Phylogeography

Jack W Sites Jr, *Brigham Young University, Provo, Utah, USA*

Mariana Morando, *Centro Nacional Patagónico-Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn, Chubut, Argentina*

Advanced article

Article Contents

- Introduction
- Early Empirical Studies
- Transitions in the Discipline
- Parallel Development of Coalescent Theory
- A Cloud of Gene Histories
- Comparative Phylogeography – Analyses of Codistributed Species to Recover Shared Patterns
- Quantitative Phylogeography
- Nested Clade Phylogeographic Analysis
- Statistical Phylogeography
- Future Directions in Phylogeography

Online posting date: 15th December 2009

Phylogeography is a relatively young discipline, having been introduced into the literature in 1987. Its original focus was the analysis of gene trees (derived from molecular sequence data) in spatial geographic contexts, and for almost a decade, the field was dominated by the use of the mitochondrial deoxyribonucleic acid (DNA) locus (in animals). Because phylogeography contains an explicit tree-based focus on population genetic questions, it has successfully linked this discipline to the previously disconnected domain of phylogenetic systematics. From a largely descriptive beginning, phylogeography has become more rigorous by the inclusion of independent nuclear gene loci, more quantitative by incorporation of various statistical methods (nested clade phylogeographic analysis, statistical phylogeography) and more synthetic by incorporation of coalescent theory (using gene trees to estimate species trees) and data and methods from disciplines such as landscape genetics, palaeoecology and palaeoclimatology. Modern phylogeography has many applications to other disciplines of biology.

Introduction

As noted in recent reviews (Beheregaray, 2008; Avise, 2009), most species show intraspecific population genetic structure that can be interpreted in geographic and chronological contexts. Estimating and quantifying the spatial and temporal components of intraspecific population structure, and interpreting the evolutionary and ecological processes responsible for it, are major goals of phylogeography. The word phylogeography was introduced by Avise *et al*. (1987) to describe a research programme based on the analysis of gene trees in explicit geographic contexts and focusing on relationships among populations within species, or among closely related species. At that time animal mitochondrial deoxyribonucleic acid (mtDNA) was promoted as the ideal genetic marker for multiple reasons (Avise, 2000, 2007). Because this tree-based approach to the study of population structure deviated dramatically from more traditional approaches to population genetics (Hey and Machado, 2003), Avise emphasized the mtDNA-based approach as a bridge to provide a means for 'for expanding communication between systematists and population geneticists. With empirical and conceptual channels opened, it might be possible to reconsider various connections between micro- and macroevolutionary change as interpreted against a continuous genealogical backdrop'. Avise *et al*. (1987) also noted that their 'review will be a success if it stimulates further dialogue in these areas'.

Most would agree now that phylogeography has matured into a major research endeavour (Avise, 2000, 2009; Beheregaray, 2008) and that it is expanding rapidly and becoming more integrative as additional markers, new data from other disciplines and conceptual advances add to its richness. Because it focuses on the population–species interface and demographic/evolutionary history, phylogeography has provided valuable contributions to studies of geographic variation, speciation, historical biogeography, human evolution, conservation biology, biodiversity research/taxonomy, palaeoecology, palaeoclimatology and volcanology (Beheregaray, 2008). Here we review some of the history of phylogeography and summarize a few of the important advances in its remarkable development.

See also: Geographical Variation; Speciation: Introduction

ELS subject area: Ecology

How to cite:
Sites Jr, Jack W; and, Morando, Mariana (December 2009) Phylogeography. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0003352

Early Empirical Studies

Several of the studies summarized in Avise *et al*. (1987) were based on vertebrate taxa in the southeastern United States, using mtDNA haplotypes identified via mapped restriction sites. Because mtDNA inheritance is both asexual and haploid and lacks the recombination associated with nuclear genes during sexual reproduction, mutations alone normally account for mtDNA genetic variability commonly documented in animal populations. An early study on the pocket gopher *Geomys pinetis* (Avise *et al*., 1979) revealed that (1) mtDNA variation was high (20 haplotypes in 87 individuals), (2) all haplotypes were spatially localized, (3) closely related haplotypes (i.e. those differing by only one or two mutation sites) were in close geographic proximity and (4) a deep genealogical split separated gophers in the eastern versus western parts of the species' range. Hundreds of studies of animal populations have confirmed this same pattern for low-vagility species: closely similar or identical haplotypes are geographically localized, and a small number (typically one to six) of deeper genealogical splits usually distinguish regional groups of populations that reflect older divergence events (Avise, 2000, 2007; Beheregaray, 2008). These studies also revealed that highly vagile or migratory species often show mtDNA genealogically structured populations when historic barriers are insurmountable (i.e. the relatively recent uplift of the Panamanian isthmus and the deep divergence between Caribbean and Pacific populations of marine animals), or if a species is philopatric to particular localities, as in the case of female nest-site fidelity of marine turtles (Avise, 2009).

