found in the Supplementary Material (Appendix S1), along with details of the downscaling and calibration procedure used (Appendix S2). This data set is formatted for use with the WorldClim (Hijmans *et al.*, 2005) current climate layers.

Two important caveats associated with palaeodistribution modelling are that a literal interpretation of the projected past distribution assumes that: (1) the palaeoclimate predictions are accurate, and (2) the physiological limits of species are constant (Hadly *et al.*, 1998; Davis & Shaw, 2001). Whereas recent work has demonstrated niche conservatism in several groups (Peterson & Vieglais, 2001; Martínez-Meyer *et al.*, 2004; Kozak & Wiens, 2006; Martínez-Meyer & Peterson, 2006), it is not known whether this assumption holds true for most organisms.

## From palaeodistributions to testable hypotheses

When a species distribution model is projected onto palaeoclimate estimates, the result is a GIS layer with continuous values indicating the predicted suitability of each cell for the species at one time in the past (i.e. a palaeodistribution model). Regions of core habitat (red in Fig. 1f), other less suitable areas (yellow in Fig. 1f), as well as regions that would probably have been uninhabitable (white in Fig. 1f) can be inferred from these continuous predictions, or, if desired, the predictions can be converted into binary presence–absence maps by setting minimum thresholds for species distributions (see Liu *et al.*, 2005, for a comparison among various types of thresholds and their applications).

By providing a range of predicted areas of low and high suitability (e.g. from 0 to 100), a palaeodistribution estimate facilitates the formulation of alternative models of historical population structure. For example, consider the current-day distribution prediction in Fig. 1e. The four most suitable areas are discrete and similar in size. Since not only the contemporary geographic configuration and associated demographic impacts, but also past population distributions may leave a genetic signature on patterns of genetic variation, such historical population structure needs to be taken into account. A palaeodistribution model (e.g. Fig 1f) can add this critical historical perspective, providing information about past population associations that might have contributed to patterns of genetic variation. Given the palaeodistribution model in Fig. 1f, we might hypothesize that current-day populations *i* and *ii* were descendant from a refugial population in area W, population *iii* from the refugial population X, and populations *iv* and *v* from refugial population Y, corresponding to the left-hand model of population structure in Fig. 1g. Alternatively, population *ii* could have descended from refugial population X along with population *iii*, corresponding to the right-hand model in Fig. 1g. Alternative hypotheses, such as one in which population *v* descended from refugial population Z as opposed to Y, could be envisioned and tested as well. In the next section we will describe how the alternative hypotheses generated from palaeodistribution models can be tested statistically using coalescent simulations and empirical genetic data.

## COMPONENT II: TESTING ALTERNATIVE BIOGEOGRAPHICAL HYPOTHESES

Once a set of historical biogeographical hypotheses has been identified, the next step is to evaluate statistically the extent to which the empirical genetic data support a given hypothesis. A variety of statistical phylogeographic approaches could be used to test alternative population models. Here we emphasize those approaches that employ summary statistics (reviewed in Knowles, in press), as opposed to evaluating the full probability of the observed genetic data (reviewed in Excoffier & Heckel, 2006), because of the great flexibility and ease of computation that the summary-statistic approach offers. In this section we highlight the biogeographical hypotheses that

can be addressed, with reference to some recent empirical investigations, to illustrate the synergy that results when palaeodistribution models are used to generate a predictive framework that can be tested in statistical phylogeographic studies. We then provide a step-by-step guide to the process of testing alternative hypotheses with coalescent simulations, mentioning the available software.

