doi:10.1046/j.1420-9101.2003.00644.x

MINI REVIEW

The burgeoning field of statistical phylogeography

L. L. KNOWLES

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Keywords:
biogeography;
bottlenecks;
coalescence;
genealogy;
genetic parameters;
historical demography;
phylogeography;
speciation;
vicariance.

Abstract

In the newly emerging field of statistical phylogeography, consideration of the stochastic nature of genetic processes and explicit reference to theoretical expectations under various models has dramatically transformed how historical processes are studied. Rather than being restricted to *ad hoc* explanations for observed patterns of genetic variation, assessments about the underlying evolutionary processes are now based on statistical tests of various hypotheses, as well as estimates of the parameters specified by the models. A wide range of demographical and biogeographical processes can be accommodated by these new analytical approaches, providing biologically more realistic models. Because of these advances, statistical phylogeography can provide unprecedented insights about a species' history, including decisive information about the factors that shape patterns of genetic variation, species distributions, and speciation. However, to improve our understanding of such processes, a critical examination and appreciation of the inherent difficulties of historical inference and challenges specific to testing phylogeographical hypotheses are essential. As the field of statistical phylogeography continues to take shape many difficulties have been resolved. Nonetheless, careful attention to the complexities of testing historical hypotheses and further theoretical developments are essential to improving the accuracy of our conclusions about a species' history.

Introduction

Phylogeography is undergoing a fundamental conceptual and methodological shift. Inferences about a species' history are now based on statistical tests of historical hypotheses and estimates of demographical parameters. This contrasts with the phylogeographical tradition of gathering data and then inferring something about the causes of an association (or lack thereof) between the observed patterns of genetic variation and the geographical distribution of populations (reviewed in Avise, 1998). The most pervasive and general difficulty with making inferences about such processes by interpreting a gene-tree unerringly is the potential for pronounced over interpretation of the data that can be extremely misleading (Edwards & Beerli, 2000; Knowles & Maddison, 2002; Wakeley, 2002; Hudson & Turelli, 2003). Consequently, despite the appeal and increasing popularity of methods that make detailed historical inferences, such as nested-cladistic analysis (Templeton *et al.*, 1995), such approaches nonetheless belie the fundamental and well established principles that caution against inferring causation based on resolute interpretation of a gene tree (Table 1) (e.g. Pamilo & Nei, 1988; Takahata, 1989; Hudson, 1990; Maddison, 1997; Avise, 2000; Ray *et al.*, 2003). In fact, the transition from *describing* to *testing* hypotheses about the processes underlying patterns of genetic variation has actually necessitated a concomitant shift in how such historical inferences are made – namely, the explicit consideration of stochastic variance and reference to predictions based on models that are defined *a priori*. Consequently, not only do the procedures of statistical phylogeography differ from earlier traditions, so too do the challenges (Table 2).

The topics discussed in this review highlight the many complex, and often interrelated, issues involved in statistical phylogeography. Rather than providing an

Correspondence: L. Lacey Knowles, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA.
Tel.: 734 763 5603; fax: 734 763 4080; e-mail: knowlesl@umich.edu

Table 1 Some reasons why historical inferences derived directly from a gene genealogy may be inaccurate or misleading.

Potential problems with historical inferences based on the structure of a gene tree
• The stochastic process of lineage sorting produces a discord between the population's and gene's history
• The actual history is obscured by deep coalescence of gene lineages
• The resolution of the genetic marker is insufficient for recovering the population history from a gene tree estimate
• The gene genealogy reflects the action of selection rather than the population's history
• Alternative hypotheses are indistinguishable because of high stochastic variance of trees of independent genes

Table 2 The transition to testing hypotheses caused a fundamental shift in how statistical phylogeographical inferences are made compared with the traditional descriptive approaches, as well as a new set of challenges.

Key components of statistical-phylogeographical tests
Specifying alternative historical scenarios
Challenge: to define hypotheses simple enough that they can be discriminated with the data available, yet still capture the essence of the biologically interesting problem
Deciding on the model's complexity
Challenge: to determine how complex a model can be fit without making overly simplified assumptions that might potentially affect the accuracy of the conclusions
Integrating external information
Challenge: to develop more testable and biologically relevant hypotheses by incorporating external data, including information from other disciplines

in-depth review of specific methods, the goal is to address the various conceptual difficulties of testing historical hypotheses. What emerges are some specific challenges that revolve around three key steps in statistical phylogeography: how to define a set of hypotheses, decide on the model's complexity, and integrate information from external data (e.g. bioclimatic and biogeographical data). An awareness of these difficulties associated with these procedures identifies those aspects of statistical phylogeographical study that require careful attention, thereby providing an invaluable guide to testing historical hypotheses, which is especially critical as the gap between the relative novice and what can be incredibly sophisticated computational methods grows. These challenges also illustrate the current limitations in this nascent field. Nevertheless, methodological advances will no doubt continue to propel the field of statistical phylogeography to its central and integrative position within the evolutionary and ecological sciences (Table 3) (e.g. O'Ryan *et al.*, 1998; Wakeley & Hey, 1998; Kliman *et al.*, 2000; Wall, 2000; Beaumont *et al.*, 2001; Knowles, 2001a,b; Clegg *et al.*, 2002; Gaggiotti *et al.*, 2002).

Table 3 Examples of various hypotheses about the evolutionary and ecological factors shaping patterns of genetic variation and species divergence that can now be tested in a statistical phylogeographical framework.