A radical notion at the time of these early studies was that each individual could be treated as an operational taxonomic unit (OTU) in the analysis, whereas in previous more traditional population genetic studies (Hey and Machado, 2003), Mendelian factors (diploidy, independent assortment) required that population samples rather than individuals be treated as the focal units of analysis. This in turn required a priori identification of population samples, which introduced an undesirable element of circularity in population genetic studies. Introduction of the mtDNA methods allowed matrilineal histories of individual organisms to be estimated without the earlier circularity of predefined population samples and also permitted application of explicit phylogenetic methods to questions of intraspecific evolution. Phylogenetic methods were not considered relevant within sexually reproducing species in which pathways of descent were better described as interwoven networks rather than branching trees, but the nonrecombining matrilineal inheritance of the mtDNA locus side-stepped these limitations, making tree-based phylogenetic methods relevant to reconstruction of intraspecific phylogeographic relationships.

Early studies of plants revealed that their mitochondrial genomes were highly variable in size and that plant mtDNA evolves slowly with respect to nucleotide substitutions; it is approximately 100-fold slower, making it poorly suited for within-species phylogeographic studies (Palmer, 1990). The other plant organellar genome, chloroplast DNA (cpDNA), varies moderately in size and evolves faster than plant mtDNA but still much slower than animal mtDNA. Plant organellar genomes may be transmitted maternally in some groups, biparentally in some and paternally in others (i.e. gymnosperms), which requires caution in their use for phylogeographic assessments.

If the mtDNA clades within a species reflected the boundaries of local populations, then phylogeography was about the evolution of meta-population lineages within a species (Avise, 2000). However, a common point in most of these early studies was the absence of quantification of geographic patterns revealed by the mtDNA gene trees; most studies simply made a qualitative link between the gene tree and the spatial distribution of the clades on the basis of a visual inspection of the mtDNA clades plotted on maps.

Transitions in the Discipline

In an issue of *Molecular Ecology* celebrating the first decade of phylogeography (Avise, 1998), Bermingham and Moritz (1998) summarized several emerging patterns. First, comparative mtDNA studies of North American birds, South American rodents and marsupials and frogs and lizards across the Australian Wet Tropics, collectively discounted Late Pleistocene models of speciation in these groups by suggesting that many species pairs significantly predated the Pleisocene. Second, phylogeographic studies yielded vastly improved descriptions of distributions and relationships among lineages of animals, which in turn led to a better understanding of regional biogeography and areas of endemism. Finally, they identified three future challenges to comparative phylogeographic studies: (1) adding unlinked nuclear markers, along with improved analytical methods for testing the congruence between nuclear and organelle genes; (2) incorporation of tools from coalescent theory into estimates of population histories and demographic parameters (effective population sizes, mutation rates, gene flow, etc.) and (3) improving the precision for estimating the timing of divergence events. They also redefined phylogeography 'to test the congruence between the evolutionary, demographic and distributional histories of taxa against the particular geological and ecological setting of a region and to determine the chronology of evolutionary diversification'. Comparative phylogeography was then described as study of the evolution of landscapes and the effects of history and geography on population and community structure at local and regional levels.

Parallel Development of Coalescent Theory

In parallel with the development of phylogeography, demographers studying the dynamics of surname turnover

ENCYCLOPEDIA OF LIFE SCIENCES © 2009, John Wiley & Sons, Ltd. www.els.net

in human populations developed models of lineage branching processes that could be transferred with little modification to studies of gene branching processes of the mtDNA locus (Avise, 2000, pp. 28–32). This led to a rapid development of a body of analytical and statistical methods of 'coalescent theory' that explores connections between population demography and gene genealogies within and among closely related species (Hudson, 1990; Kingman, 1982a, b; Tajima, 1983). Unlike traditional 'forward-looking' population genetic theory designed to predict changes in allele frequencies over time (Ewens, 1972), coalescent theory projects gene genealogies backwards in time to estimate parameters such as historical population sizes, divergence times and migration rates, while accounting for the stochastic processes of lineage sorting within and among gene trees (Wakeley, 2008).

