Statistical phylogeography

L. LACEY KNOWLES and WAYNE P. MADDISON
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721–0088, USA

Abstract

While studies of phylogeography and speciation in the past have largely focused on the documentation or detection of significant patterns of population genetic structure, the emerging field of statistical phylogeography aims to infer the history and processes underlying that structure, and to provide objective, rather than *ad hoc* explanations. Methods for parameter estimation are now commonly used to make inferences about demographic past. Although these approaches are well developed statistically, they typically pay little attention to geographical history. In contrast, methods that seek to reconstruct phylogeographic history are able to consider many alternative geographical scenarios, but are primarily nonstatistical, making inferences about particular biological processes without explicit reference to stochastically derived expectations. We advocate the merging of these two traditions so that statistical phylogeographic methods can provide an accurate representation of the past, consider a diverse array of processes, and yet yield a statistical estimate of that history. We discuss various conceptual issues associated with statistical phylogeographic inferences, considering especially the stochasticity of population genetic processes and assessing the confidence of phylogeographic conclusions. To this end, we present some empirical examples that utilize a statistical phylogeographic approach, and then by contrasting results from a coalescent-based approach to those from Templeton's nested cladistic analysis (NCA), we illustrate the importance of assessing error. Because NCA does not assess error in its inferences about historical processes or contemporary gene flow, we performed a small-scale study using simulated data to examine how our conclusions might be affected by such unconsidered errors. NCA did not identify the processes used to simulate the data, confusing among deterministic processes and the stochastic sorting of gene lineages. There is as yet insufficient justification of NCA's ability to accurately infer or distinguish among alternative processes. We close with a discussion of some unresolved problems of current statistical phylogeographic methods to propose areas in need of future development.

Keywords: biogeography, coalescence, genealogy, phylogeography, population genetics, speciation

Received 20 June 2002; revision received 21 August 2002; accepted 21 August 2002

Introduction

Studying patterns of genetic variation in a geographical context via gene trees (i.e. phylogeography) has contributed considerably to our understanding of what factors have influenced population structure and species divergence (e.g. Avise 1994). An explicit focus on the species' history, in particular the biogeographical past, has set phylogeography apart from classical population genetics. Phylogeographic inferences are usually derived

by studying the reconstructed genealogical histories of individual genes (gene trees) sampled from different populations. However, because many events may have occurred in the past (e.g. population expansion, bottlenecks, vicariance, and migration), and gene lineages may be lost by chance, a species' history might not be easily inferred from a gene genealogy. Nonetheless, the field of phylogeography is growing quickly, with many papers publishing interpretations of past events based on patterns of genetic variation. Our goal in this paper is to consider how interpretations of population history are best made, and whether current enthusiasm should be more strongly tempered with caution.

Correspondence: L. Lacey Knowles.
E-mail: knowles@u.arizona.edu

While there is an obvious trend towards formalizing tests of phylogeographic scenarios (Stone 2000; Emerson *et al.* 2001; Hare 2001; Wakeley 2002), depending upon the particular methodological approach, there are fundamental differences among these historical inferences. They range from largely qualitative, *post hoc* explanations of historical processes (e.g. Avise 1994) and tests of significant population structure (e.g. Excoffier *et al.* 1992), to coalescent-based tests of specific hypotheses about the biogeographical and demographic past (e.g. Edwards 1993; Knowles *et al.* 1999; Kliman *et al.* 2000; Knowles 2000; Milot *et al.* 2000; Emerson *et al.* 2001; Hare 2001). The reconstructed gene tree may be used directly to make an inference or just as an intermediate step towards the derivation of an historical estimate (Knowles 2003; Wakeley 2002). Such inferences may be based on either a single locus, such as mitochondrial DNA, or multiple loci. However, irrespective of which approach is used, an explicit consideration of its limitations is not only important for selecting an appropriate method, but it is also integral to making accurate historical inferences, and in particular, avoiding over-interpreting the analyses. For example, was the error associated with the historical reconstruction assessed? Might alternative hypotheses have fitted the data equally well? Were assumptions made about the species history accurate?

Recent population genetic models based on coalescent theory (Kingman 1982) do provide a statistical framework for estimating demographic parameters, such as effective population size, migration rates, divergence times, and population growth or decline, as well as hypothesis testing (e.g. Takahata *et al.* 1995; Kuhner *et al.* 1998; Beerli & Felsenstein 1999, 2001; Bahlo & Griffiths 2000; Nielsen & Slatkin 2000; Wakeley 2001). However, despite the statistical power and computational sophistication of these models for parameter estimation (e.g. joint likelihood estimation of divergence times and migration rates), a simple underlying population history is typically assumed (e.g. simple divergence of two undivided populations of constant size). Attention to the possibility of more complex histories (e.g. a series of sequentially diverging populations with varying migration rates) could not only result in improved estimates of parameters but also demand that new methods be developed to reconstruct such complex histories. Population genetic theory has paid little attention to the inference of complex, explicitly geographical population histories. For example, the few models that consider population subdivision can not be used to detect isolation by distance (Bahlo & Griffiths 2000; Wakeley 2001; for an exception, see Beerli & Felsenstein 1999).

In contrast, a method like Templeton's nested cladistic analysis (NCA) considers and seeks to distinguish among a diverse array of historical processes (Templeton *et al.* 1995). Rather than making presumptions about the under-lying processes, NCA attempts to derive a historical reconstruction using an inference key. However, as we discuss later in more detail, NCA does not attempt to distinguish statistically among alternative interpretations, nor does it provide an estimate of the uncertainty in its conclusions. Thus, for any interpretation derived from Templeton's inference key, we do not know the confidence limits on the reconstructed history, whether they are so broad as to include many unconsidered alternatives, or if an alternative hypothesis would be almost equally well supported by the data. We are not implying that the results from a method such as NCA will necessarily be inaccurate. In some cases, the detailed history suggested by NCA may indeed be an accurate reconstruction; however, there is no way to tell if it is.

In this paper we advocate the use and development of phylogeographic methods that make both explicit statistical links between process, prediction and test (like the coalescent-based population genetic models) and consider a diverse array of processes and histories (like NCA). With such methods, explicit phylogeographic histories can be proposed, stochastic expectations derived and compared against data in formal statistical tests. Thus, not only can errors associated with a specific hypothesis be assessed, but support for alternative explanations can also be evaluated. Such a framework is critical for addressing particular phylogeographic hypotheses, such as whether contemporary populations are derived from the same or different ancestral populations (e.g. Takahata 1993; Knowles 2001b), but it is also key for accurately estimating demographic parameters, such as the effective size of an ancestral population (e.g. Takahata *et al.* 1995; Wakeley & Hey 1998; Wakeley 2001).

