Objective choice of phylogeographic models

CrossMark

Bryan C. Carstens[a,*], Ariadna E. Morales[a], Nathan D. Jackson[b], Brian C. O'Meara[b,*]

[a] Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, Columbus, OH 43210-1293, United States
[b] Department of Ecology and Evolutionary Biology, University of Tennessee-Knoxville, United States

ARTICLE INFO

ABSTRACT

Phylogeography seeks to discover the evolutionary processes that have given rise to organismal and genetic diversity. This requires explicit hypotheses (i.e., models) to be evaluated with genetic data in order to identify those hypotheses that best explain the data. In recent years, advancements in the model-based tools used to estimate phylogeographic parameters of interest such as gene flow, divergence time, and relationships among groups have been made. However, given the complexity of these models, available methods can typically only compare a handful of possible hypotheses, requiring researchers to specify in advance the small set of models to consider. Without formal quantitative approaches to model selection, researchers must rely on their intuition to formulate the model space to be explored.

We explore the adequacy of intuitive choices made by researchers during the process of data analysis by reanalyzing 20 empirical phylogeographic datasets using PHRAPL, an objective tool for phylogeographic model selection. We show that the best models for most datasets include both gene flow and population divergence parameters, and that species tree methods (which do not consider gene flow) tend to be overly simplistic for many phylogeographic systems. Objective approaches to phylogeographic model selection offer an important complement to researcher intuition.

1. Models and phylogeographic inference

In the early years of phylogeographic research, inferences about evolutionary history were largely derived qualitatively, after considering the estimated gene tree in combination with the geographic distribution of alleles (e.g., Avise et al., 1987). As the field progressed and incorporated multilocus data, a renewed interest in quantifying the influence of population-level evolutionary processes has led to the increased use of analytical methods that incorporate evolutionary parameters such as gene flow, genetic drift, and population size change. Rather than genealogical patterns, estimates of parameters associated with evolutionary processes now form the basis for most inferences in phylogeography. This expansion of the phylogeographic toolbox has improved the discipline, with recent studies providing clear examples of the importance of these processes to various population dynamics (e.g., Khatchikian et al., 2015; Weir et al., 2015). However, the increased complexity of these models has an associated cost; in general, analytical solutions are not available, and thus approximate approaches such as Approximate Bayesian Computation (ABC; Csilléry et al., 2010) or complex computational machinery such as Markov chain Monte Carlo (MCMC) methods are required to estimate the parameters of interest. These parameters quantify particular evolutionary processes, but a given model cannot include all conceivable processes because the computational machinery becomes intractable in such a case. Researchers are thus forced to choose among an array of methods that each include some, but not all, of the possible parameters of interest. Models of historical demography are central to phylogeographic research (e.g., Knowles, 2009). However, the inferences derived from model-based analyses are conditional on the model used to analyze the data, which may or may not be appropriate to a particular system (e.g., Nielsen and Beaumont, 2009). Choosing inappropriate models can lead to inaccurate estimates of the parameter of interest because they fail to account for other evolutionary processes that have also influenced allele frequencies (e.g., Koopman and Carstens, 2010).

There are three general approaches that one could take to identify a model for the purposes of data analysis (Table 1). Most researchers identify a model for analysis intuitively, using their knowledge of the focal taxon to identify the evolutionary processes that they believe to be important. They then identify a software package that includes parameters that represent these processes. For example, a researcher who is interested in landscape-level source-sink dynamics might use Migrate-*n* (Beerli and Felsenstein, 2001) to estimate gene flow, while a researcher who is motivated to assess the species status of several allopatric populations may choose ∗BEAST (Heled and Drummond, 2010) to

Table 1
Three general approaches can be used by researchers to identify a model for phylogeographic data analysis. Each emphasizes different components as the basis for the resulting inference.