Figure 1 shows a hypothetical case of the stochastic sorting of haplotypes for a single gene tree within a species tree. The topology of the species tree is pectinate due to sequential speciation events along the same lineage: These proceed in the order A, B and C, from the oldest to the most recent and produce a species topology of (((1, 2) 3) 4). In contrast, the origin and branching of haplotypes (alleles) on the gene tree, shown by lower case letters a, b and c (from oldest to most recent mutations), reflect the stochastic nature of the sorting process that does not track the population splits in the species tree. This produces a symmetrical gene tree topology ((1, 2) (3, 4)) that recovers terminal 3 as the sister to terminal 4, instead of sister to clade (1, 2) in the species relationships. The result of this 'lineage sorting' process is that if different alleles within a population to not coalesce back to a Most-Recent-Common-Ancestor (MRCA) within their population (or species), but rather this MRCA coalescence is older than the speciation event, then some individuals within the same population/species will be more closely related to individuals in another population/species, than to individuals within their own group.

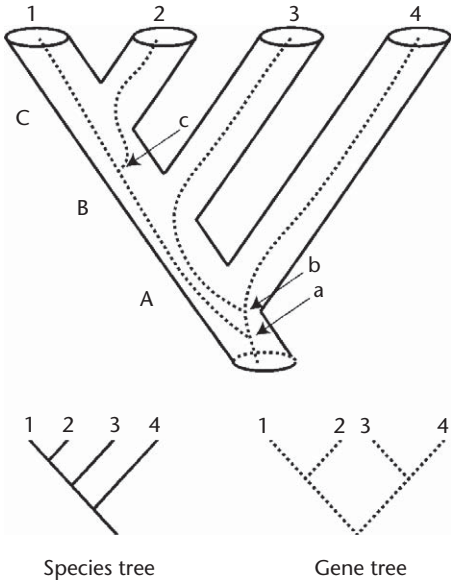

Figure 1 An example of species tree–gene tree discordance; in the main figure the open cylinders represent species relationships produced by a series of three speciation events along a single pathway of descent. The speciation events are indicated in the order from oldest to the most recent (upper-case letters A→B→C), and the relationships are summarized by the pectinate topology in the lower left (((1, 2) 3) 4). The dotted lines inside of the cylinders represent allele, or haplotype relationships in a single gene tree, with new alleles originating by mutation events identified from the oldest to the most recent, with the lower case letters a→b→c. The gene tree topology thus differs from the species tree with respect to placement of the allele at terminal 3; relationships among haplotypes are symmetrical in the gene tree because haplotype b is sister to haplotype a ((1, 2) (3, 4)), whereas species 3 is the sister terminal to the (1, 2) species clade. The text discusses the common mechanisms by which the sorting of alleles in a gene tree may not match the splitting events in the species tree.

The critical point is that matrilineal (and patrilineal, such as mammalian Y-chromosome loci) genealogies are tree-like (branched and nonreticulate) despite being embedded within a sexually reproducing species pedigree that is overall highly reticulate (due to interbreeding). These asexually transmitted markers are 'self-pruning' by lineage sorting in which some haplotypes are lost by drift from a pedigree (female lines die out), while others proliferate (due to higher fecundity in other females) and new ones appear by mutation (see review by Nielsen and Beaumont, 2009).

In real populations of sexually reproducing organisms, lines of descent are dominated by nuclear genes that are usually transmitted through both parents and may also undergo recombination. Real pedigrees in noninbred populations are therefore composed of multiple gene genealogies, most of which show reticulate patterns of descent. In each generation the number of transmission pathways for haplotypes at any Mendelian (nuclear autosomal) locus is 4-fold greater (M→M, F→F, M→F, or F→M) than for mtDNA (F→F) or Y-chromosome loci (M→M). This implies that the mean coalescent time back to a common ancestor is 4-fold greater for a Mendelian locus relative to the unisexually transmitted mtDNA or Y-chromosome loci. The general point is that unlinked loci (maternal, paternal or Mendelian) may have different transmission histories (and therefore different tree shapes) within the same organismal pedigree.