Potential insights of statistical phylogeography
• Founder events are involved in speciation
• Reproductive isolation is a by product of the gradual accumulation of species differences by genetic drift
• Biogeographical barriers contribute to regional patterns of endemism
• Population differentiation reflects differing selective pressures rather than extrinsic impediments to gene flow
• Regional areas identified by phylogeographical congruence across species reflect areas of historical persistence indicating their importance for conservation
• Historical biogeographical factors are relatively more important than species-specific ecological and behavioral characteristics in structuring genetic variation

Inherent difficulties of statistical phylogeography

Phylogeographical studies are confronted with two major difficulties: the stochasticity of genetic processes (Hudson, 1990), and the potentially complex and varied history of different species (Knowles & Maddison, 2002). Both issues pose specific challenges to testing phylogeographical hypotheses and can have a substantial impact on the accuracy of our interpretations about a species' history.

High stochastic variance

The stochasticity inherent to the coalescence of gene lineages and to the mutational process among unlinked genes has significant consequences for testing historical hypotheses as well as determining the domain of application for different methodological approaches. This high stochastic variance underscores the potential for uncertainty of historical estimates given the broad confidence limits on historical inferences (Hudson, 1990; Kuhner *et al.*, 1998; Arbogast *et al.*, 2002), but also translates into significant difficulties for distinguishing among alternative hypotheses (Edwards & Beerli, 2000; Pritchard *et al.*, 2000; Beaumont *et al.*, 2002; Knowles & Maddison, 2002).

The extent to which this stochastic variance will complicate phylogeographical interpretations depends on both the underlying population processes and the parameter values of such processes (Fig. 1). Consequently, the historical context itself also determines whether a single gene (vs. multiple loci) or the structure of the gene tree (as opposed to integrating over all possible gene trees) would be more or less likely to provide an accurate statistical estimate of a species' history (Wakeley, in press). For example, consider attempts to estimate the timing of species or population

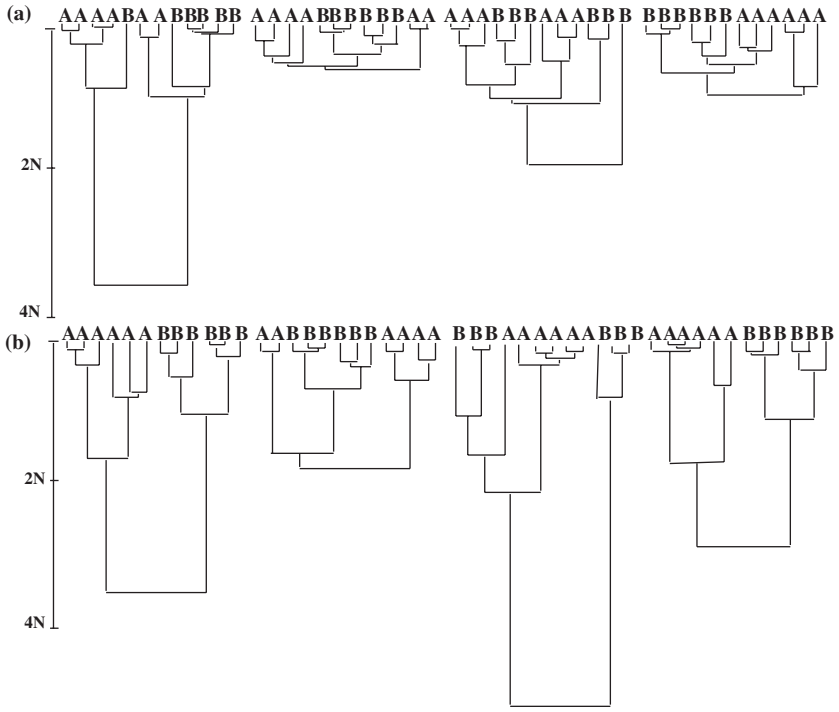

Fig. 1 The eight gene trees were generated using the same model of population divergence for six gene copies sampled from each population (i.e. population A and population B) with 1N gene copies. Coalescent simulations were done at two divergence times: (a) a recent divergence of 0.5N generations ago, and (b) an older divergence of 1.5N generations ago. For a specific time of divergence, the variation among the trees reflects chance alone. These differences illustrate the inherent problems of interpreting gene trees unerringly or using methods that do not take into account the stochastic variance. Comparison between the two sets of trees (a) and (b) also illustrates how specific details of the history (in this case, the parameter τ, the time of population divergence) influence the degree to which this stochasticity complicates historical inference, including estimates of the time of divergence as well as the process of allopatric divergence without gene flow.

divergence, τ, from estimates of gene divergence, T (e.g. Edwards & Beerli, 2000). Population divergence times can be significantly overestimated (Fig. 2) when ancestral polymorphism is ignored in a model, especially when coalescence times are lengthened in the ancestral population by factors such as population substructure or asymmetric migration (Wakeley, 2000, 2001; Beerli & Felsenstein, 2001). The degree of overestimation is much greater for recent divergences compared with older ones (Fig. 2) because the discrepancy between the gene (T) and population (τ) divergences because of the inherent stochasticity of genetic processes represents a greater proportion of the estimated divergence time. With data from a single gene, it is not possible to measure the degree of over estimation, whereas with multiple loci, the stochasticity of the coalescent can be accounted for, thus improving the accuracy of any estimate of divergence times (Arbogast *et al.*, 2002).