Approach	Contribution of biogeographic or life history information	Role of simulation	Role of parameters	Potential problems	Example references
Intuitive model identification	Existing information interpreted by scientist to identify analytical model from available set of models	Optimal parameter estimates often derived from MCMC	Parameter estimates are evaluated and serve as the basis for phylogeographic inference	Identification of models with a poor fit to the history of the system can lead to confirmation bias	Dolman and Moritz (2006)
Phylogeographic hypothesis testing	Existing information used to design model(s) to be tested	Simulation used to construct null or prior distribution	Parameter estimates (or parameter priors) used in simulations	Does not consider unexpected evolutionary processes or biogeographic scenarios	Knowles (2001), Fagundes et al. (2007)
Objective model selection	Does not contribute to the set of evaluated models	Simulation used to approximate the probability of each model	Parameters are optimized in order to calculate approximate likelihoods	Model space is effectively infinite, best model may not be represented	Morales et al. (2017)

estimate depth of lineage divergence or IMa2 (Hey, 2010) to estimate divergence with gene flow. In such cases, the researchers will be able to estimate the parameters that enable them to make an inference about the biological process that is considered to be important, but will not explore how other biological processes, unquantified by the measured parameters, might influence the evolution of their system. Such intuitive model choice may result in a sort of confirmation bias, where the researcher confirms that the processes that they think were occurring in the system are important by failing to consider alternative processes (e.g., Nickerson, 1998).

Researcher intuition is also of central importance to simulation-based approaches that have long been applied in the testing of phylogeographic hypotheses. Knowles (2001) tested phylogeographic hypotheses by designing two models of refugial population structure and subsequently using parametric simulation to generate replicate datasets corresponding to each model. As presented by Knowles (2001), phylogeographic hypothesis testing allows researchers to reject models that are not realistic given their data. However, it forces them to evaluate models on an individual basis, and to conform to a statistical framework that may not be suited to a science that lacks the experimental controls used in classical hypothesis testing (e.g., Cleland, 2001). In historical sciences such as phylogeography, it is more useful to evaluate multiple working hypotheses, where each represents a plausible historical scenario. A similar approach has long been advocated in the geological sciences (e.g., Chamberlin, 1890), which are also concerned with how processes acting over long time periods shape systems that are observed in the present. ABC approaches (Csilléry et al., 2010) are also used to test phylogeographic hypotheses, and also rely on user intuition to design the models that are used to approximate the posterior distribution of parameter values. Phylogeographic investigations incorporating ABC have been conducted in many systems, with results generally consistent with other sources of data in model systems (e.g., Fagundes et al., 2007). However, identifying the optimal model using ABC has inherent complications (see Robert et al., 2011), and error rates likely increase as the number of models in the analysis increases (Pelletier and Carstens, 2014).