A Cloud of Gene Histories

Maddison (1997) explored the relationship between gene trees and species trees and introduced the idea of phylogeny as a diffuse cloud of gene histories. He described a phylogeny (an intra- or interspecific genealogy) as a statistical distribution of gene trees with a central tendency and a variance due to the diversity of tree topologies. Thus, some gene genealogies, those that are part of the variance, may not be congruent with the species history. He also reviewed processes by which gene tree–species tree discordance can arise, and identified three main factors: horizontal gene transfer, lineage sorting or deep coalescence and gene duplication/deletion. Horizontal transfer may be accomplished by a

vector and the transferred genes also have to be incorporated as functioning components of the recipient genome. With respect to lineage sorting, ancestral polymorphisms can persist through several speciation events, and gene trees may differ from the species due to the demographic processes described earlier. If we see this process backwards in time, this issue of 'deep coalescence' means that the point back to the MRCA of gene copies at a single locus extends deeper than speciation events. As an example, **Figure 2** summarizes some of the points made by Maddison (1997) and shows three hypothetical gene trees embedded in a species tree. Two of the gene trees (A and B) have the same topology as the species tree (and each other), but differ in the depth of coalescent times; gene tree A for example has a deeper MRCA between alleles B and C, one which considerably predates the speciation event, relative to gene tree B.

Topological differences among gene trees (gene tree C in **Figure 2**) may arise because gene trees sampled from different individuals in a population are random realizations of a stochastic allele-sorting process. Gene trees relating individuals within a population will differ depending on which individuals are sampled, because gene tree evolution is a random process owing to both variance in reproductive success and random segregation of alleles in diploid organisms (Nielsen and Beaumont, 2009). Further, even if all individuals in a population could be sampled, the time at which all alleles would coalesce back to an MRCA, and other properties of any gene tree, would still have a strong random component that is heavily influenced by the ancestral effective population size (N_E) and the time between divergence events (as in gene trees A and B in **Figure 2**). Thus gene trees with topologies identical to the species tree may differ substantially in coalescent depths, and this 'branch length heterogeneity' has only recently been emphasized as an intrinsic property of natural populations (Edwards, 2008).

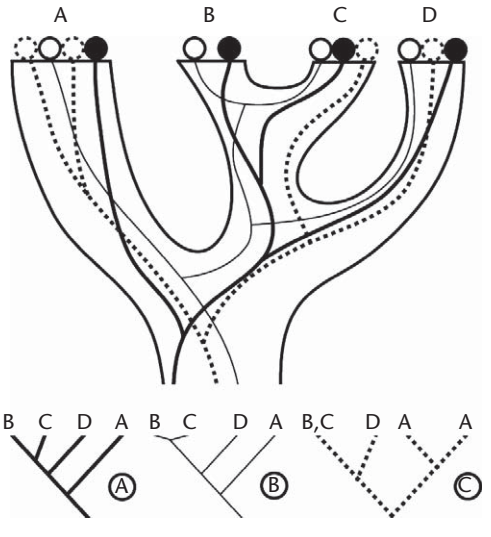

Figure 2 A hypothetical species tree (open pathways delimited by heavy lines) that includes three gene trees reflecting different patterns of haplotype evolution. Haplotypes identified by solid circles connected along the genealogy identified by heavy lines and those identified by open circles along the genealogy of light lines both evolved along a pathway that matches the species tree in the sequence of branching points (((B, C) D) A), but these differ in the depths of their coalescent points (compare gene tree topologies A and B). Haplotypes identified by dotted circles and connected along the genealogy dotted lines, evolved along a pathway discordant with the species tree (gene tree topology C). Modified from Maddison (1997). Reproduced with permission from The Society of Systematic Biologists.

Each of the three events identified by Maddison (1997) likely responsible for gene tree–species tree discordance depends on different circumstances for its occurrence. For example, at very recent temporal scales near the species level, gene duplication may be unlikely and either deep coalescence and/or hybridization could be assumed. In **Figure 1** for example, small N_E in the common ancestor of each species split, and long intervals of time (long branch lengths) between speciation events, will accelerate the sorting process for any gene tree to a condition known as reciprocal monophyly; coalescence to the MRCA occurs first within each of two sister clades and then between them. The reverse conditions, with larger population sizes and shorter branch lengths, increase the time required for coalescence to occur.