Potentially complex species' histories

Species histories can be as varied as they are complex. Not only might they involve a large variety of processes, such as migration, admixture, isolation by distance,

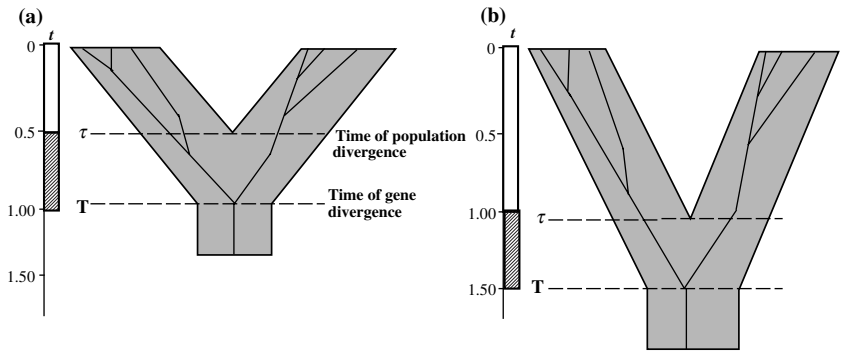

Fig. 2 Overestimates of recent population divergence times because of variance in coalescence times in the ancestral population (see hatched bar). The degree of overestimation is disproportionately larger for recent divergences (a) compared with older divergences (b) because the discrepancy between the gene (T) and population divergence (τ) represents a greater proportion of the estimated population divergence time (i.e. the estimate of the gene divergence is nearly two times greater than the actual population divergence time for the recent divergence, but the degree of the overestimate is substantially less for the older divergence).

divergence in isolation or with gene flow, bottlenecks, and population growth, but these various processes may only occur during specific time periods. This diversity and potentially complex configuration of processes operating at the population level presents a major challenge to testing historical hypotheses (Wakeley & Hey, 1998; Arbogast *et al.*, 2002; Knowles & Maddison, 2002).

Statistical phylogeographical methods that encompass the wide array of potential processes have not yet been developed (Stephens, 2001; Knowles & Maddison, 2002), unlike the heuristic explanations that are applied with approaches that take a gene tree more or less at face value (that is, they do not explicitly consider the stochasticity of genetic processes) (e.g. Avise, 1989; Templeton *et al.*, 1995). Tests of phylogeographical hypotheses therefore can be sensitive to which and how many parameters are included in the model depending on the extent that (a) the models accurately represent the history and (b) the alternative hypotheses can be distinguished statistically with the available data (Nielsen & Slatkin, 2000; Rosenberg & Nordborg, 2002; Stephens, 2001; Beaumont *et al.*, 2002). For example, mitochondrial (mtDNA) sequence data was collected to address the question of whether or not anatomically modern humans and Neanderthals interbred (Kringes *et al.*, 1997). To actually distinguish between these alternative hypotheses, the data would have to be able to distinguish between a model in which there was ancient gene flow vs. one in which there was none. Reconstruction of the mtDNA genealogy showed that the modern-human mtDNAs were monophyletic and quite distinct from the Neanderthal sequence (Fig. 3). However, the mtDNA gene tree is not a phylogenetic tree of Neanderthals and humans – the question is whether such a tree would exist under a history with interbreeding or one of isolation. Nordborg (1998) demonstrated that the mtDNA data were also consistent with a history with substantial interbreeding (e.g. exchanges of 25%). Thus, it was not until the question was modelled with well-constructed hypotheses and tested using coalescent simulations that it became apparent that the issue of interbreeding was beyond the scope of the available data.

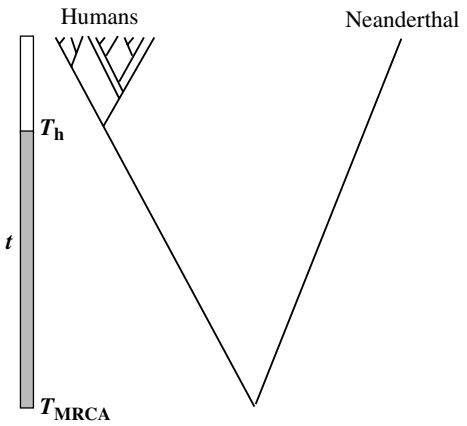

Fig. 3 Schematic genealogy of modern-human mtDNAs and a single Neanderthal mtDNA. The inferred most recent common ancestor (MRCA) of the entire sample (T_{MRCA}) was inferred to be more than four times greater than the MRCA of the human sample (T_h) (see Nordborg, 1998). Because of the amount of genetic drift taking place during the time interval *t*, the available data could not be used to distinguish between the alternative hypotheses of some vs. no interbreeding.

Complex histories in particular are also difficult to test. Specific events may not leave a genealogical signature (Hudson, 1990; Avise, 2000). Even assuming the predicted patterns of genetic variation differ substantially such that the alternative scenarios could be discerned, because of the number of parameters involved, tests of complex histories require large amounts of data and can be difficult to implement (Rannala & Mountain, 1997; Wakeley & Hey, 1998; Edwards & Beerli, 2000; Nielsen & Slatkin, 2000; Pritchard *et al.*, 2000; Beerli & Felsenstein, 2001; Stephens, 2001; Rosenberg & Nordborg, 2002; Takahata & Satta, 2002; Brumfield *et al.*, 2003).