We will refer to this general approach as, 'statistical phylogeography'. Falling within its scope are a variety of existing methods, including those that estimate basic population genetic parameters, which are of course components of the population history (e.g. Kuhner *et al.* 1998; Beerli & Felsenstein 1999; Bahlo & Griffiths 2000; Wakeley 2001; Tang *et al.* 2002). However, a review of these methods is beyond the realm of this paper, nor are we advocating any method in particular. Rather, our intent is to identify, through various examples, some of the conceptual issues that pertain to how historical inferences are made, especially in reference to considering the stochasticity of population genetic processes and assessing the confidence of various phylogeographic conclusions. To this end, the examples we present use statistical tests based on computer simulations while addressing a diversity of questions. We then discuss the utility of a statistical phylogeographic approach, and contrast results from a coalescent-based approach to those from NCA with computer simulations. We close with a discussion of some of the unresolved problems of current methods to propose areas in need of future developments.

Examples of statistical phylogeographic approaches

In this section we present examples to clarify how inferences are made in statistical phylogeography. It should not be assumed that methods not presented here are necessarily excluded from 'statistical phylogeography'. We also do not intend to imply that all aspects of the presented examples are necessary or even optimal for addressing statistical phylogeographic questions. Rather, we use a diversity of examples merely to show common features of statistical phylogeographic methods. Indeed, population histories other than strictly branching trees should be considered in many circumstances, as well as uncertainty in the reconstruction of the gene tree, and other test statistics may prove more powerful.

The presence of significant population genetic structure has been demonstrated in a diversity of species with disparate ecologies and natural histories (Avise 1994). Statistical phylogeography is confronted with the challenge of explaining what has caused a particular structure, and endeavors to provide objective, rather than *ad hoc* explanations. Of the examples that we present, two focus on tests of historical hypotheses, specifically the biogeographical context of population divergence (Milot *et al.* 2000; Knowles 2001b). The other emphasizes how specific parameters can be estimated; in this case the timing of population divergence (Edwards & Beerli 2000).

Tests of historical biogeographic hypotheses

Because biogeographical events are reflected in the branching patterns of gene trees, genealogies are now frequently used to study the processes generating species divergence (Avise 1989; Harrison 1991; Riddle 1996), as well as population differentiation (Barton & Wilson 1995; Stone 2000). For example, Milot *et al.* (2000) used a genealogical approach to test whether patterns of divergence within the yellow warbler (*Dendroica petechia*) supported Mengel's model of warbler diversification that proposed that speciation resulted from the disruption of ancestrally contiguous populations into allopatric populations during the Pleistocene glaciations (Mengel 1964). While an explicit phylogenetic prediction of Mengel's hypothesis is that western species are derived from eastern forms, when applied at the intraspecific level, this pattern could be a result of gene flow rather than ancestry.

To distinguish between these hypotheses, Milot *et al.* estimated the population origin of specific mutations (Fig. 1). If mutations at the node uniting the E2 and W1 haplotypes originated in the eastern populations, then the ancestry hypothesis would be supported, whereas an origin in the western population would support the migration
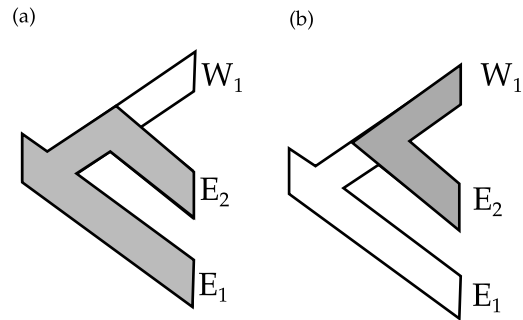

Fig. 1 The two phylogeographic scenarios Milot *et al.* (2000) tested by determining whether mutations at the node uniting E2 and W_1 originated within (a) the eastern warbler populations, supporting the ancestral hypothesis, or (b) the western population, supporting the migration hypothesis.

hypothesis. The origin of these mutations was estimated using maximum-likelihood estimates of parameter values from an evolutionary model based on the coalescent process (using GENETREE & Griffiths 1998), conditioned on estimates of migrations rates (using MIGRATE, Beerli & Felsenstein 1999) and growth rates (using FLUCTUATE, Kuhner *et al.* 1998) within each subpopulation. Simulations showed a significant probability (> 95%) that the mutations originated within the eastern populations, thus supporting the ancestral hypothesis, and thereby confirming Mengel's proposed model of warbler speciation.

In another study on the effects of the Pleistocene glacial cycles on diversification, Knowles (2001b) used a coalescent-based approach to test whether contemporary populations of a montane grasshopper species (*Melanoplus oregonensis*) were derived from different ancestral populations, as predicted if divergence took place among allopatric glacial refugia. This species inhabits the 'sky-islands', or mountaintops, of western North America. Although these grasshoppers are geographically isolated today among the montane meadows of the northern Rocky Mountains, they were displaced by glaciers during the Pleistocene.

In this study, F_{ST}-analyses indicated there was significant population structure, and the mitochondrial gene tree also provided evidence of phylogeographic structuring of haplotypes. However, none of the regional phylogeographic groups were monophyletic, and haplotypes from some populations occurred in more than one regional group. While it was possible that the regional groups reflected the founding of contemporary sky-island populations from different ancestral-refugial source populations, the phylogeographic structure might have arisen from the fragmentation of a widespread ancestor (i.e. a single ancestral population), followed by the subsequent loss of haplotypes from some sky-island populations by drift.

To distinguish among these hypotheses, the coalescent process was used to evaluate the probability that *n* gene
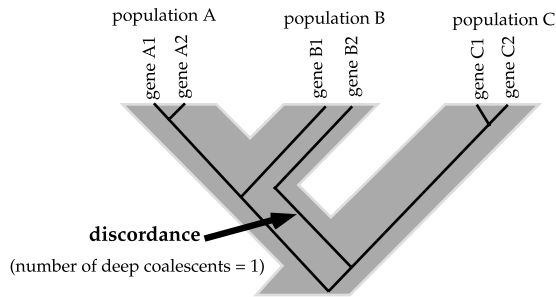
Fig. 2 Schematic representations of a gene tree contained within a population tree; an arrow identifies the discordance between the gene and population trees. This discord can easily be quantified by counting the number of deep coalescents (Maddison 1997), or the failure of gene lineages to coalesce within their respective population lineages (assuming that there is no gene flow among the populations).

Fig. 3 The two phylogeographic hypotheses Knowles (2001b) tested with coalescent simulations to determine if contemporary sky-island populations of montane grasshoppers were derived from (a) a single ancestral population as represented by a fragmentation model, or (b) multiple ancestral populations, using an allopatric refuge model (groupings of populations were identified from previous phylogeographic analyses (see Figs 3 and 4, Knowles 2001b).

lineages would fail to coalesce within population lineages representing purported allopatric ancestral refugial-source populations, hence, producing a discordance between the gene and population tree (Fig. 2). Alternative historical scenarios were represented as different population trees (Fig. 3). The null hypothesis of a single ancestral source population was modelled where each individual, contemporary sky-island population corresponded to a population in the population tree. The alternative hypothesis of multiple ancestral populations, was modelled with groups of sky-island populations treated essentially as a single population if they were colonized from the same ancestral refugial-source population. Gene trees simulated by a neutral coalescent process under the fragmentation model (Fig. 3a) were used to derive an expected distribution for the amount of discord that would be predicted between a genealogy and a population tree representing the allopatric refugia model (Fig. 3b), when the data have actually evolved under a history in which the contemporary populations were derived from a single ancestral source. The discord between the reconstructed gene tree and the allopatric refuge model was then compared to this expected distribution. The simulations showed that the observed value was significantly lower than predicted ($P < 0.05$) if the populations had been founded by a single ancestral population, and that this conclusion was robust over a range of different times of population splitting. Thus, the sky-island populations of *M. oregonensis* were most likely colonized from multiple ancestral-source populations, supporting the role of allopatric glacial refugia in promoting divergence.