## Using coalescent models to test alternative hypotheses

Coalescent models have proved to be a useful tool for phylogeographic research even in the absence of explicit reconstructions of species past ranges (e.g. Milot *et al.*, 2000; Knowles, 2001; Carstens *et al.*, 2005b; DeChaine & Martin, 2005; Russell *et al.*, 2005; Steele & Storfer, 2006). For example, using a statistical evaluation of five separate potential population models, Steele & Storfer (2006) were able to show that populations of Pacific giant salamander (*Dicamptodon tenebrosus*) were isolated in separate glacial refugia during the Pleistocene glaciation. Whereas the structure of the genealogy was suggestive of this disjunction, the coalescent modelling provided details that would not otherwise have been known. These included an evaluation of the timing of divergence, which was consistent with a mid-Pleistocene divergence, thereby providing corroborative evidence for the biogeographical hypothesis of divergence among Pleistocene refugia. By using a coalescent framework, the authors could be assured that the observed geographic distribution of genetic variation reflected the population history, rather than simply the stochasticity of genetic processes. Nonetheless, palaeodistribution modelling could have added rigour to this (and other) phylogeographic studies by guiding the formation of realistic alternative hypotheses. In the case of Steele & Storfer's (2006) study, this information would ensure that the specified refugia probably contained suitable habitat for the focal species, as well as facilitating inferences about the sizes and locations of other putative refugia. Consequently, inferences about the relative contributions of past events, such as the effect of climate-induced shifts in species distributions, on population genetic structure would be not only more accurate, but also more detailed.

The potential benefits of this approach extend to comparative phylogeographic studies, in which general regional hypotheses provide a metric for comparisons among organisms with different life-history traits (Arbogast & Kenagy, 2001). For example, Carstens & Richards (2007) generated palaeodistribution models for eight codistributed lineages from the Pacific Northwest mesic forests of North America and used the fit of genetic data to the alternative models, as determined with coalescent simulations, to evaluate whether there was congruence in the location and structure of Pleistocene refugia and post-Pleistocene dispersal corridors among the taxa. Such a framework is critical for identifying whether differences in the patterns of genetic variation among species reflect varying responses to common historical events,

or, despite shared distributions today, reflect incongruence among the species past distributions.

### A step-by-step guide to testing alternative hypotheses with coalescent models

The alternative population structures suggested by palaeodistribution models can be evaluated by constructing null distributions for expected patterns of genetic variation (or a summary statistic that is used to characterize the data) from data simulated by a neutral coalescent process under a specific population model. For example, at least two testable hypotheses are suggested by the model shown in Fig. 1f. Coalescent models that correspond to these hypotheses may be conceptualized by the respective population trees, in which branch lengths reflect the timing of divergence and branch widths correspond to the effective population size (Fig. 1g). Coalescent models can be designed with varying degrees of complexity. However, an excessively complex model may have limited utility because the available genetic data for evaluating such models may not be sufficient – complex models can require large amounts of genomic data (Knowles & Maddison, 2002). Furthermore, since the use of summary statistics necessarily involves a loss of information, the ability to distinguish among various complicated models of a species history may not be possible because the expected value of the summary statistic may not differ between the models (Wakeley, 2003). The key is to identify the simplest model that captures the relevant features of the organism's history (Knowles, 2004).

Whereas the palaeodistribution models provide crucial information for erecting a coalescent model that captures the geography of divergence, as illustrated in Fig. 1g, there are demographic aspects of the population history that are also important as they too influence the pattern of genetic variation across the landscape by influencing the rate of gene-lineage loss (i.e. the amount of genetic drift). These include the timing of divergence, as well as the effective population size, which may or may not have been constant over time. Whereas the timing of divergence may be derived from the palaeoclimatic information (e.g. the last or preceding glacial maxima), other demographic parameters are estimated directly from the genetic data. For example, the effective population size,  $N_e$ , can be calculated from the population-mutational parameter  $\theta$ , which is  $4N_e\mu$ , when there is an estimate of the mutation rate  $\mu$  (e.g. the commonly used rate of divergence of 2% per million years for insect mitochondrial DNA). The parameter  $\theta$  might be estimated using a coalescent-based program (e.g. using LAMARC: Kuhner, 2006), as might an estimate of a population growth parameter in the event that a constant effective population size is not a reasonable assumption. Otherwise,  $\theta$  might be estimated directly from the distribution of segregating sites (e.g. using Watterson's estimator of  $\theta$ ) or the pairwise differences (e.g. based on nucleotide diversity  $\pi$ ) among DNA haplotypes (e.g. using DNASP: Rozas *et al.*, 2003).