Contending with the complexities of statistical phylogeography

While a statistical framework may permit hypothesis testing and parameter estimation, it nonetheless does not exclude the possibility that the method used to make the inference significantly affected the *reliability* and *usefulness* of a study's conclusions (Penny *et al.*, 1995; Rousset & Raymond, 1997; Huelsenbeck & Imennov, 2002). However, the negative consequences that can arise from the methodological procedure itself can be avoided, or at least minimized. This requires recognizing not only the key steps involved in testing phylogeographical hypotheses, but also the challenges associated with each component of the procedure (Table 2). As discussed in detail below, these include: (a) specifying alternative historical scenarios, (b) deciding on the model's complexity and (c) integrating external information into the statistical phylogeographical test.

Specifying alternative historical hypotheses

Decisions regarding the hypotheses to be tested are as pivotal as methodological decisions about how to conduct the statistical phylogeographical test (e.g. whether to use summary statistics or actually evaluate the likelihood of observing a model given the data). Hypotheses need to be statistically tractable, generating distinct predictions such that the historical scenarios can be distinguished (Knowles & Maddison, 2002; Wakeley, in press). However, the possibility of rejecting a hypothesis must also be tempered by the potential

biological insights afforded by the test. A study's conclusions can be profoundly affected by how and which hypotheses are formulated (Nielsen & Slatkin, 2000; Nielsen & Wakeley, 2001; Ray *et al.*, 2003) and only a small subset of potential histories can typically be considered. At this time, it is not generally feasible to consider the full array of possible histories, or even necessarily the most probable one (Knowles, 2001b; Huelsenbeck & Imennov, 2002).

The challenge is to define hypotheses simple enough that they can be discriminated with the data available, yet still capture the essence of the biologically interesting problem. An overemphasis on statistical formalization without consideration of the biological plausibility of specific models or the consequences of severe oversimplifications can lead to misleading interpretations or those of little substantive value (e.g. Penny *et al.*, 1995; Rousset & Raymond, 1997). Moreover, because both the sampling scheme and types of data collected influence the power of statistical–phylogeographical tests, the hypotheses need to be defined *a priori*. How many individuals per population and the number of loci per individual, as well as the use of DNA sequences vs. polymorphism frequencies, can affect the ability to distinguish among hypotheses and the precision of parameter estimates (Takahata, 1989; Pluzhnikov & Donnelly, 1996; Kuhner *et al.*, 1998; Nielsen & Slatkin, 2000; Arbogast *et al.*, 2002; Wakeley, in press). For example, to test for population admixture, a large number of unlinked loci and representative sampling of the purported parental and admixed populations are necessary (Pritchard *et al.*, 2000; Wall, 2000; Chikhi *et al.*, 2001; Wang, 2003). Consequently, the actual design of a study follows from the specific questions being addressed. This contrasts with the tradition of first collecting data, and then asking what the data mean (e.g. Avise, 1989; Templeton *et al.*, 1995).

Deciding the model's complexity

While the high stochastic variance of genetic processes and typically broad confidence limits on parameter estimates place a premium on seeking powerful approaches to distinguish among hypotheses (Hudson, 1990), it is neither the only, nor is it necessarily, the primary consideration. Species' histories can be complex and involve varying combinations of different processes. Although a method may be computationally eloquent, any test or parameter estimate can only be accurate if the model is a reasonable representation of a species' history (Edwards & Beerli, 2000; Nielsen & Wakeley, 2001; Knowles & Maddison, 2002).

Models with rather restrictive assumptions may be overly simplistic for inferences about natural populations (Hudson, 1990; Nielsen & Slatkin, 2000; Pritchard *et al.*, 2000; Beerli & Felsenstein, 2001; Stephens, 2001), or unable to test specific historical scenarios, especially if they do not explicitly consider the geographical context of population structure and divergence (Wakeley, 2001). Even if a particular method is robust to certain demographical conditions, that is no assurance that the parameter estimate will not be significantly affected if other assumptions are violated. For example, if estimation of a population divergence time is fairly insensitive to demographical events such as strong bottlenecks or rapid population expansion, this issue may nonetheless be irrelevant if the major departure from the model's assumptions is the presence of geographical substructure. A flexible model capable of specifying a large number of historical scenarios with a rich variety of population genetic processes may indeed be desirable. While potentially being biologically more realistic, the utility of a complex model, as discussed earlier, may be offset by the concomitant increases in the amount of data required to distinguish among alternative hypotheses or estimate the additional parameters (Wakeley & Hey, 1998; Beaumont *et al.*, 2002; Knowles & Maddison, 2002), as well as the increased computational requirements that make such tests difficult to implement (Kuhner *et al.*, 1998; Nielsen & Slatkin, 2000; Pritchard *et al.*, 2000; Rosenberg & Nordborg, 2002).

The challenge is to decide how complex a model can be fit to the available data without losing the ability to distinguish among alternative hypotheses or compromising the insights that well-constructed, biologically-meaningful hypotheses can provide (Penny *et al.*, 1995; Rousset & Raymond, 1997). For example, to test historical hypotheses, a model may make assumptions about mutational processes, diverging populations and migration (i.e. the models parameters). Depending on the time scale of interest, mutation may have a negligible effect on the patterns of genetic variation (O'Ryan *et al.*, 1998; Nielsen & Slatkin, 2000). If the divergence time is much smaller than the reciprocal of the mutation rate (i.e. $T \ll 1/\mu$), and the migration rate is substantially larger (i.e. $m \gg \mu$), ignoring mutational input is most likely not only justified, but it significantly reduces computational challenges while also avoiding assumptions about the demographical history of the species before divergence (i.e. assumptions about allele frequencies in the ancestral population) (Nielsen & Wakeley, 2001).