Maddison stressed the need to combine data from multiple genes to arrive at a good estimate of species tree at shallow levels of divergence and proposed species tree estimation procedures by incorporating coalescent processes into maximum parsimony or maximum likelihood algorithms. Of all the processes generating gene tree–species tree discordance, deep coalescence is perhaps the most problematic because it is expected to be ubiquitous in sexual species. Maddison makes clear that this is not a 'problem'; instead deep coalescence is a natural outgrowth of our view of evolution, and as such a tree is a broad-scale, low-resolution view of the genetic connections from one generation to the next. He referred to this as the 'realized genetic history': a summary of the history of the passage of all the genes through the generations.

This proposal by Maddison – viewing population and species histories as 'clouds of gene histories' – was a major conceptual advance because it pointed out limitations of the 'mtDNA only' studies. This locus, even though for various reasons, should track divergence events with greater fidelity than any single-nuclear gene tree, can be overinterpreted as a species genealogy because it alone cannot accommodate the stochasticity of coalescent processes (Edwards and Beerli, 2000). Others have also realized the same limitations of single gene tree studies, and that performance is improved for many analytical methods when these are based on multilocus data sets (Wakeley and Hey, 1997; Hudson and Turelli, 2003).

Comparative Phylogeography – Analyses of Codistributed Species to Recover Shared Patterns

Bermingham and Avise (1986) first studied multiple codistributed species (freshwater fish of the southeastern

USA), to assess shared signal of underlying historical events driving divergence in unrelated taxa. **Figure 3** provides a hypothetical example in which two species–species complexes, identified by trees 1 and 2, are approximately codistributed across three isolated habitats (areas 1 through 6). Each species is sampled from the same number of localities in each region (species 1 in the open areas, species 2 from shaded areas; see figure caption for details in **Figure 3**). The strongest signal for shared historical events (orogenies, river captures, glacial cycles, etc.) on unrelated taxa would be evidence of statistically significant spatial and temporal codivergence. In this example, species 1 and 2 would provide evidence for codivergence in both time and space in a comparison of trees 1 and 2. Here topologies and branch lengths (surrogates for time since divergence) are equal, which is good initial evidence for spatial and temporal codivergence in these two groups. Two alternatives to this scenario can be imagined; first, the two groups may not have matching phylogeographic histories (i.e. their population/species trees have different topologies), or they may show spatial codivergence (identical trees) but not temporal codivergence. An example of this second case is illustrated by a comparison of trees 1 and 3; here the spatial divergence has been the same, but tree 3 implies a much longer divergence time – one in which the splits happened earlier in time – than the one implied by tree 1. In this case we would conclude that the groups codiverged on the same spatial scales, but the divergence history of species 3 was much earlier than that for species 1. Recent quantitative methods (see section on Future Directions in Phylogeography) allow for discrimination of different divergence times across the same spatial barriers (Hurt *et al*., 2009).

Quantitative Phylogeography

The earliest phylogeographic studies were simply qualitative visualizations of gene trees with the geographic distribution of collecting localities and traditional population genetic methods based only on the spatial distribution of alleles frequencies (Fst and related estimators), do not use historical information contained in the gene trees that are the core component of phylogeography. Quantitative approaches have been developed along a number of traditions, and this continues to be an extremely active area of research (Nielsen and Beaumont, 2009).

Nested Clade Phylogeographic Analysis

The 'visual inspection' approach of how geography overlays a haplotype tree cannot assess adequacy of sample sizes or sampling locations for distinguishing among potential causes of geographical associations. The method of Nested Clade Analysis (NCA) was proposed to identify and distinguish between historical and recurrent demographic processes influencing the spatial distribution of genetic variation (Templeton *et al*., 1995). This method has been expanded to incorporate multiple loci and is now called Nested Clade Phylogeographic Analysis (NCPA) (Templeton, 2004). The method integrates information from gene genealogies, haplotype frequencies and geographic localities and first requires a haplotype network estimated by a parsimony algorithm. The network is used to group haplotypes into nested clades in a relative temporal dimension, and this information is combined with the geographic locations of the haplotypes to determine if a statistical association exists between haplotypes and geography. The geographical data are quantified as either: (1) the clade distance, Dc, which measures the geographical range of a particular clade relative to its mean location; or (2) – the nested clade distance, Dn, which measures the geographic spread of members of a clade relative to the mean location of all members of the nesting clade.