Coalescent simulations are used to evaluate the fit of the empirical data to a particular historical model (Fig. 1h,i). For such tests, the data should be simulated under conditions that mirror all aspects of the empirical data, including the amount of data and mutational model underlying the observed patterns of genetic variation. For example, if a researcher sequenced 983 basepairs from a gene, which evolved under an HKY+ $\Gamma$  model of sequence evolution, from 129 individuals, the simulated data should share these characteristics. The program *ms* (Hudson, 2002), in combination with *SEQ-GEN* (Rambaut & Grassly, 1997), allows users to specify  $\theta$ , the number of basepairs, the model of sequence evolution, and the number of individuals in order to generate simulated data that provide an expectation for the pattern of genetic variation under a specific population history. *MESQUITE* (Maddison & Maddison, 2006) includes modules with similar capabilities along with several analytical tools that allow users to calculate a summary statistic (such as the number of deep coalescences or Slatkin and Maddison's  $s$ ) for each simulated data set that can then be used to construct a null distribution for the summary statistic (Fig. 1j).

This coalescent-based hypothesis-testing process involves first generating a large number of genealogies simulated by a neutral coalescent process under each model of population history (e.g. Fig. 1h). Sequence data are then simulated on these genealogies (Fig. 1i). A summary statistic is calculated for each replicate data set, and together (e.g. considering the values of the summary statistic from each of 1000 simulated data sets) they generate a null distribution for the summary statistic (Fig. 1j) (see Voight *et al.*, 2005, and Hickerson *et al.*, 2006, for examples in which multiple summary statistics are considered simultaneously). When the value of the summary statistic estimated from empirical genetic data is compared with the null distribution, it provides a statistical framework for evaluating the fit of the data to one or more models (for example, the red and green distributions in Fig. 1j reflect the expected number of deep coalescences under the respective population models, Fig. 1g). For example, the number of deep coalescences observed in the empirical data in Fig. 1 differs significantly from what would be expected had the data evolved under a model in which population  $i$  was not colonized from the same ancestral population as population  $ii$  (i.e., the population model on the right in Fig. 1g) – less than 5% of the simulated data sets exhibited a value for the number of deep coalescences that was equal to or greater than what was observed for the empirical data. However, the data are consistent with the alternative population model (i.e. the population model on the left in Fig. 1g), as the probability of observing the number of deep coalescences that was calculated for the empirical data was less than 5% (i.e.,  $P < 0.05$ ).

### BENEFITS OF THE INTEGRATIVE APPROACH

Whereas the integration of palaeodistribution and coalescent modelling techniques represents a new and informative development in biogeographical research (Stigall & Lieberman,

2006), it has yet to be widely employed. However, predictive models of the type advocated here can lead to important biogeographical insights at a variety of spatial and temporal scales. This is because the genetic data are used to test hypotheses built with explicit reference to the species under study, as opposed to relying on generic models. Moreover, the coupling of palaeodistribution and coalescent models provides a flexible framework with which to evaluate patterns of genetic variation under the diverse and varied historical conditions that have contributed to contemporary patterns of species diversity.

There are challenges associated with palaeodistribution modelling (Araújo & Guisan, 2006; Hijmans & Graham, 2006), as well as with statistical phylogeographic tests (reviewed in Knowles, in press), but these difficulties are offset by the potential benefits of improving studies of the population processes that contribute to regional patterns of biodiversity. Indeed, it is only when present and historical geo-spatial and genetic data are integrated in such a predictive, hypothesis-testing framework that the discipline of phylogeography will fulfill its promise as an integrative field capable of connecting microevolutionary processes to macroevolutionary patterns (Bermingham & Moritz, 1998).