Integrating external information

The potential for parameter estimates and tests of historical hypotheses to be influenced by statistical phylogeographical procedures emphasizes the need to incorporate other sources of data (Romauldi *et al.*, 2002). These inherent difficulties also illustrate why, even without contradictory evidence, inferences based on a limited range of specific features in the genetic data are necessarily weakened (Penny *et al.*, 1995). In

contrast, conclusions that are supported by different methodological approaches and corroborated by information other than the genetic data used in the test, perhaps from external sources such as paleoecological or bioclimatic data (e.g. Kidd & Ritchie, 2000; Gaggiotti *et al.*, 2002; Hugall *et al.*, 2002), are not only more robust, but are also potentially more testable. For example, consistent results from multiple tests about the origin of humans have provided support for the out-of-Africa scenario. These include tests that allow for phylogenetic uncertainty so the conclusion does not depend on any single tree being correct, as well as the integration of diverse types of information into a single analysis, such as when results from other studies are used to construct a 'prior' for a Bayesian analysis (e.g. Huelsenbeck & Imennov, 2002). Similarly, corroboration of inferences about human population structure has also been possible by using a combination of approaches for assigning individuals to populations. These include defining populations *a priori* on the basis of geography (Penny *et al.*, 1995), *a posteriori* group designations based on inferring the most likely number of groups and assigning individuals on the basis of probabilities estimated from a set of independently transmitted loci (e.g. Pritchard *et al.*, 2000), as well as a discrimination analysis in which the most likely geographical origin of individuals are inferred from the genotypes (e.g. Rannala & Mountain, 1997).

In some cases, integration of external information into a statistical framework is essential to testing specific hypotheses (Gaggiotti *et al.*, 2002), including some that address fundamental evolutionary and ecological theories. For example, by comparing the structure of gene trees across a number of different species and reconstructions of the history of fragmentation of the North Queensland rain forest during the Pleistocene based on bioclimatic data, Hugall *et al.* (2002) were able to test if climatic changes affected species similarly, or if species-specific ecological requirements determined how taxa responded to forest fragmentation. Likewise, Wilding *et al.* (2001) used patterns of molecular differentiation in an effort to test what role selection had played in the divergence between populations with two distinct morphs of snails. Comparisons between the observed data and the theoretical expectations identified specific loci that exhibited more differentiation between morphs than predicted by chance alone, suggesting these loci were subject to strong disruptive selection. To confirm that this interpretation was not an artefact of the model used in the statistical phylogeographical procedure, they demonstrated that the same set of highly differentiated loci was consistently divergent across multiple populations and that the patterns of molecular differentiation corresponded to the ecologically divergent morphs and not the geographical locality of individuals.

Brief overview of statistical phylogeographical methods

In contrast to the fairly restrictive set of conditions of classic approaches for parameter estimation, the new methods encompass a wide range of demographical and biogeographical scenarios. These include models with varying population size, asymmetric migration rates, population admixture and structure (e.g. Hudson, 1998; Kuhner *et al.*, 1998; Edwards & Beerli, 2000; Nielsen & Slatkin, 2000; Wakeley, 2000, 2001; Beerli & Felsenstein, 2001; Beaumont *et al.*, 2002), as well as those that explicitly consider the geographical configuration (e.g. Pritchard *et al.*, 2000; Beerli & Felsenstein, 2001; Gaggiotti *et al.*, 2002) or the history of population associations (e.g. Milot *et al.*, 2000; Knowles, 2001b), although the latter are less developed.

Statistical phylogeographical methods generally fall into one of two categories: those that take a summary-statistic approach vs. likelihood analysis of evolutionary models (Table 3). Likelihood analyses (where the probability of observing the data under alternative models is calculated) take advantage of all the information in the data, unlike the summary-statistic approach (where a simple statistic summarizes data) (Felsenstein, 1992). However, evaluating the likelihood of the data for some models is computationally demanding and in some cases intractable. To circumvent the problem of analysing all possible genealogies and allelic configurations (it is not computationally possible to evaluate the likelihood of the data for large sample sizes), Markov chain Monte Carlo (MCMC) and importance sampling (IS) can be used (Stephens, 2001; Rosenberg & Nordborg, 2002). In contrast, summary-statistic approaches are relatively easy to implement, but their interpretations are very sensitive to the defined models. Moreover, very different population histories can produce the same summary statistic (because they do not necessarily take advantage of all the information in the data). For example, a low F_{st}-value could indicate an older population divergence with substantial gene flow, or a recent split with no gene flow where the shared lineages between populations reflect their common ancestry. Similarly, the power to distinguish between hypotheses can be sensitive to how well the summary statistic extracts information relevant to the question of interest (Beaumont *et al.*, 2002).