Time since population divergence

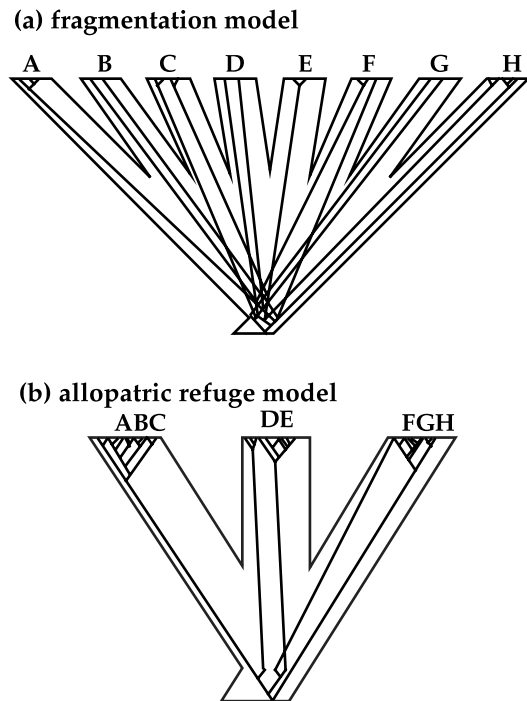Estimating the time of population or species splitting is notoriously difficult (Arbogast *et al.* 2002) Not only are

there issues surrounding the use and calibration of molecular clocks, but a perhaps more insidious problem stems from differences between the timing of gene and population divergence (Fig. 4). The former can be surmounted through the use of taxon-specific local clocks (e.g. Beerli *et al.* 1996) which have been tested empirically (Huelsenbeck & Crandall 1997) or methods for dating divergences in the absence of rate constancy (Sanderson 1997; Kishino *et al.* 2001). The extent to which estimates of divergence times derived from gene trees approximate population or species divergence depends upon both the timing of divergence and the ancestral population from which the populations diverged (Maddison 1997; Takahata 1989, 1991; Wakeley 2000). When divergence times between species are large, the discrepancy between the time of gene and species divergence becomes inconsequential (Fig. 4a). However, this distinction is critical to estimates of divergence times for recent events (Fig. 4b) such as diversification during the Pleistocene (Riddle 1996; da Silva & Patton 1998; Schneider *et al.* 1998; Knowles 2000, 2001a,b; Maddison & McMahon 2000) or the origin of humans (Takahata *et al.* 1995; Takahata & Satta 1997).
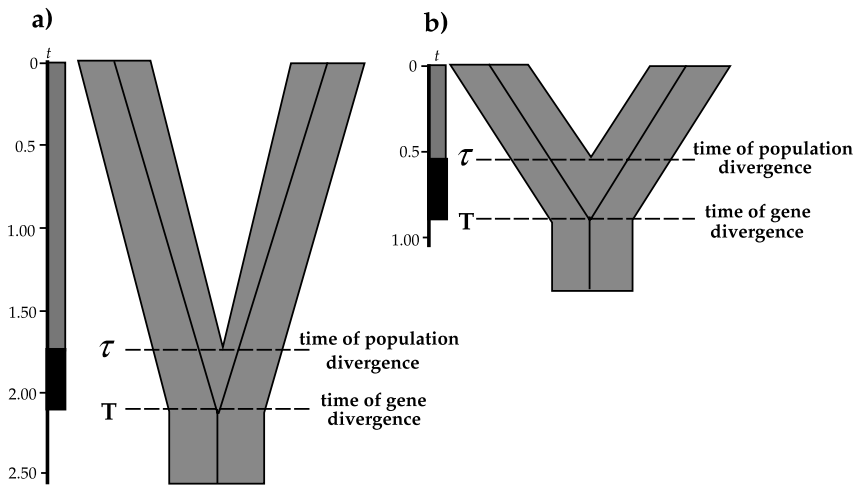
Fig. 4 The timing of gene divergence generally overestimates population divergence times. The discrepancy between gene and population divergence is less for events in the more distant past (a) compared to the effect on estimates of recent divergence times (b) where the difference between τ and T represents a much larger proportion of the estimated population divergence time.

The degree of overestimation of population divergence times can be substantial, increasing dramatically when ancestral coalescence times are lengthened by demographic factors such as population substructure within the ancestral population (Slatkin 1987; Wakeley 2000; reviewed in Edwards & Beerli 2000). Therefore, even though variance in gene divergence times among codistributed species in phylogeographic studies are often interpreted as differences in the timing of species divergence, it is possible that the difference actually reflects variance in coalescence times or ancestral population sizes among species.

Because the variance in gene divergence times, T, between a pair of populations contributed by the effects of drift in the ancestral population can be substantial (Fig. 4), Edwards & Beerli (2000) suggested that an appropriate null hypothesis in comparative phylogeographic studies should be equivalence of τ between populations. Using the divergence times estimated by Klicka & Zink (1997) for 35 species pairs of North American birds, Edwards and Beerli asked whether the data implied the 'same' divergence for the avian pairs. Klicka and Zink had calculated that the mean divergence of 5.1%, corresponding to an average divergence of 2.5 million years ago (Ma), with a range of coalescence times from 200 000 years to 5.55 Ma, assuming a 2%/million years mitochondrial clock. Klicka and Zink then interpreted the data as evidence against a Pleistocene model of speciation. However, Edwards and Beerli's analyses showed that with an ancestral N_f of 1.5×10^6 females the standard deviation of the coalescence times of these 35 species was consistent with a single vicariant event in which all the population divergences were no older than the minimum estimated T of 200 000 years. Even if this value is rejected as being biologically implausible, Edward and Beerli's study shows that for single locus data for which reciprocal monophyly has been achieved, it is not possible to tell how much of the total gene divergence

between species (or populations), or differences among codistributed species, actually results from divergence since species separation, τ (Fig. 4).

Limitations and prospects of statistical phylogeography

The examples we present each exemplify a statistical approach to phylogeographic inference that considers the stochasticity of population genetic processes and offer an explicit assessment of the confidence in any specific conclusion. They illustrate the broad array of historical scenarios that can be modelled, including different geographical scenarios (e.g. Milot *et al.* 2000; Knowles 2001b), timing of population divergence (e.g. Edwards & Beerli 2000), and estimates of effective population sizes, migration and growth (e.g. Milot *et al.* 2000). They also differ with respect to how the genealogy is used, addressing evolutionary questions directly from the structure of gene tree itself, as opposed to using the properties of gene trees to obtain a genetic parameter estimate (Knowles 2003; Wakeley 2002).