## ACKNOWLEDGEMENTS

Training in species distribution modelling was provided to C.L.R. by the Center for Biodiversity and Conservation at the American Museum of Natural History and was funded by the University of Michigan's Rackham Graduate School. The research was funded by a National Science Foundation grant (DEB-0447224) to L.L.K.

## REFERENCES

- Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modeling. *Journal of Biogeography*, **33**, 1677–1688.
- Arbogast, B.S. & Kenagy, G.J. (2001) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, **28**, 819–825.
- Austin, M.P. (1985) Continuum concept, ordination methods and niche theory. *Annual Review of Ecology and Systematics*, **16**, 39–61.
- Avise, J.C. (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**, 1192–1208.
- Avise, J.C. (1992) Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, **63**, 62–76.
- Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.
- 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.
- Bermingham, E. & Moritz, C. (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, **7**, 367–369.
- Bernatchez, L. & Wilson, C.C. (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**, 431–452.
- Bond, J.E., Beamer, D.A., Lamb, T. & Hedin, M. (2006) Combining genetic and geospatial analyses to infer population extinction in mygalomorph spiders endemic to the Los Angeles region. *Animal Conservation*, **9**, 145–157.
- Brunhoff, C., Galbreath, K.E., Fedorov, V.B., Cook, J.A. & Jaarola, M. (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, **12**, 957–968.
- Busby, J.R. (1991) Bioclim – a bioclimatic analysis and prediction system. *Plant Protection Quarterly*, **6**, 8–9.
- Cane, M.A., Braconnot, P., Clement, A., Gildor, H., Joussame, S., Kageyama, M., Khodri, M., Paillard, D., Tett, S. & Zorita, E. (2006) Progress in paleoclimate modeling. *Journal of Climate*, **19**, 5031–5057.
- Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2**, 667–680.
- Carstens, B.C. & Richards, C.L. (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.
- Carstens, B.C., Stevenson, A.L., Degenhardt, J.D. & Sullivan, J. (2004) Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology*, **53**, 781–792.
- Carstens, B.C., Brunsfeld, S.J., Demboski, J.R., Good, J.D. & Sullivan, J. (2005a) Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution*, **59**, 1639–1652.
- Carstens, B.C., Degenhardt, J.D., Stevenson, A.L. & Sullivan, J. (2005b) Accounting for coalescent stochasticity in testing phylogeographic hypotheses: modeling Pleistocene population structure in the Idaho Giant Salamander *Dicamptodon aterrimus*. *Molecular Ecology*, **14**, 255–265.
- Chamberlin, T.C. (1890) The method of multiple working hypotheses. *Science (old series)* **15**, 92–96; reprinted 1965, **148**, 754–759.
- Cleland, C.E. (2001) Historical science, experimental science, and the scientific method. *Geology*, **29**, 987–990.
- Cognato, A.I., Harlin, A.D. & Fisher, M.L. (2003) Genetic structure among Pinyon pine beetle populations (Scolytinae: *Ips confusus*). *Environmental Entomology*, **32**, 1262–1270.
- Cook, J.A., Bidlack, A.L., Conroy, C.J., Demboski, J.R., Fleming, M.A., Runck, A.M., Stone, K.D. & MacDonald, S.O. (2001) A phylogeographic perspective on endemism in the