Contrasts in these distance measures between tip clades and the clades immediately interior to them in the cladogram are important in discriminating the potential causes of geographical structuring of the genetic variation. The statistical significance of the different distance measures and the interior-tip contrasts are determined by random permutation testing to simulate the null hypothesis of a random geographical distribution for all clades within a nesting category, given the marginal clade frequencies and sample sizes per locality. If a statistical association is significant, then inferences about the population history and gene flow can be made by use of an inference key (December 2008: http://darwin.uvigo.es/software/geodis.html) which provides qualitative interpretations. This key is structured as follows (abbreviated):

1. Are all clades within the nesting clade found in separate areas with no overlap?
   NO – go to step 2.
   YES – go to step 19
2. Are both of the following conditions satisfied:
   (a) The clades with significantly small Dc or Dn values have ranges that are completely or mostly non-overlapping with the other clades in the nested group
   (b) The pattern of completely or mostly non-overlapping ranges in the above condition represents a break or reversal from lower-level trends within the nested clade series
   NO – Restricted Gene Flow With Isolation-by-Distance
   YES – go to step 9.

The NCPA has become controversial for several reasons (summarized in Nielsen and Beaumont, 2009), but more importantly because of the possible ambiguity in the interpretation of the inference key. Among other objections, simulation studies show that the inference key may misassign a demographic process to explain a given genetic pattern (Knowles and Maddison, 2002; Knowles, 2008; but see Templeton, 2009a, b). Simulation studies by Panchal

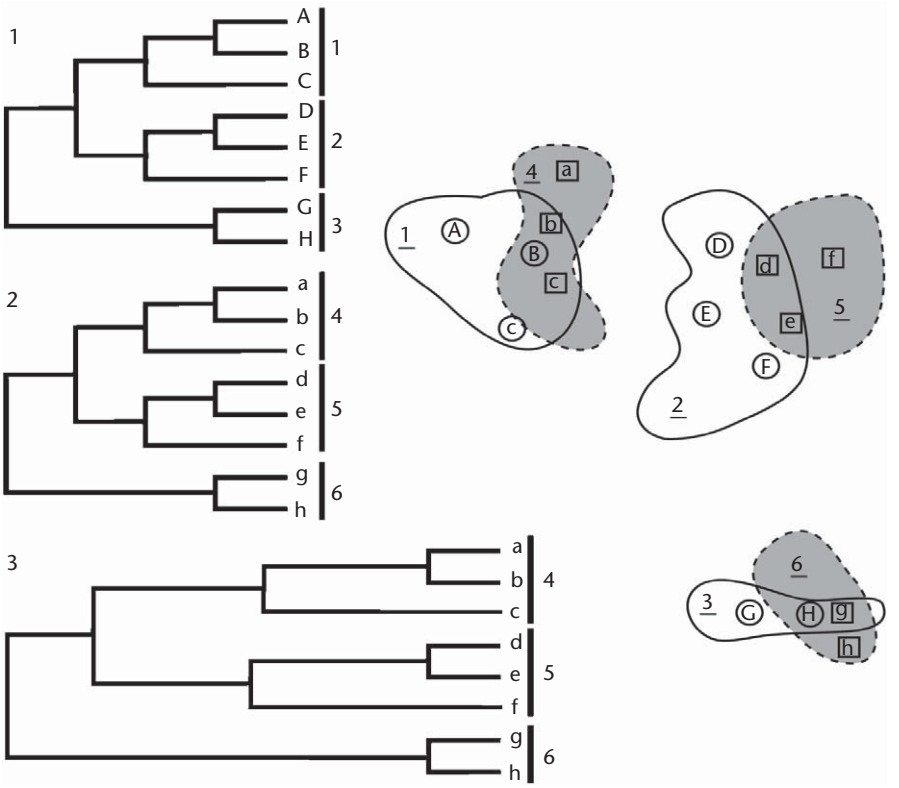

Figure 3 An hypothetical comparative phylogeographic scenario in which two unrelated groups of organisms, represented by upper- ('birds') and lower-case ('lizards') letters, respectively, are codistributed in partially overlapping distribution; the birds from allopatric populations confined to the open areas and the lizards from allopatric populations confined to the shaded areas. The upper and lower-case letters identify distinct haplotypes sampled from these areas, for the same gene region, and trees 1, 2 and 3 present topologies for the birds (tree 1) and two alternative lizard topologies (trees 2 and 3). For the sake of argument, branch lengths are proportional to time in all gene trees, the gene trees accurately reflect the population (or species) relationships and all nodes are strongly supported.