We do not want to leave the reader with the impression that the methods exemplified above are flawless, or that we are advocating them in particular. Ideally, a method needs to be able to provide an accurate representation of the past, consider a diverse array of processes, and yield estimates of the history with some measure of the error of that estimate. Such a method does not yet exist.

We will now describe some unresolved problems with phylogeographic methods, in general, so as to point to where future developments are needed, and to emphasize what considerations must be kept in mind when attempting to make historical inferences. We will also reconsider nested cladistic analysis, its current limitations, and to what extent it may (or may not) provide a solution to some of the challenges statistical phylogeography poses.

Extracting information from the data

To distinguish among historical scenarios, theoretical expectations must first be derived either explicitly by simulation, or implicitly in analytical calculations. These predictions need to be quantitatively compared against the observations. That is, a test-statistic needs to be calculated to compare the observed data to the expected, but the actual test-statistic used will vary (Table 1), and will depend on how best to extract the information relevant to the question being addressed. Therefore, the choice of a test-statistic is an important aspect of a test's design, for it can determine how efficiently the data are used. For example, we expect that the s-statistic developed by Slatkin & Maddison (1989) will discard information about deeper population divergences, and hence will not yield tests of high power for some questions.

A test having low power could do more than frustrate our ability to detect a pattern. If the alternate scenarios are branching population trees, the power to distinguish among them may differ depending upon specific branching patterns of the population tree and the specific test statistic that is used. Consequently, difficulties with effectively testing some hypotheses could bias the apparent relative frequency of some historical processes. However, the potential for such a bias can be addressed for any specific study by comparing the rejection rate of the null hypothesis when it is false (i.e. the power of the test) (Huelsenbeck & Crandall 1997).

The fact that different test statistics may extract information relevant to different questions highlights the need to consider carefully what information could recover the signal of multiple and varied processes. For this reason NCA examines specific measures of the geographical distribution of haplotypes and clades that are expected to yield different values under the various possible processes it considers. However, even if a test-statistic might be capable of distinguishing among alternative explanations for patterns of genetic variation, resolution of specific demographic events may be limited, as may temporal delineation of sequential events during a species history. Only to the extent that past events leave a signature on the gene tree can these historical processes be identified, where the impact on the genealogy will depend on the duration and magnitude of the event (e.g. the length and severity of a bottleneck). Furthermore, much of the history in a genealogy reflects events in the distant past (Barton & Wilson 1995). Consequently, the genealogy should not be interpreted as a precise record of a species' demographic history. Likewise, for species with fairly complex histories, the signature of these past events will not necessarily be delineated as discrete, orderly partitions across the gene tree, or disjunct temporal sequence of past events.

Depending on the historical process being inferred, the domain in which the structure of the gene tree itself may be informative also varies. For example, population subdivision can have dramatic effects on the shape of trees, whereas changes in population size affect tree length by making the coalescence rate time dependent but do not influence the topology of trees (Hudson 1990). Furthermore, integrating over gene trees rather than using a reconstructed gene tree itself may be more conducive to phylogeographic inference under certain circumstances (Wakeley 2002). For example, when populations conform to the standard coalescent model, the effects of random genetic drift dominate and the branching patterns of a gene tree does not contain much information about the population history (e.g. Kuhner *et al*. 1998; Beerli & Felsenstein 2001; Wakeley 2002).

Multiple processes and complex histories

Each of the examples described above focuses on a limited number of possible processes. None of the examples simultaneously consider population divergence, changes in population size, ongoing gene flow, occasional long distance migration, and so on. In principle, however, there is no reason that elaborate historical scenarios involving multiple processes could not be studied with similar methods.

Since a primary goal of a phylogeographic study is to yield an inferred population history, models with rather restrictive assumptions obviously may be overly simplistic (Nielsen & Slatkin 2001), or unable to test specific historical

example test-statistics	references
segregating sites	Wakeley (2001); Wakeley & Hey (1998)
s-statistic	Slatkin & Maddison (1989); Knowles (2001b)
number of deep coalescents	Maddison (1997)
number of rare alleles	Bertorelle & Rannala 1998
distribution of pairwise differences	Slatkin & Hudson (1991)
lineage-through-time plots	Nee *et al*. (1995)
D_c, D_n and *I-T*	Templeton *et al*. (1995)
intra-allelic variability	Slatkin & Bertorelle (2001)

Table 1 Some examples of different test-statistics used in phylogeographic inferences in which theoretical expectations are derived analytically or by simulation for comparison with an observed value from the data

scenarios if they do not consider patterns of genetic variation in an explicit geographical context (Wakeley 2001). Moreover, even if a study's focus is to estimate a simple population genetic parameter and not the phylogeographic history of populations, attention to the complexity of history may be necessary. Any parameter estimate, no matter how precise, is only useful to the extent that the historical process matches the model on which the estimate was based (Wakeley & Hey 1998; Nielsen & Slatkin 2000; Wakeley 2001). Accurate historical inferences therefore require thoughtful consideration of the assumptions, that is, the fixed parameters in such models. For example, the time of population divergence estimated from a model that assumed there was no migration could be very misleading if the populations did indeed exchange migrants at any appreciable rate. Similarly, a method may fail to provide a realistic estimate if the ancestral population was highly subdivided when it was assumed to be panmictic.

Deciding how many and what parameters are included in a model, as well as which are fixed or allowed to vary, is no trivial matter, and such decisions should not be dictated arbitrarily by model availability. However, in designing tests of historical scenarios, it is not enough to have flexibility in specifying the historical scenario and a rich model of population genetics processes. While it is desirable for methods to accommodate (as does NCA) a diverse array of processes, the costs associated with increased versatility may offset any potential gains (e.g. Wakeley & Hey 1998). With increased model complexity, not only do more parameters have to be estimated, but the utility of complex models is also limited by the extent to which the models' expectations differ, such that alternative hypotheses can be distinguished statistically. Alternative histories must make different predictions about sampled genetic variation for a successful statistical phylogeographic analysis.

There are several difficulties in undertaking this inference. Even if an efficient statistical approach for extracting information from the data has been identified, an enormous number of alternative histories could be considered. The inferred history may claim, for example, that a large ancestral population fragmented into three, of which one underwent a severe bottleneck, and the remaining two continued to exchange occasional migrants. Obviously, many other histories are conceivable, involving various combinations of processes.

In principle, one could imagine a computer program sifting through many alternative population histories involving varied processes, similar to a tree search in phylogenetics. However, in practice the number of alternative histories will be enormous, and the search difficult; hence the limited number of alternative explanations statistically explored in the examples we presented above.

Secondly, the choice of histories cannot be based solely on the fidelity with which a hypothetical history predicts the observations. The complexity of a hypothetical history must also be used as a criterion in the inference procedure, with simpler hypotheses preferred. Otherwise, trivial solutions will be found by proposing terribly complex histories. For example, for any particular gene tree, a population history that would very precisely predict the observed data would consist of a branching tree of populations of extremely small effective size, such that the population tree and gene tree had the same form (i.e. one gene copy per very small population). While this population tree might very precisely predict the gene tree, it would be an unnecessarily complex hypothesis, and choosing it would be like choosing a tenth order polynomial for bivariate data. Such a complex model with many instances of fluctuation, migration and divergence, would soon have more parameters than the number of data points available.