With methods that rely upon summary statistics (e.g. Hudson, 1998; Wakeley & Hey, 1998; Kliman *et al.*, 2000; Knowles, 2001a,b; Wakeley, in press), data sets simulated by a neutral-coalescent process are typically used to estimate the distribution of an appropriate test statistic under a specific historical model, to which the value of the statistic calculated for the observed data is compared (e.g. Knowles & Maddison, 2002). For example, Wakeley's (in press) test-statistic max p_i, an interlocus measure of concordance with respect to a specific genealogical

split, can be used to statistically distinguish between a model of geographical isolation and a model of divergence with gene flow. Examination of the distributions of max p_i under the respective models generated from coalescent simulations shows how differences in theoretical expectations for this test statistic could be used to test the alternative hypotheses.

Alternatively, estimates of parameters of interest as well as tests of hypotheses can be conducted using likelihood methods where the probability of observing the exact sample configuration is calculated (e.g. Nielsen & Slatkin, 2000; Beaumont *et al.*, 2001; Nielsen & Wakeley, 2001; Gaggiotti *et al.*, 2002; Huelsenbeck & Imennov, 2002; Edwards & Beerli, 2000), while taking into account both sampling and genealogical variation in gene frequencies (Griffiths & Tavaré, 1994; Kuhner *et al.*, 1998). For example, by constraining the migration parameter (i.e. $m = 0$) compared with a model in which m is not constrained, a likelihood-ratio test can be used to determine whether population divergence occurred in isolation or with gene flow while also generating estimates of the time of divergence (e.g. Nielsen & Wakeley, 2001).

In addition to testing hypotheses and estimating parameters, these methods also make it possible to investigate whether a particular history can be expected to have left a trace in the data (Knowles & Maddison, 2002; Rosenberg & Nordborg, 2002; Wakeley, in press). Consequently, not only is it possible to evaluate the feasibility of addressing specific hypotheses (e.g. Pritchard *et al.*, 2000), but the coalescent simulations can also provide useful guidance about how many individuals, populations, and loci need to be sampled to answer the questions of interest (Pluzhnikov & Donnelly, 1996; Wall, 2000; Wakeley, in press).

Choosing a method

No single method possesses all of what might be considered ideal qualities for testing historical hypotheses, namely: provides an accurate representation of the past, considers a diverse array of processes, and still yields a statistical estimate of that history (Stephens, 2001; Knowles & Maddison, 2002). Nevertheless, empiricists are faced with a bewildering array of statistical phylogeographical methods (Table 4). The diversity of methods as well as their computational sophistication can be rather intimidating to the uninitiated. Yet by simply considering the three key steps (Table 2) involved in the statistical phylogeographical approach in concert with

Table 4 Examples of some different programs used in statistical phylogeography*.

Potential programs for parameter estimation and tests of hypotheses
FLUCTUATE (Kuhner *et al.*, 1998) and MIGRATE (Beerli & Felsenstein 1999):
Likelihood-based methods using MCMC to calculate likelihood surfaces of model parameters. FLUCTUATE can estimate exponential expansion or decline for a panmictic population with sequence data. MIGRATE includes models for migration between subpopulations and includes the option of inputting a geographical matrix to incorporate the distribution of subpopulations; it is suitable for both sequence data and microsattelites. http://evolution.genetics.washington.edu/lamarc.html
GENETREE (e.g. Griffiths & Tavaré, 1994):
Likelihood-based method using IS and includes migration and growth rates in structured populations using sequence data, also allows incorporation of geographical information. http://www.stats.ox.ac.uk/mathgen/griff/software.html
BATWING (Wilson *et al.*, 2003):
Flexible Bayesian MCMC method for modelling population divergence, population size and growth, and mutation rates, providing investigators a range of probability distributions to choose priors from; it is suitable for single nucleotide polymorphisms as well as microsatellites. http://www/maths.abdn.ac.uk/ijw/downloads/download.htm
MDIV (Nielsen & Wakeley, 2001):
Likelihood MCMC approach that will simultaneously estimate divergence times and migration rates between two populations and test divergence with gene flow models under the infinite-sites and finite-sites model (HKY). http://www.biom.cornell.edu/Homepages/Rasmus_Nielsen/files.html
BEAST (Drummond & Rambaut, 2003):
Bayesian MCMC analysis of molecular sequences for estimating divergence dates, population size and growth, using flexible models of mutation that can incorporate interlocus differences in substitution processes. http://evolve.zoo.ox.ac.uk/beast/
STRUCTURE (Pritchard *et al.*, 2000):
Bayesian MCMC model-based clustering method for inferring population structure without specifying populations *a priori*, testing models of admixture, estimating admixture proportions, identifying immigrant individuals while also permitting the incorporation of geographical information into the inference process; uses multilocus data, including microsatellites, RFLPs and SNPs. http://pritch.bsd.uchicago.edu/
MESQUITE (Maddison & Maddison, 2000):
Flexible program for testing different evolutionary models of population divergence, including vicariance, fragmentation, and isolation by distance, using coalescent simulations; suited for sequence data. Various summary-statistics are supported or simulations can be exported to other programs for likelihood analysis. http://mesquiteproject.org

*See also SITES (e.g. Wakeley & Hey, 1998), LEADMIX (Wang, 2003), and others listed at http://evolution.genetics.washington.edu/lamarc/popgensoftware.html

the assumptions associated with the varying methods, one can make informed decisions about which methods would be appropriate given the questions of interest.