- Alexander Archipelago of southeast Alaska. *Biological Conservation*, **97**, 215–227.
- Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**, 673–678.
- DeChaine, E.G. & Martin, A.P. (2005) Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. *Journal of Biogeography*, **32**, 1943–1956.
- Edwards, S.V. & Beerli, P. (2000) Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Excoffier, L. & Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**, 745–758.
- Fedorov, V.B. & Stenseth, N.C. (2002) Multiple glacial refugia in the North American Arctic: inference from phylogeography of the collared lemming (*Dicrostonyx groenlandicus*). *Proceedings of the Royal Society B: Biological Sciences*, **269**, 2071–2077.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for assessment of predictive errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Gaston, K.J. (1996) Species-range-size distributions: patterns, mechanisms, and implications. *Trends in Ecology and Evolution*, **11**, 197–201.
- Graham, C.H., Ron, S., Santos, J.C., Schneider, C.J. & Moritz, C. (2004a) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781–1793.
- Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004b) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, **19**, 497–503.
- Graham, C.H., Smith, T.B. & Languy, M. (2005) Current and historical factors influencing patterns of species richness and turnover of birds in the Gulf of Guinea highlands. *Journal of Biogeography*, **32**, 1371–1384.
- Graham, C.H., Moritz, C. & Williams, S.E. (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences USA*, **103**, 632–636.
- Guisan, A. & Zimmerman, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modeling*, **135**, 147–186.
- Hadly, E.A., Kohn, M.H., Leonard, J.A. & Wayne, R.K. (1998) A genetic record of population isolation in pocket gophers during Holocene climatic change. *Proceedings of the National Academy of Sciences USA*, **95**, 6893–6896.
- Hewitt, G.M. (2000) The genetic legacy of the Quaternary ice age. *Nature*, **405**, 907–913.
- Hickerson, M.J., Dolman, G. & Moritz, C. (2006) Comparative phylogeographic summary statistics for testing simultaneous vicariance. *Molecular Ecology*, **25**, 209–223.
- Hijmans, R.J. & Graham, C.H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12**, 2272–2281.
- 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.
- Hudson, R.R. (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 1–44.
- Hudson, R.R. (1992) Gene trees, species trees and the segregation of ancestral alleles. *Genetics*, **131**, 509–512.
- Hudson, R.R. (2002) Generating samples under a Wright–Fisher neutral model. *Bioinformatics*, **18**, 337–338.
- Hudson, R.R. & Coyne, J.A. (2002) Mathematical consequences of the genealogical species concept. *Evolution*, **56**, 1557–1565.
- Hudson, R.R. & Turelli, M. (2003) Stochasticity overrules the three-times rule: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182–190.
- Hugall, A., Moritz, C., Moussalli, A. & Stanicic, J. (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Arosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences USA*, **99**, 6112–6117.
- Kidd, D.M. & Ritchie, M.G. (2006) Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography*, **33**, 1851–1865.
- Kingman, J.F.C. (1982) The coalescent. *Stochastic processes and their Applications*, **13**, 235–248.
- Knowles, L.L. (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, **10**, 691–701.
- Knowles, L.L. (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.
- Knowles, L.L. (in press) Statistical phylogeography: interpreting evolutionary history from analyses of genetic variation. *Annual Review Ecology, Evolution and Systematics*.
- Knowles, L.L. & Carstens, B.C. (2007) Estimating a geographically explicit model of population divergence. *Evolution*, **61**, 477–493.
- Knowles, L.L. & Maddison, W.P. (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.
- Knowles, L.L. & Richards, C.L. (2005) Genetic drifts importance during Pleistocene divergence as revealed by analysis of genomic variation. *Molecular Ecology*, **14**, 4023–4032.
- Knowles, L.L., Carstens, B.C. & Keat, M.L. (2007) Coupled genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, **17**, 1–7.