The challenge of seeking the simplest population history that best explains the data is daunting and simply may not be feasible in some cases. An alternative approach would be to reduce the scope of the inference itself by using information that is unrelated to the genetic data to generate the models for hypothesis testing. For example, palaeoclimatic and geotectonic information (e.g. Kidd & Ritchie 2000), or even comparative phylogeographic patterns themselves (e.g. Hugall et al. 2002), might be used to derive an explicit historical model. At the very least, this information and/or data on the species, ecology might be useful in constraining the universe of potential population histories to be explored (Knowles 2001b). Different criteria can then be used to judge among alternative histories. For instance, a proposed population history would confer probabilities on various possible gene trees and gene sequences. Maximum likelihood could therefore be used — the population history inferred being that which maximized the probability of the observed data (Huelsenbeck & Crandall 1997). This would allow a fully statistical approach and perhaps allow the specification of confidence limits on the inferred population histories. Various methods might also be used in a complementary fashion, potentially increasing the accuracy of historical inferences. For example, before the parameters of interest are estimated, a statistical phylogeographic method could initially be used to evaluate the relative probabilities of different potential histories, thereby determining which parameters might realistically be fixed (e.g. determine the likelihood of a single vs. multiple ancestral populations). The accuracy of the parameter estimates derived from subsequent analyses, such as joint likelihood of divergence time and migration rate, will most likely be greater than if the model's assumptions had gone unchecked (i.e. specific parameters arbitrarily fixed).

By constraining the universe of potential histories that we might search, we can not be assured of finding the history that best fits the data. Nonetheless, rejection of alternative histories can still provide important insights, as our

examples, as well as many others demonstrate (e.g. Avise & Ball 1990; Hudson 1990; Slatkin & Hudson 1991; Barton & Wilson 1995; Takahata *et al*. 1995; Kuhner *et al*. 1998; Edwards & Beerli 2000; Wakeley 2001). There is definitely room for improvement, particularly with respect to three areas that are vital to the development of phylogeography and historical studies of speciation: (i) parameterization of our models; (ii) strategies for searching through alternative histories; and (iii) criteria for judging the adequacy of their explanations of the data.

The importance of assessing error

One of the primary messages of this paper is that inferences involving gene histories, which are so strongly influenced by chance events, should carefully consider the probable error in the inference. Since the confidence limits are indeed expected to be very broad given the high variance of the coalescent process (Donnely & Tavaré 1995; Hudson 1990), phylogeographic interpretation needs to be made with considerably more caution than has to date been fashionable.

Inferences, especially those about recent events, should be made cautiously since information from any one genealogy may be misleading (Barton & Wilson 1995; Maddison 1997; Rosenberg 2002). If the estimated gene tree itself is used in subsequent analyses (e.g. Bahlo & Griffiths 2000; some examples presented herein), as opposed to integrating over all possible trees (e.g. Kuhner *et al*. 1998; Beerli & Felsenstein 2001), uncertainty in the gene tree should also be considered. Increasing the number of loci is also critical to decreasing the uncertainty in estimates. Not only can multilocus data increase the accuracy of parameter estimates where the summary properties of gene trees are inferred over many loci (Arbogast *et al*. 2002; Kuhner *et al*. 1998; Wakeley & Hey 1998;), but reference to multiple loci are required to address evolutionary questions based on comparative analyses among species. For example, differences in the ancestral effective population sizes or mode of divergence may account for the variation in the geographical distribution or timing of divergence among codistributed species, without invoking differences in the species' reponses to a shared common history. Reference to multilocus data is integral to evaluating hypotheses regarding the temporal and biogeographical congruence of species or population divergence because it provides a framework for considering multiple sources of variance in an estimate, including the stochastic genealogical component, or variability among species attributable to the coalescent process (Arbogast *et al*. 2002; Edwards & Beerli 2000; Wakeley 2002).

Although we have portrayed how, in principle, statistical phylogeography could consider complex histories with many processes involved, such a method does not yet exist.

We recognize that this claim seems to be at odds with statements in the literature regarding Templeton's NCA (e.g. Templeton 1998a, 1998b, 2002a, 2002b; Cruzan & Templeton 2000). In fact, Templeton recently commented on NCA as using objective and quantifiable 'statistical inferences or criteria' (Templeton 2002a, 2002b, respectively), but it is important to consider exactly what aspects of NCA involve statistical inference. It is true that the test for population structure is statistical, and that the rejection of the null model of no geographical structure can be made with statistical confidence. However, the inference key goes far beyond rejection of this null model, and selects a particular process to which the pattern is attributed. This selection of a particular process, over others, is not done statistically. The fact that one can conclude statistically that there must be a process generating a pattern does not say what process in particular generated that pattern. For example, while there are 'statistically significant inferences' in Templeton's new analysis of recent human evolution (Fig. 1, Templeton 2002a) these do not pertain to particular biological processes, such as the two major range expansions that were inferred to have occurred after an original range extension of *Homo erectus* out of Africa. Rather, the statistical inference only applies to the more basic question of whether there is some association between genetic variation and geographical locality. Thus, NCA does not distinguish statistically among restricted gene flow, long distance migration, past fragmentation or range expansion, or any other processes, even though the method is often used as a means to distinguish among these processes.

Nested cladistic analysis has recently increased in popularity, which may be due in part to the notion that it does not make *a priori* assumptions about the past or that it is essentially model free (Templeton 1998a,b). Does NCA then fill the role desired of a flexible, statistical inference procedure?

There are several difficulties in assessing NCA, the first of which is that it is not analytically derived from an explicit, stochastic model. That is, it has not been proven mathematically that under some model of evolution the full NCA method (including inference key) would result in inferences with desirable statistical properties (e.g. powerful, unbiased). An alternative means of assessment would be to simulate evolution under various historical scenarios to see how accurately NCA could recover the known history. Unfortunately, the fact that a key step of NCA — the inference key — must currently be executed by hand makes testing the method extremely labour-intensive. Nevertheless, we have attempted on a small-scale, such a study.

Simulation Study. In order to explore the behaviour of NCA, the accuracy of historical inferences derived from NCA (i.e. data analysed using TCS, Clement *et al*. 2000;

GeoDis, Posada *et al.* 2000; Templeton's inference key, Templeton *et al.* 1995) was evaluated using computer simulations. Ten genealogies were simulated by a neutral coalescent process ($N_e = 10\ 000$) under a history of allopatric divergence involving two vicariant events, an initial split at *t*1 followed by subsequent splits in each population lineage at *t*2 (Fig. 5). These genealogies were then used to evolve 1000 base pairs using a model of among-site rate heterogeneity ($\alpha = 0.3$; ts: tv = 3) assuming all sites were potentially variable. The 10 data matrices were each analysed with NCA (simulation details, access to data matrices, and NCA analyses of individual data sets are available at http://mesquiteproject.org/knowles). A limited number of data sets were analysed because of the lack of automation for NCA.