The differing assumptions associated with various methods include not only those concerning the parameters included in the model (e.g. methods that can be used to test hypotheses or estimate population divergence times may or may not include a migration parameter), but also implicit assumptions the user makes by applying the method. Because specific details of the history, including the values of the model's parameters, affect the ability and power of different methods to test statistical phylogeographical hypotheses, the applicability of the methods will vary. To illustrate how important these implicit assumptions are, consider the case where the choice is made to use the reconstructed gene tree itself to test a phylogeographical hypothesis, or in the latter example to use a Bayesian method.

The structure of any single gene genealogy can be quite informative under both migration and isolation. However, the confidence surrounding an inference about population structure will differ under the respective models, and also depends on the migration rate and divergence time (Takahata & Slatkin, 1990). As migration decreases or divergence times increase, the structure of gene trees from independent loci will become more concordant. Yet, the rate of convergence among loci on a common topological structure (like a population division) is slower under a history of divergence with gene flow compared with divergence in isolation. Consequently, a gene tree reconstructed for one locus will generally be more informative about population structure under isolation than under migration, all else being equal (Wakeley, in press). Therefore, the historical context itself dictates whether a single or multiple gene trees would be more appropriate, or if the structure of the reconstructed gene trees should even be considered.

Relying exclusively on the structure of a single gene tree itself arguably may place too much emphasis on an estimate that may or may not accurately reflect a species' past, whereas by ignoring genealogical structure, a method may discard valuable historical information (Knowles & Maddison, 2002; Wakeley, in press). For example, when populations conform to the standard coalescent model, the branching pattern of a gene tree does not contain much information about the population history (Hudson, 1990). Nevertheless, statistics related to tree length might be useful to testing different hypotheses about historical population size or expansion (e.g. Slatkin & Hudson, 1991). When the structure of the gene tree does not contain information relevant to the statistical phylogeographical test, or where there are concerns about how robust the conclusions are to assumptions about the particular structure of a genealogy (e.g. Slatkin & Hudson, 1991; Huelsenbeck & Imennov, 2002), integration over all possible genealogies (e.g. Kuhner *et al.*, 1998; Beerli & Felsenstein 2001; Nielsen & Slatkin, 2000) may be preferable. In this case, the gene trees function as theoretical tools for deriving parameters of interest rather than as the basis for the inference itself (Hudson, 1998; Rosenberg & Nordborg, 2002).

Similarly, while Bayesian methods are particularly appealing in that they allow 'prior' information to be incorporated into the model when evaluating the likelihood of the data (including information from other sources such as geographical data; e.g. Pritchard *et al.*, 2000; Gaggiotti *et al.*, 2002), because they also require the incorporation of 'prior information', they may not always be appropriate when we have no information to specify the prior. For example, Wang (2003) demonstrated that assuming that allele frequencies in the parental populations are independent (e.g. Chikhi *et al.*, 2001) biases the estimation of admixture proportions, especially when the parental populations are not completely differentiated.

So it is not the case that a particular method will necessarily always be better than another – the issue is whether a specific method is appropriate given the question at interest, irrespective if it is a summary-statistic approach or a full-likelihood analysis. Therefore successful application of these methods (e.g. Table 4) requires careful consideration of whether the assumptions of the methods are reasonable on a study-by-study basis (Stephens, 2001; Knowles & Maddison, 2002). The methods cannot be treated as a black-box from which 'the' answer will emerge.

Conclusions

With careful attention to decisions about how we test historical hypotheses, statistical phylogeographical approaches can provide unprecedented views into a species' history. This information is essential to understanding what factors shape patterns of population genetic variation and divergence, but also the processes underlying the speciation process as well. In the future, methodological developments that permit testing of complex models while incorporating information from multiple loci and data from external sources will no doubt become commonplace. Thus, the accuracy and potential insights of statistical phylogeographical tests will continue to increase, thereby making statistical phylogeography a vital and integrative link between evolutionary and ecological processes.

Acknowledgments

I would like to thank Wayne Maddison for thoughtful discussions about conceptual issues related to historical inference. I would also like to thank Josepha Kurdziel and Peter Midford for helpful comments on earlier drafts of the manuscript, and the suggestions of Roger Butlin and two anonymous reviewers that significantly improved this mini-review.

References

Arbogast, B.S., Ewards, S.V., Wakeley, J., Beerli, P. & Slowinski, J.B. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic time scales. *Annu. Rev. Ecol. Syst.* **33**: 707–740.

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

Avise, J.C. 1998. The history and purview of phylogeography: a personal reflection. *Mol. Ecol.* **7**: 371–379.

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

Beaumont, M., Barrat, E.M., Gottelli, D., Kitchener, A.C., Daniels, M.J., Pritchards, J.K. & Bruford, M.W. 2001. Genetic diversity and introgression in the Scottish wildcat. *Mol. Ecol.* **10**: 319–336.

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

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.

Brumfield, R.T., Beerli, P., Nickerson, D.A. & Edwards, S.V. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.* **18**: 249–256.

Chikhi, L., Bruford, M.W. & Beaumont, M.A. 2001. Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. *Genetics* **158**: 1347–1362.

Clegg, S.M., Degnan, S.M., Kikkawa, J., Moritz, C., Estoup, A. & Owens, I.P.F. 2002. Genetic consequences of sequential founder events by an island-colonizing bird. *Proc. Natl. Acad. Sci. USA* **99**: 8127–8132.

Drummond, A.J. & Rambaut, A. 2003. *BEAST v 1.0*, http://evolve.zoo.ox.ac.uk/beast/.