Whether researchers employ null hypothesis testing, ABC, or other methods (e.g., Carstens et al., 2013a), it is common to include only a handful of hypotheses (~2–6) that are chosen because they represent evolutionary scenarios that are considered to be plausible by the researcher. In model systems, these scenarios are likely derived from a great deal of existing information, but in non-model systems, where previous work may be limited, they are typically derived intuitively from general knowledge of the focal taxon in addition to the biogeographic and geological history of its range. Consequently, there is no guarantee that any set of hypotheses will include a scenario that approximates the actual evolutionary history, or even a parameter corresponding to the evolutionary process that has most influenced genetic variation. This is potentially problematic because the number of possible models is very large, particularly when they include both phylogenetic divergence and gene flow among some of the branches. As a result, the most appropriate model may not be included in the set of models devised by the researcher. To some scientists (e.g., Templeton, 2004), this shortcoming represents a fatal flaw of phylogeographic hypothesis testing: one model will necessarily be identified as being most probable given the data, but this model might not be as appropriate as a model that was not included in the model set, and as such the choice of the best model from a collection of poor models may lead to false confidence on the part of the researcher. Simply expanding the model set to include every possible model may not be plausible for a variety of reasons, including statistical considerations related to testing larger numbers of models (e.g., Nakagawa, 2005). A framework for objective phylogeographic model selection, particularly one that includes a large numbers of possible models, allows researchers to identify a reasonably good model for their focal system, which in turn could be used to estimate both the parameters of interest as well as
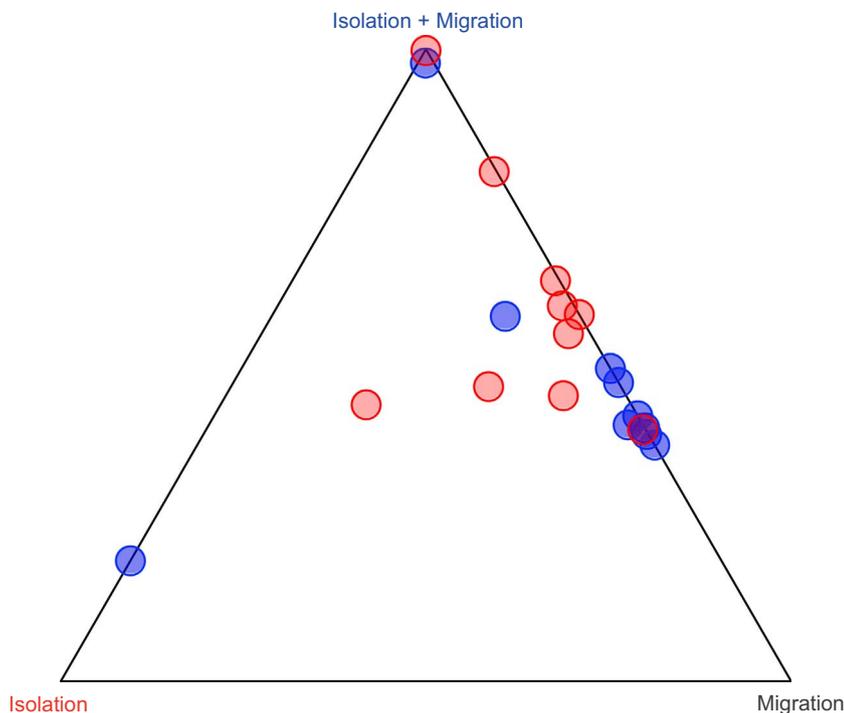
Fig. 1. Triangle plot showing the weighted probabilities from a PHRAPL analysis of 20 datasets. Each of the vertices represents a commonly used approach to phylogeographic data analysis. The top corresponds to IMa2 (Hey, 2010), the lower right to Migrate-n (Beerli and Felsenstein, 2001) and the lower left to a species tree analysis (i.e., *BEAST; Heled and Drummond, 2010). Datasets are colored on the basis of the analytical model chosen intuitively by the authors of each investigation. The probabilities of each analysis are shown decreasing from the respective vertex in increments of 0.1. Weighted probabilities were corrected for the unevenness of the model space in respect to the three model classes such that a dataset with equivalent probability for each of the models would appear in the center of the triangle (marked with a small yellow circle). Results indicate that there is little support for the isolation-only model (i.e., species trees) in these systems. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parameters associated with the other processes that influence the system.

Molecular systematists are familiar with objective model selection. When estimating phylogenetic trees from sequence data, researchers first use a program such as MODELTEST (Posada and Crandall, 1998) to select a model of sequence evolution, and then use other software to generate a precise estimate of the parameter of interest (i.e., the phylogeny and branch lengths). However, it is worth noting that this was not always the case. Prior to the widespread adoption of objective model selection, papers were routinely published that promoted unconventional relationships on the basis of phylogenetic trees that were poor estimates of the true phylogeny because they were estimated using inappropriate models (see Sullivan and Joyce, 2005). In molecular systematics, parameters such as gamma rate heterogeneity are typically only important for inference because they allow phylogeny (the true parameter of interest) to be estimated accurately. However, in phylogeographic research, the parameters that may be included in analytical models are decidedly not nuisance parameters, particularly the parameters that model evolutionary processes such as gene flow, genetic drift, or population history, and so which ones are most important to include is itself of biological importance. While objective approaches to model selection represent a break with the long tradition in phylogeography, it is not the case that inference is an either/or proposition. Just as researchers continued to make inferences based on the geographic pattern of genealogy after they adopted hypothesis testing approaches (e.g., Fig. 5 of Knowles, 2001), it is possible to compare the results of phylogeographic hypothesis testing, where models are chosen intuitively, with those from objective model selection exercises. While automated model selection has been criticized (e.g., Massatti and Knowles, 2014) as potentially uninformative, especially if the differences between models do not relate to key biological questions, these approaches may be the most effective safeguard against the inherent biases that a researcher may bring with them to a particular empirical system. Hypothesis testing can often become a mere exercise in comparing a model that biologists believe is realistic with a trivial null hypothesis; with objective model selection, both the alternative and null models can be included in a richer variety of models. In fact, we present evidence below that by relying on intuition about which models

are realistic, we often miss key processes.