- Kozak, K.H. & Wiens, J.J. (2006) Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution*, **60**, 2604–2621.
- Kuhner, M.K. (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.
- Kutzbach, J.E. & Guetter, P.J. (1986) The influence of changing orbital parameters and surface boundary conditions on climate simulations for the past 18,000 years. *Journal of the Atmospheric Sciences*, **43**, 1726–1759.
- Lawton, J.H. (1993) Range, population abundance and conservation. *Trends in Ecology & Evolution*, **8**, 409–413.
- Lehmann, A., Overton, J.M.C. & Leathwick, J.R. (2002) GRASP: generalized regression analysis and spatial predictions. *Ecological Modelling*, **157**, 189–207.
- Liu, C., Berry, P.M., Dawson, P.D. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.
- Loiselle, B.A., Howell, C.A., Graham, C.H., Brooks, T. & Williams, P.H. (2003) Identifying conservation priorities: sensitivities to model selection. *Conservation Biology*, **17**, 1591–1600.
- Losos, J.B. & Glor, R.E. (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution*, **18**, 220–227.
- Maddison, D.M. & Maddison, W.P. (2006) *Mesquite: a modular system for evolutionary analysis, version 1.12*. <http://www.mesquiteproject.org>.
- Martínez-Meyer, E. & Peterson, A.T. (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-recent transition. *Journal of Biogeography*, **33**, 1779–1789 [last accessed 1-Sep-07].
- Martínez-Meyer, E., Peterson, A.T. & Hargrove, W.W. (2004) Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography*, **13**, 305–314.
- McCracken, K.G., Johnson, W.P. & Sheldon, F.H. (2001) Molecular population genetics, phylogeography, and conservation biology of the mottled duck (*Anas fulvigula*). *Conservation Genetics*, **2**, 87–192.
- Milot, M., Gibbs, H.L. & Hobson, K.A. (2000) Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). *Molecular Ecology*, **9**, 667–681.
- Moritz, C. & Faith, D.P. (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology*, **7**, 419–429.
- Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. *Atlas of elapid snakes of Australia*. Australian Flora and Fauna Series No. 7 (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.
- Pamilo, P. & Nei, M. (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.
- Parra-Olea, G., Martínez-Meyer, E. & Pérez-Ponce de León, G. (2005) Forecasting climate change effects on salamander distribution in the highlands of central Mexico. *Biotropica*, **37**, 2002–2008.
- Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimatic envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.
- Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.
- Peterson, A.T. (2001) Predicting species geographic distributions based on ecological niche modeling. *Condor*, **103**, 599–605.
- Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling. *BioScience*, **51**, 363–371.
- Peterson, A.T., Martínez-Meyer, E. & González-Salazar, C. (2004) Reconstructing the Pleistocene geography of the *Aphelocoma* jays (Corvidae). *Diversity and Distributions*, **10**, 237–246.
- Phillips, S.J., Anderson, R. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Rambaut, A. & Grassly, N.C. (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics*, **13**, 235–238.
- Richards, C.L. & Knowles, L.L. (2007) Tests of phenotypic and genetic concordance and their application to the conservation of Panamanian golden frogs (Anura, Bufonidae). *Molecular Ecology*, **16**, 3119–3133.
- Riddle, B.R., Hafner, J., Alexander, L.F. & Jaeger, J.R. (2000) Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences USA*, **97**, 14438–14443.
- Riginos, C. (2005) Cryptic vicariance in Gulf of California fishes parallels vicariant patterns found in Baja California mammals and reptiles. *Evolution*, **59**, 2678–2690.
- 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 the California herpetofauna. *The American Naturalist*, **167**, 655–666.
- Roura-Pascual, N., Suarez, A.V., Gomez, C., Pons, P., Touyama, Y., Wild, A.L. & Peterson, A.T. (2004) Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2527–2534.
- Rozas, J., Sánchez-DelBarrio, J.C., Messegyer, X. & Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.
- Ruegg, K.C., Hijmans, R.J. & Moritz, C. (2006) Climate change and the origin of migratory pathways in the Swainson's thrush, *Catharus ustulatus*. *Journal of Biogeography*, **33**, 1172–1182.