To evaluate how well NCA accurately reconstructed the actual history of allopatric divergence (Fig. 5a), a number of categories were used to describe the performance of NCA. Of particular interest is: (i) the proportion of times NCA correctly identified a significant pattern; (ii) when a significant pattern was identified, was allopatric divergence vs. some other process inferred; and (iii) how frequently was the stochastic process of lineage sorting misidentified as some deterministic process. Following the convention of NCA in which permutation tests of significant structure are performed as a series of nested, hierarchical tests, all the results are reported by nesting level (i.e. separately for 1-step, 2-step and 3-step clades) (Fig. 5b).

Significant genetic structuring was detected in the majority of data sets by NCA (Table 2). However, there were two critical problems with this initial part of the analyses. First, the nested clades that were identified as having significant phylogeographic signal by the permutation tests did not correspond to the actual geographical configuration or historical association of the populations. For example, even though statistically significant structure was identified for 3-step clades in nine of the 10 data sets, only in 30% of the data sets did the reconstructed clades by

NCA correspond to a clade containing the majority of haplotypes from populations A and B, and another with haplotypes from C and D (Fig. 5b). The second problem involved the detection of significant structure, and hence inference of plausible biological causation, at the 1-step clade level when no such structure existed in the model used to simulate the data (i.e. there was no geographical substructure within any of the four populations A, B, C or D). This indicates that NCA repeatedly misidentified the stochastic process of lineage sorting, and instead, incorrectly assigned a deterministic process to patterns of genetic variation in 70% of the data sets. These problems raise the question of the biological relevance of defining clades in NCA using the one-step mutational criterion. Irrespective of appeals to avoiding *a prior* assumptions about the hierarchy among the populations, this convention tends to place too much emphasis on what are arbitrarily defined clades that may, or may not extract genealogical information pertinent to a species history. Whatever the underlying cause(s) for these problems, the simulations suggest a study of the criteria used to identify and evaluate significant phylogeographic structure in NCA is needed.

We also scored whether allopatric divergence was inferred by NCA, the only biological process generating structure in the simulated data sets. While NCA detected significant genetic structuring in almost all the data sets, it rarely interpreted allopatric divergence as the underlying cause (Table 2b). Furthermore, of the 30 separate nesting levels in which significant population structure was identified, 30% of the time NCA ascribed this genetic structuring to recurrent gene flow, as opposed to any of the historical processes (e.g. fragmentation, long-distance colonization, allopatric divergence, population expansion) (Table 2b). Thus, despite its reputation (e.g. Templeton 1998b; Emerson *et al.* 2001; Hare 2001), NCA did not effectively distinguish between recurrent gene flow and historical processes, and in most cases, it did not accurately distinguish among historical factors.
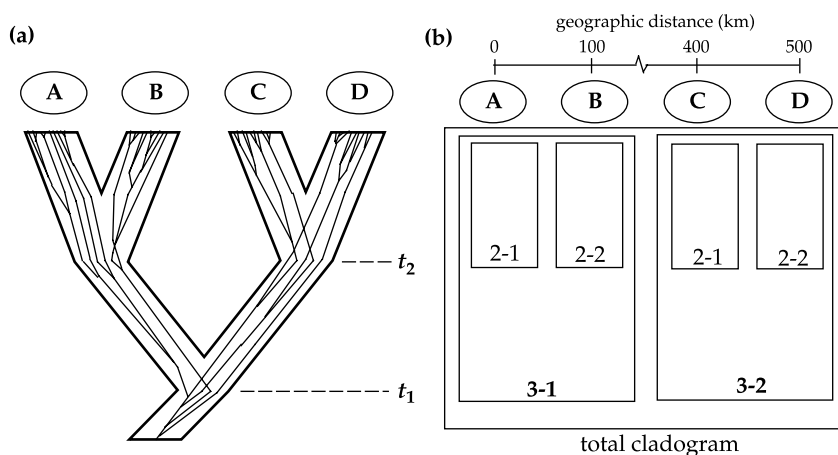

Fig. 5 History of allopatric divergence used in the simulations represented as (a) a population tree, and (b) a nested cladogram (haplotypes are not shown). The geographical distances shown between populations were those used in NCA.

Table 2a Results from permutation tests of the null hypothesis that the clades are randomly distributed geographically relative to other clades within the same nesting clade. NCA results are summarized across the 10 data sets and reported separately for each of the nesting clades

NCA results	3-step clades* nested in entire cladogram	2-step clades† nested within clades 3–1 & 3–2	1-step clades‡ nested within clades 2–1 & 2–2
significant geographical association detected	90%	65%	70%
clade-groupings accurate	30%	35%	na

Table 2b Results from inference key summarized across the 10 data sets and reported separately for each of the nesting clades. The first value represents the percentage of the 10 data sets in which a particular inference was made, whereas the one in parentheses was calculated using only the subset of the data for which significant population structure was identified by the permutation tests

results from inference key	3-step clades* nested in entire cladogram	2-step clades† nested within clades 3–1 & 3–2	1-step clades‡ nested within clades 2–1 & 2–2
allopatric divergence inferred	0%	0.5%	7.7%
	(0%)	(7.7%)	(14.3%)
recurrent force inferred (e.g. gene flow)	20%	25%	15.4%
	(22%)	(38.5%)	(28.6%)
historical process other than allopatric divergence inferred	80%	74.5%	76.9%
	(78%)	(53.8%)	(57.1%)

*corresponds to the two major lineages defined by the initial vicariance at t_1
†corresponds to the two subgroups within each of the major lineages defined by the subsequent vicariance at t_2
‡there is no geographical structure in the model used to simulate the data that corresponds to these clades.

These results obviously do not, nor were they intended to, exhaustively explore the accuracy of NCA over different parameter values. However, there is no reason to suspect that something particular to the fairly simple, and probably common, history of allopatric divergence modelled here (Fig. 5) would cause NCA to fail. This small study does illustrate one of the primary problems with NCA, which is that it does not adequately consider the stochastic nature of the coalescent process when making historical estimates.

To demonstrate another limitation of NCA, that it does not provide a framework for (i) evaluating the statistical support for any specific event, nor (ii) a means for statistically distinguishing between alternative processes, we analysed the same 10 data sets discussed above using a coalescence-based method. This comparison is not intended to evaluate which method was better at inferring the phylogeographic history. Indeed, such a comparison would be unfair since NCA considers a broader array of processes than we consider in our analyses. Rather the objective of this exercise is to demonstrate how important it is that a method based on theoretical expectations derived from the coalescence of gene lineages account for the stochastic nature of the process.

For this exercise, we considered whether we could reject statistically two historical processes, namely population fragmentation (Fig. 6a) and colonization with isolation by distance (Fig. 6b). These choices are obviously arbitrary, but they are both histories considered by NCA.