2. Objective analysis of empirical data

If the intuition of biologists is at all accurate, one would predict that researchers are skilled at identifying analytical models that contain the parameters that capture the key processes operating in a system. Since phylogeographic inference to date has been almost entirely based on parameters that were estimated using models that were selected intuitively, reanalyzing these data using an objective selection of models should, assuming that intuition is adequate and the analysis is performed correctly, identify the same models used by the researchers to analyze their data. To explore this question, we reanalyze published data (see Supporting Information 1) and compare the demographic models selected by the objective model selection method PHRAPL (see Supporting Information 2) to those chosen by researchers. *Our criteria were as follows:* We selected studies (i) that were published between January 2012 and December 2014; (ii) that used statistical models implemented using three popular methods (*Beast, IMa2, Migrate-n) chosen intuitively by researchers; (iii) that were based on sequence data from multiple loci available to us from the authors or from public databases; (iv) that analyzed their data using at least three defined lineages (to allow for the same, moderately sized model sets to be used for all datasets); and (v) sampled four or more individuals per population.

We analyzed each empirical dataset using a set of 81 models previously used in simulation testing (e.g., Jackson et al., 2017b; Morales et al., 2017), and with analysis settings that conform to the best practices of a PHRAPL analysis as currently understood (see Supporting information). Notably, this model set contains many of the models that are commonly considered in phylogeographic studies (e.g., those implemented in IMa2, *BEAST, etc.), such that for a given data set, the model(s) used by the researchers was available in the model set. Our results indicate that the intuition of the researcher is often incomplete and/or inadequate. For example, our reanalysis of recently published data suggests that the process of gene flow has been underappreciated, as it is almost always implicated as an important evolutionary process by PHRAPL (Fig. 1), despite the fact that many of the original studies

utilize only isolation models (i.e., species trees). Note that while the number of possible isolation-with-migration (IM) histories is inherently much larger than the number of *n*-island (MO) and isolation-only (IO) histories (90% of the model set was composed of IM models, whereas IO and MO models made up only 5% each), we normalized the model class probabilities depicted in Fig. 1 based on the frequency of each model class to account for this bias. Additionally, the absence of support for isolation-only models among empirical datasets suggests that reliance on species tree approaches (in our sample, the most commonly applied model to phylogeographic systems) may fail to account for important evolutionary processes that maintain population connectivity. Gene flow is a key parameter in species-level systematic investigations, and its omission from the analytical model used for parameter estimation would likely bias estimates of other parameters of interest, such as the divergence among populations (e.g., Leaché et al., 2014). For example, in each of our empirical datasets, there were four tree structures (a polytomy and three resolved trees) that were analyzed with and without migration. In 79 out of these 80 examples (i.e., the product of 20 data sets each containing four possible isolation-only models), estimates of population divergence time were lower (typically much lower), under the isolation-only model as compared to the divergence times in the best fitting migration model. Over all the 80 pairs, the median branch length of the species tree under the no-migration model (which, again, was typically not very probable given the data) was just 13% of the corresponding branch length for the best migration model. Thus, for empirical datasets, ignoring migration would produce a substantial change in the resulting inference of population divergence, even if gene flow is not explicitly a process of interest. These results are consistent with a growing appreciation among systematicists that the species tree framework may not be appropriate for all datasets (e.g., Edwards et al., 2016).