Edwards, S.V. & Beerli, P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeography studies. *Evolution* **54**: 1839–1854.

Felsenstein, J. 1992. Estimating effective population size from samples of sequenes: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res.* **59**: 139–147.

Gaggiotti, O.E., Jones, F., Lee, W.M., Amos, W., Harwood, J. & Nichols, R.A. 2002. Patterns of colonization in a metapopulation of grey seals. *Nature* **416**: 424–427.

Griffiths, R.C. & Tavaré, S. 1994. Simulating probability distribtutions in the coalescent. *Theor. Pop. Biol.* **46**: 131–159.

Hudson, R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evol. Biol.* **7**: 1–44.

Hudson, R.R. 1998. Island models and the coalescent process. *Mol. Ecol.* **7**: 413–418.

Hudson, R.R. & Turelli, M. 2003. Stochasticity overrules the ''three-times rule'': genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial data. *Evolution* **57**: 182–190.

Huelsenbeck, J.P. & Imennov, N.S. 2002. Geographic origin of human mitochondrial DNA: accommodating phylogenetics uncertainty and model comparison. *Syst. Biol* **51**: 155–165.

Hugall, A., Moritz, C., Moussalli, A. & Stanisic, J. 2002. Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail *Gnarosophia bellendenkerensis*. *Proc. Natl. Acad. Sci. USA* **99**: 6112–6117.

Kidd, D.M. & Ritchie, M.G. 2000. Inferring the patterns and causes of geographic variation in *Ephippiger ephippiger* (Orthoptera, Tettigoniidae) using geographical information systems (GIS). *Biol. J. Linnean Soc.* **71**: 269–295.

Kliman, R.M., Andolfatto, P., Coyne, J.A., Depaulis, F., Kreitman, M., Berry, A.J., McCarter, J., Wakeley, J. & Hey, J. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**: 1913–1931.

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

Knowles, L.L. 2001b. Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. *Proc. R. Soc. Lond. B* **268**: 1–6.

Knowles, L.L. & Maddison, W.P. 2002. Statistical phylogeography. *Mol. Ecol.* **11**: 2623–2635.

Kringes, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M. & Paabo, S. 1997. Neanderthal DNA sequences and the origin of modern humans. *Cell* **90**: 19–30.

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

Maddison, W.P. 1997. Gene trees in species trees. *Syst. Biol.* **46**: 523–536.

Maddison, W.P. & Maddison, D.R. 2000. *Mesquite: A Modular Programming System for Evolutionary Analysis*. http://mesquite-project.org.

Milot, E., Gibbs, H.L. & Hobson, K.A. 2000. Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). *Mol. Ecol.* **9**: 677–681.

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

Nielsen, R. & Wakeley, J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.

Nordborg, M. 1998. On the probability of Neanderthal ancestry. *Am. J. Hum. Genet.* **63**: 1237–1240.

O'Ryan, C., Harley, E.H., Bruford, M.W., Beaumont, M., Wayne, R.K. & Cherry, M.I. 1998. Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Animal Conservation* **1**: 85–94.

Pamilo, P. & Nei, M. 1988. Relationships between gene trees, and species trees. *Mol. Biol. Evol.* **5**: 568–583.

Penny, D., Steel, M., Waddel, P.J. & Hendy, M.D. 1995. Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Mol. Biol. Evol.* **12**: 863–882.

Pluzhnikov, A. & Donnelly, P. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* **144**: 1247–1262.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inferences of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Rannala, B. & Mountain, J.L. 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* **94**: 9197–9201.

Ray, N., Currat, M. & Excoffier, L. 2003. Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* **20**: 76–86.

Romauldi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A. & Barbujani, G. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Research* **12**: 602–612.

Rosenberg, N.A. & Nordborg, M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphism. *Nature Rev. Genet.* **3**: 380–390.

Rousset, F. & Raymond, M. 1997. Statistical analyses of population genetic data: new tools, old concepts. *Trends Ecol. Evol.* **12**: 313–317.

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

Stephens, M. 2001. Inference under the coalescent. In: *Handbook of Statistical Genetics* (D. J. Balding, M. Bishop & C. Cannings, eds), pp. 213–238. John Wiley & Sons, LTD, New York.

Takahata, N. 1989. Gene genealogy in tree related populations: consistency probability between gene and population trees. *Genetics* **122**: 957–966.

Takahata, N. & Satta, Y. 2002. Pre-speciation coalescence and the effective size of ancestral populations. In: *Modern Developments in Theoretical Population Genetics* (M. Slatkin & M. Veuille, eds), pp. 52–71. Oxford University Press, Oxford.

Takahata, N. & Slatkin, M. 1990. Genealogy of neutral genes in two partially isolated populations. *Theoret. Pop. Biol.* **38**: 331–350.

Templeton, A.R., Routman, E. & Phillips, C.A. 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. *Theor. Popul. Biol.* **59**: 133–144.

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

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

Wall, J.D. 2000. Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* **154**: 1271–1279.

Wang, J. 2003. Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics* **164**: 747–765.

Wilding, C.S., Butlin, R.K. & Grahame, J. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* **14**: 611–619.

Wilson, I., Weale, M. & Balding, D. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic math probabilities. *J. Royal Stat. Soc. Series A* **166**: 155–188.

Received 9 August 2002; revised 4 April 2003; accepted 9 August 2003