- Russell, A.L., Medellín, R.A. & McCracken, G.F. (2005) Genetic variation in the Mexican free-tail bat (*Tadarida brasiliensis mexicana*). *Molecular Ecology*, **14**, 2207–2222.
- Schneider, C.J., Cunningham, M. & Moritz, C. (1998) Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics of Australia. *Molecular Ecology*, **7**, 487–498.
- da Silva, M.N.F. & Patton, J.L. (1998) Molecular phylogeography and the evolution and conservation of Amazonian mammals. *Molecular Evolution*, **7**, 475–486.
- Steele, C.A. & Storfer, A. (2006) Coalescent-based hypothesis testing supports multiple Pleistocene refugia in the Pacific Northwest for the Pacific giant salamander (*Dicamptodon tenebrosus*). *Molecular Ecology*, **15**, 2477–2487.
- Stigall, A.L. & Lieberman, B.S. (2006) Quantitative paleobiogeography: GIS, phylogenetic biogeographical analysis, and conservation insights. *Journal of Biogeography*, **33**, 2051–2060.
- Stockwell, D.R.B. & Peters, D. P. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Systems*, **13**, 143–158.
- Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modeling*, **148**, 1–13.
- Sullivan, J., Arellano, E.A. & Rogers, D.S. (2000) Comparative phylogeography of Mesoamerican highland rodents: concerted versus independent responses to past climatic fluctuations. *The American Naturalist*, **155**, 755–768.
- Swenson, N.G. (2006) GIS-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. *Journal of Evolutionary Biology*, **19**, 717–725.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Takahata, N. (1989) Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics*, **122**, 957–966.
- Tribsch, A. & Schonswetter, P. (2003) Patterns of endemism and comparative phylogeography confirm paleoenvironmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon*, **52**, 477–497.
- Voight, B.F., Adams, A.M., Frisse, L.A., Qian, Y., Hudson, R.R. & Di Rienzo, A. (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proceedings of the National Academy of Sciences USA*, **12**, 18508–18513.
- Wakeley, J. (2003) Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. *The evolution of population biology* (ed. by R. Singh and M. Uyenoyama), pp. 193–213. Cambridge University Press, Cambridge.
- Wakeley, J. (2007) *Coalescent theory: an Introduction*. Roberts and Co., New York.
- Weins, J.J., Graham, C.H., Moen, D.S., Smith, S.A. & Reeder, T.W. (2006) Evolutionary and ecological causes of latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *The American Naturalist*, **168**, 579–596.
- Wright, H.E., Kutzbach, J.E., Webb, T., Ruddiman, W.F., Street-Perrott, F.A. & Bartlein, P.J. (eds) (1993) *Global climates since the Last Glacial Maximum*. University of Minnesota Press, Minneapolis, MN.

## SUPPLEMENTARY MATERIALS

The following supplementary material is available for this article online:

**Appendix S1** An example set of palaeoclimate GIS layers, generated using the CCM1 model (Kutzbach & Guetter, 1986; Wright *et al.*, 1993) for the last glacial maximum (21,000 yr BP).

**Appendix S2** Details of the downscaling and calibration procedures used to create the palaeoclimate layers in Appendix S1.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2699.2007.01814.x>

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

**BIOSKETCHES**

**Corinne Richards** is a PhD candidate at the University of Michigan whose dissertation research integrates studies of molecular and phenotypic variation among populations of Panamanian golden frogs (*Atelopus varius* and *A. zeteki*). She is interested in the application of phylogeography and landscape genetics to conservation, the role of selection in the evolution of morphological variation, and the effects of climate change and disease on declining amphibian populations.

**Bryan Carstens** is interested in the evolution of ecological communities and the methodological approaches used in comparative phylogeography.

**L. Lacey Knowles'** studies of the processes that initiate or contribute to population divergence span a wide range of temporal and spatial scales. Her primary research interests include the relative contributions of selection and drift to speciation, the evolution of reproductive isolation, the processes generating macroevolutionary patterns of diversity, and the use of statistical approaches (especially coalescent models) to infer the biogeographical, demographic and temporal contexts of lineage divergence.

---

Editor: Michael Patten