The results presented above illustrate how phylogeographic inference can be improved by more objective methods of model selection, as such methods consider models that represent a wider range of evolutionary processes than is typically considered by researchers. However, objective model selection should complement, not replace, the intuition of researchers, and is an important addition to the toolkit of phylogeographic analytical methods. Our results suggest that there are two advantages to looking at multiple objective hypotheses. The first is surprise: rather than learning about whether our biological intuition can outperform a null model, we could identify unexpected processes to be important in a given system. The second is a greater appreciation of uncertainty: at phylogeographic scales, there can be a lot of complexity with a rather modest amount of data available to serve as the basis for inference. In many cases, an informed model will clearly outperform a simple null, but it might not outperform a model of equal complexity that includes other parameters. This recovery of uncertainty is itself a discovery about the system, and prevents overconfidence that can lead to incorrect beliefs about relevant factors. This may be especially relevant for phylogeographic questions relating to conservation. Objective model choice can be easily incorporated into the phylogeographic workflow. For example, one could select from a broad set of models using something like PHRAL, and then use other methods (e.g., MCMC-based or ABC) to more precisely estimate parameters based on that model. The exploration of model space, because it is effectively infinite, requires flexibility in both the models considered by the researcher and the methods that implement these models.

3. Outstanding challenges

Apart from the technical and computational challenges associated with objective model selection in general, which are considerable (e.g., Jackson et al., 2017a, 2017b), a number of outstanding theoretical questions pertaining to objective model choice deserve consideration. One is defining and navigating the full complexity of the model space. For example, the possible number of topologies for *n* sampled populations is slightly higher than the number of resolved and unresolved phylogenetic trees for the same number of OTUs (the difference comes from allowing histories with at least some non-coalescing populations as in a typical island population model). This can be very high. Even for a simplified system with three populations, 1–2 population divergence events, and a single migration rate, there are 1462 possible histories. To reduce the complexity of model space, constraints can be added. For example, users might specify that population coalescence or migration is to occur only between geographically proximate populations, or that the overall population sizes are constant, but it is unclear how the introduction of such constraints might bias the results. Consequently, identifying the optimal approach to exploring this model space is of key concern. Another challenge is to determine how many loci are required to ensure that the optimal model in a given data set is identified. Results from simulated data (e.g., Jackson et al., 2017a) suggest that this number may be ~20 loci, while results from a recent empirical study suggest that it may be higher (Morales et al., 2017). PHRAPL scales well with genomic data and has been used to analyze ~1000 loci in recent studies (e.g., Morales et al., 2017). Finally, understanding the biases that can potentially result when comparing models with vastly different dimensions, such as panmixia in a single population vs. an IM model with three populations, remains a substantial challenge, and one that has particular practical ramifications to species delimitation (Carstens et al., 2013b).

4. Conclusions

The basic inference process used by evolutionary geneticists is to (i) collect samples and genetic data, (ii) identify a statistical model for data analysis, (iii) estimate one or more parameters associated with the above processes from their data using the model, and finally to (iv) make inferences on the basis of these parameter estimates. Genomic data have never been easier to collect and are increasingly abundant even in non-model systems (Garrick et al., 2015). Since parameter estimates and the resulting inferences are required for publication, it is easy to overlook the step that identifies the analytical model. However, good data analyzed using inappropriate models will result in parameter estimates that fail to account for important biological processes, and thus produce inferences that can mislead researchers (e.g., Thomé and Carstens, 2016). Indeed, our reanalysis of 20 empirical datasets highlights the utility of phylogeographic model selection by demonstrating that the intuition of researchers (including some of the authors of this paper) is sometimes flawed in regards to choosing the models used to analyze data from empirical systems. Phylogeography has long been promoted as a 'bridge' between population genetics and systematics (Avise et al., 1987), but the reliance on the species tree has constrained the discipline to one side of this divide. Advances in phylogeographic model selection will be key to reducing the entrenched reliance on intuition. They enable researchers to select demographic models for data analysis without relying on guesswork or potentially incorrect assumptions about the processes likely to be important to their systems. In many cases, the intuition of researchers is likely correct, either in part or in whole. If this is the case, objective methods such as PHRAPL will confirm the developing understanding of the evolutionary history of the focal system.

Data archiving

Data will be archived on Dryad upon acceptance.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.08.018. Scripts used in the PHRAPL analysis are available on Dryad: http://dx.doi.org/10.5061/dryad.3d120.

References

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. Ann. Rev. Ecol. Evol. Syst. 18, 489–522.

Beerli, P., Felsenstein, J., 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc. Natl. Acad. Sci. USA 98, 4563–4568.