Following the procedure discussed with the *Melanoplus* grasshopper example earlier, we first derived theoretical expectations for the amount of discord (Fig. 2) that would be observed between a genealogy and the population tree representing a history of allopatric divergence, when the
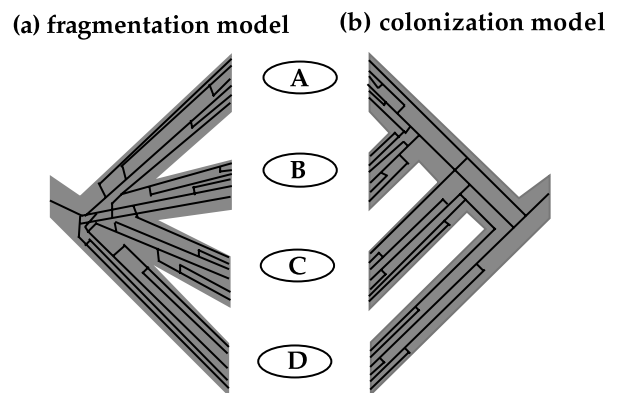
(a) fragmentation model   **(b) colonization model**

Fig. 6 A history of fragmentation (a) and colonization with isolation by distance (b) modelled by coalescent simulations of gene trees within the respective population trees.

data had actually evolved under either a history of fragmentation or colonization. The amount of discord was quantified as the number of deep coalescents (i.e. the failure of gene lineages to coalesce within the respective population lineages) because gene flow was not included in the models.

To generate an expected distribution for the test statistic, one hundred gene trees were simulated by a neutral coalescent process (N_e = 10 000) under a history of fragmentation and colonization (Fig. 6) for each of the 10 data sets using the program MESQUITE (Maddison & Maddison 2000). Then the number of deep coalescents between these simulated gene trees and the population tree representing a history of allopatric divergence (Fig. 5a) was quantified, and used to derive an expected distribution for evaluating the hypothesis that the reconstructed gene trees (i.e. genealogies reconstructed by maximum-likelihood for each of the 10 simulated data matrices) had evolved under either a history of fragmentation or colonization. The number of deep coalescents between the reconstructed gene tree and the model of allopatric divergence (Fig. 5a) was individually quantified, and compared to the expected distributions for the number of deep coalescents (Fig. 7). The null models of fragmentation and colonization were rejected when the observed value was less than 95% of those calculated from the simulated gene trees (i.e. $P < 0.05$). For the 10 analysed data sets, the history of allopatric divergence was accepted over the fragmentation and colonization histories only 40 and 30% of the time, respectively.

The poor performance of the tests suggests that the number of deep coalescents was not a particularly powerful test-statistic, and that some other statistic might have better extracted information pertinent to tests of these particular phylogeographic hypotheses. However, while the coalescent simulations, like NCA, failed to identify allopatric divergence in the majority of data sets, there are fundamental differences between the failures of NCA and the coalescent simulations. When the coalescent method 'failed', the conclusion was that we were unable to distinguish among alternative phylogeographic hypotheses, whereas with NCA, since there was no means for evaluating the
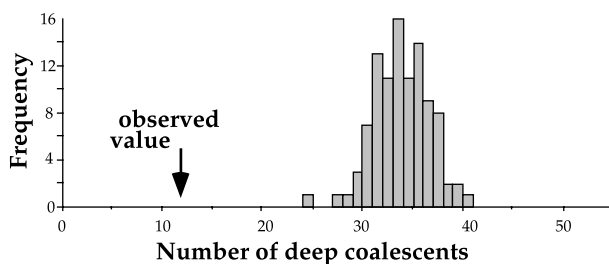
statistical confidence of any conclusions derived from the inference key, an incorrect process was consequently inferred (Table 2b). We are not questioning NCA's test-statistics that are based on the geographical distribution of haplotypes and clades (e.g. D_c, D_n, and I-T-values) and which most likely do capture information relevant to many phylogeographic hypotheses. However, until the biological insights represented in the inference key are transferred to a more fully statistical model that recognizes the stochasticity of the processes involved, the simulations indicate that conclusions from NCA may very well be misleading.

One might argue that the simulations were too simple, but if NCA can't recover simplicity, can we hope that it will recover complexity? Similarly, one might argue that the performance of NCA might be evaluated with empirical examples by comparing the detailed history recovered from the inference key to those processes inferred from independent criteria, such as information based on biogeographical or climatic data, rather than by simulations. To date, this comparison has not been conducted. While frequently cited as 'the' test of NCA's accuracy and validity (e.g. Templeton 2002a,b), Templeton (1998a) did not test whether the multiple processes inferred for any individual species' histories were indeed accurate. Rather, the survey of empirical studies tested if NCA was able to accurately reconstruct a single process, that of range expansion. Appeals to this simple test as validation for NCA as a method that can effectively and accurately estimate complex histories involving many processes would seem insufficient. Therefore, it remains yet to be determined whether NCA can indeed accurately infer 'the combination of factors that best explains the current distribution of genetic variation (Templeton 1998a)'.

Conclusions

The field of phylogeography has expanded rapidly, accumulating vast quantities of data on a great diversity of organisms (Stone 2000). This wealth of information has spurred the need for methods to statistically test hypotheses generated by these studies. Parameter estimation methods from population genetics are well developed statistically, but pay little attention to the possibility of more complex phylogeographic histories. Nested cladistic analysis is rich in the processes and histories it considers as well as the biological insight it embodies, but it conveys no information about the error in its inferences, nor does it offer an assessment of population histories other than those inferred. What is needed is a fusion of these two traditions, yielding a statistical approach to phylogeography that promotes the generation of explicit evolutionary models involving geography and history while providing the tools to reject hypotheses.

Fig. 7 Example of an expected distribution for the number of deep coalescents that was derived by simulations.

The tremendous variety of historical scenarios that can be considered by such approaches makes them broadly applicable, and their ability to incorporate biological information (e.g. climatological data that bears on past species distributions) is also a highly desirable attribute. However, it is imperative that a foundation for choosing a specific model, especially with regards to the number of parameters used in a model, as well as a means for searching among and evaluating the many possible alternative hypotheses be developed. Ideally these developments will also provide a framework for multiple loci and considerations of uncertainty in the reconstructed gene tree.

How well phylogeographic methods succeed in inferring population histories remains for the future to discover. The success of phylogenetic biology at inferring evolutionary histories at deeper time scales is encouraging, but there is no guarantee that that will translate into an ability to infer histories at shallow time scales, within and among populations. It is vital that more effort is put into understanding the errors in our phylogeographic inferences.

Acknowledgements

References

Arbogast B, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic time scales. *Annual Review of Ecology and Systematics*, in press.

Avise JC (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**, 1192–1208.

Avise JC (1994) *Molecular Markers, Natural History, and Evolution*. Chapman & Hall, London.

Avise JC, Ball RM (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 43–67.

Bahlo M, Griffiths RC (2000) Inference from gene trees in a subdivided population. *Theoretical Population Biology*, **57**, 79–95.

Barton NH, Wilson I (1995) Genealogies and geography. *Philosophical Transactions of the Royal Society of London B*, **349**, 49–59.