Carstens, B.C., Brennan, R.S., Chua, V., Duffie, C.V., Harvey, M.G., Koch, R.A., McMahan, C.D., Nelson, B.J., Newman, C.E., Satler, J.D., Seeholzer, G., Posbic, K., Tank, D.C., Sullivan, J., 2013a. Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. Mol. Ecol. 22, 4014–4028.

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013b. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383.

Chamberlin, T.C., 1890. The method of multiple working hypotheses. Science 15, 92–96.

Cleland, C.E., 2001. Historical science, experimental science, and the scientific method. Geology 29, 987–990.

Csilléry, K., Blum, M.G., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian computation (ABC) in practice. Trends Ecol. Evol. 25, 410–418.

Dolman, G., Moritz, C., 2006. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). Evolution 60, 573–582.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94, 447–462.

Fagundes, N.J., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F.M., Bonatto, S.L., Excoffier, L., 2007. Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. USA 104, 17614–17619.

Garrick, R.C., Bonatelli, I.A., Hyseni, C., Morales, A., Pelletier, T.A., Perez, M.F., Rice, E., Satler, J.D., Symula, R.E., Thomé, M.T.C., Carstens, B.C., 2015. The evolution of phylogeographic data sets. Mol. Ecol. 24, 1164–1171.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Hey, J., 2010. Isolation with migration models for more than two populations. Mol. Biol. Evol. 27, 905–920.

Jackson, N.D., Carstens, B.C., Morales, A.E., O'Meara, B.C., 2017a. PHRAPL: phylogeographic inference using approximate likelihoods. Syst. Biol. http://dx.doi.org/10.1093/sysbio/syx001. (in press).

Jackson, N.D., Carstens, B.C., Morales, A.E., O'Meara, B.C., 2017b. Species delimitation with gene flow. Syst. Biol. 66, 799–812. http://dx.doi.org/10.1093/sysbio/syw117.

Khatchikian, C.E., Prusinski, M.A., Stone, M., Backenson, P.B., Wang, I.N., Foley, E., Seifert, S.N., Levy, M.Z., Brisson, D., 2015. Recent and rapid population growth and range expansion of the Lyme disease tick vector, *Ixodes scapularis*, in North America. Evolution 69, 1678–1689.

Knowles, L.L., 2001. Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. Mol. Ecol. 10, 691–701.

Knowles, L.L., 2009. Statistical phylogeography. Ann. Rev. Ecol. Evol. Syst. 40, 593–612.

Koopman, M.M., Carstens, B.C., 2010. Conservation genetic inferences in the carnivorous pitcher plant *Sarracenia alata* (Sarraceniaceae). Cons. Genet. 11, 2027–2038.

Leaché, A.D., Harris, R.B., Rannala, B., Yang, Z., 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63, 17–30.

Massatti, R., Knowles, L.L., 2014. Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. Evolution 68, 2833–2846.

Morales, A.E., Jackson, N.D., O'Meara, B.C., Carstens, B.C., 2017. Speciation with gene flow in North American *Myotis* bats. Syst. Biol. 66, 440–452.

Nakagawa, S., 2005. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behav. Ecol. 5, 1044–1045.

Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gener. Psych. 2, 175–220.

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

Pelletier, T.A., Carstens, B.C., 2014. Model choice in phylogeography using a large set of models. Mol. Ecol. 23, 3028–3043.

Posada, D., Crandall, K.A., 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Robert, C.P., Cornuet, J.-M., Marin, J.-M., Pillai, N.S., 2011. Lack of confidence in approximate Bayesian computation model choice. Proc. Natl. Acad. Sci. USA 108, 15112–15117.

Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. Ann. Rev. Ecol. Evol. Syst. 31, 445–466.

Templeton, A.R., 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. Mol. Ecol. 13, 789–809.

Thomé, M.T., Carstens, B.C., 2016. Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. Proc. Natl. Acad. Sci. USA 113, 8010–8017.

Weir, J.T., Faccio, M.S., Pulido-Santacruz, P., Barrera-Guzmán, A.O., Aleixo, A., 2015. Hybridization in headwater regions, and the role of rivers as drivers of speciation in Amazonian birds. Evolution 69, 1823–1834.