Beerli P, Felsenstein J (1999) Maximum likelihood estimation of a migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

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

Beerli P, Hotz H, Uzzell T (1996) Geologically dated sea barriers calibrate a protein clock for Aegean water frogs. *Evolution*, **50**, 1676–1687.

Bertorelle G, Rannala B (1998) Using rare mutations to estimate population divergence times: a maximum likelihood approach. *Proceedings of the National Academy of Sciences of the USA*, **95**, 15452–15457.

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 167–1660.

Cruzan MB, Templeton AR (2000) Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution*, **15**, 491–496.

Donnely P, Tavaré S (1995) Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.*, **29**, 401–421.

Edwards SV (1993) Mitochondrial gene genealogy and gene flow among island and mainland populations of a sedentary songbird, the Grey-crowned Babbler, *Pomatostomus temporalis*. *Evolution*, **47**, 118–1137.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Emerson BC, Paradis E, Thébaud C (2001) Revealing the demographic histories of species using DNA sequences. *TREE*, **16**, 707–716.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Griffiths RC (1998) GENETREE. *Available from* http:// http:// www.maths.monash.edu.au/~mbahlo/mpg/gtree.html.

Hare MP (2001) Prospects for nuclear gene phylogeography. *TREE*, **16**, 700–706.

Harrison R (1991) Molecular changes at speciation. *Annual Review of Ecology and Systematics*, **22**, 281–308.

Hudson R (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 1–44.

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, **28**, 437–466.

Hugall A, Mortiz C, Moussalli A, Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail *Gnarosophia bellendenkerensis*. *Proceedings of the National Academy of Sciences of the USA*, in Press.

Kidd DM, Ritchie MG (2000) Inferring the patterns and causes of geographic variation in *Ephippiger ephippiger* (Orthoptera, Tettigoniidae) using geographical information systems (GIS). *Biological Journal of the Linnean Society*, **71**, 269–295.

Kingman JFC (1982) The coalescent. *Stochastic Process. Appl.*, **13**, 235–248.

Kishino H, Thorne JL, Bruno WJ (2001) Paeformance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution*, **18**, 352–361.

Klicka J, Zink RM (1997) The improtance of recent ice ages in speciation: a failed paradigm. *Science*, **277**, 1666–1669.

Kliman RM, Andolfatto P, Coyne JA *et al.* (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics*, **156**, 1913–1931.

Knowles LL (2000) Tests of Pleistocene speciation in montane grasshoppers from the sky islands of western North America (genus *Melanoplus*). *Evolution*, **54**, 1337–1348.

Knowles LL (2001a) Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. *Proceedings of the Royal Society of London Series B*, **268**, 1–6.

Knowles LL (2001b) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, **10**, 691–701.

Knowles LL (2003) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, in press.

Knowles LL, Futuyma DJ, Eanes WF, Rannala B (1999) Insights into speciation mode from historical demography in the phytophagous beetle *Ophraella. Evolution*, **53**, 1846–1856.

Kuhner M, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Maddison W (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Maddison WP, Maddison DR (2000) MESQUITE: a modular programming system for evolutionary analysis. *Prototype* v.0.9d17. http://mesquiteproject.org.

Maddison W, McMahon M (2000) Divergence and reticulation among montane populations of a jumping spider (*Habronattus pugillis* Griswold). *Systematic Biology*, **49**, 400–421.

Masta SE, Maddison WP (2002) Sexual selection driving diversification in jumping spiders. *Proceedings of the National Academy of Sciences of the USA*, **99**, 4442–4447.

Mengel RM (1964) The probable history of species formation in some northern wood warblers (Parulidae). *Living Bird*, **3**, 9–43.

Milot E, Gibbs HL, Hobson KA (2000) Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). *Molecular Ecology*, **9**, 677–681.

Nee S, Holmes EC, Rambaut A, Harvey PH (1995) Inferring population history from molecular phylogenies. *Philosophical Transactions of the Royal Society of London Series B*, **349**, 25–31.

Nielsen R, Slatkin M (2000) Likelihood analysis of ongoing gene flow and historical association. *Evolution*, **54**, 44–50.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Riddle BR (1996) The molecular phylogeographic bridge between deep and shallow history in continental biotas. *Trends in Ecology and Evolution*, **11**, 187–228.

Rosenberg NA (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, **61**, 225–247.

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution*, **14**, 1218–1231.

Schneider CJ, Cunningham M, Moritz C (1998) Comparitive phylogeography and the history of endemic vertebrates in the wet tropic rainforests of Australia. *Molecular Ecology*, **7**, 487–498.

da Silva MNF, Patton JL (1998) Molecular phylogeography and the evolution and conservation of Amazonian mammals. *Molecular Ecology*, **7**, 475–486.

Slatkin M (1987) The average number of sites separating DNA sequences drawn from a subdivided population. *Theoretical Population Biology*, **32**, 42–49.

Slatkin M, Bertorelle G (2001) The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics*, **155**, 865–874.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–613.

Stone G (2000) Phylogeography, hybridization and speciation. *Trends in Ecology and Evolution*, **15**, 354–355.

Takahata N (1989) Gene genealogy in 3 related populations — consistency probability between gene and population trees. *Genetics*, **122**, 957–966.

Takahata N (1991) Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. *Genetics*, **129**, 585–595.

Takahata N (1993) Allelic genealogy and human evolution. *Molecular Biology and Evolution*, **10**, 2–22.

Takahata N, Satta Y (1997) Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proceedings of the National Academy of Sciences of the USA*, **94**, 4811–4815.

Takahata N, Satta Y, Klein J (1995) Divergence time and population size in the lineage leading to modern humans. *Theoretical Population Biology*, **48**, 198–221.

Tang H, Siegmund DO, Peidong S, Oefner PJ, Feldman MW (2002) Estimation of coalescence times from nucleotide sequence data using a tree-based partition. *Genetics*, **161**, 447–459.

Templeton AR (1998a) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR (1998b) The role of molecular genetics in speciation studies. In: *Molecular Approaches to Ecology* (eds DeSalle R, Schierwater B), pp. 131–156. Birkhäuser-Verlag, Basel.

Templeton AR (2002a) Out of Africa again and again. *Nature*, **416**, 45–51.

Templeton AR (2002b) 'Optimal' randomization strategies when testing the existence of a phylogeographic structure: a reply to Petit and Grivet. *Genetics*, **161**, 473–475.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographic distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma Tigrinum. Genetics*, **140**, 767–782.

Wakeley J (2000) The effects of subdivison on the genetic divergence of populations and species. *Evolution*, **54**, 1092–1101.

Wakeley J (2001) The coalescent in an island model of population subdivision with variation among demes. *Theoretical Population Biology*, **59**, 133–144.

Wakeley J (2002) Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. In: *The Evolution of Population Biology — Modern Synthesis* (eds Singh R, Uyenoyama M, Jain S), Cambridge University Press, Cambridge, in press.

Wakeley J, Hey J (1998) Testing speciation models with DNA sequence data. In: *Molecular Approaches to Ecology* (eds DeSalle R, Schierwater B), pp. 157–175. Birkhäuser-Verlag, Basel.