and Beaumont (2007) showed incorrect historical processes were inferred in over 75% of the datasets. This last work motivated a series of papers, providing opposite positions regarding the value of the method (Petit, 2008; Garrick *et al.*, 2008; Knowles, 2008). Templeton (2008) addresses points where he saw flaws in the commentaries made by Knowles and Maddison (2002) and Panchal and Beaumont (2007), and he explores the issue of high false-positive rates. Templeton (2008, 2009a, 2009b) develops further statistical and biological arguments to show that the method has been subjected to extensive validation through the analysis of actual data sets in a manner that satisfies what opponents point out, and that it is robust if based on multiple loci (Templeton, 2002) and inferences from two or more loci suggest the same process at the same location. The NCPA advanced phylogeography by making a large part of the analysis statistical, instead of simple visualizations of genes and geographic distributions.

Statistical Phylogeography

In a rather different tradition, statistical phylogeogaphy (SP) was used by Knowles and Maddison (2002) to describe an approach based on methods that make both explicit statistical links between process, prediction and test and to also incorporate a diverse array of processes and histories. They presented two examples on how to test historical biogeographic hypotheses and one on how to estimate parameters under this statistical approach, which considers the stochasticity of population genetic processes and offers an explicit assessment of the confidence in any specific conclusion. They recognized that no optimum method yet existed, but emphasized what limitations have to be kept in mind when trying to make historical inferences. They identified three main areas vital for phylogeographic studies that needed further development: (1) parameterization of the models, (2) strategies for searching through alternative histories and (3) criteria for judging the adequacy of explanations of the data. With the expansion of coalescent theory, SP represents a shift from a posteriori interpretation of patterns to testing hypotheses based on models that are defined a priori, and these present new challenges (Knowles, 2004). Species histories can be complex, and there are challenges related to three key steps in SP: (1) how to define a realistic set of alternative historical hypotheses, (2) how to decide on the complexity of any model and (3) how to integrate information from external

data. The stochasticity inherent to the coalescence processes described earlier, especially for recently diverged populations/species, has significant consequences for testing historical hypotheses and is one the most challenging methodological aspects for statistical phylogeography (Nielsen and Beaumont, 2009).

To explore how much information can be extracted from gene trees to reconstruct species trees, Maddison and Knowles (2006) tested two methods, one of which was a parsimony algorithm that minimizes the number of deep coalescences (Maddison, 1997). The second method clustered species directly by their most similar sequences, and although both are crude, simulations over a wide range of N_E sizes, generation times, numbers of loci and numbers of individuals sampled, showed that gene sequences retain enough signal to make accurate estimates of a species phylogeny despite incomplete lineage sorting. In general, at shallower levels, more individuals than loci give better results and the inverse for deeper divergences. Statistical phylogeography advanced the discipline by providing a means to evaluate alternative hypotheses of the evolution of populations or species and then select the one best fitting the data. The caveat is that the true history frequently will be much more complex than any of those simulated in alternative coalescent models, and therefore not captured by this approach (Templeton, 2009b).

Future Directions in Phylogeography

We have only briefly touched on some of the basic components of contemporary phylogeography, and indeed the number of recent reviews attests to the rapid growth of the field in many directions (Beheregaray, 2008; Brito and Edwards, 2009; Nielsen and Beaumont, 2009). The increasing availability of nuclear genetic markers, recent advances in landscape genetics, coalescent theory and new tools for generating ecological niche and palaeoclimate models are shifting the direction of phylogeographic studies by making them more synthetic. These include methods that can test explicit a priori hypothesis when available or estimate phylogeographic history in the absence of such hypotheses in a statistical framework (Lemmon and Lemmon, 2008). We have neither discussed the rapid growth of the application of coalescent methods to species delimitation (Carstens and Knowles, 2007; Edwards *et al*., 2007; Liu and Pearl, 2007; Brumsfield *et al*., 2008) nor applications of comparative phylogeographic methods to conservation of evolutionary processes (Davis *et al*., 2008). Other studies incorporate external climatic and geologic data to generate a priori predictions (e.g. Richards *et al*., 2007; Knowles and Carstens, 2007), and this approach has been successfully extended to multispecies comparative studies (Carnaval *et al*., 2009; Hurt *et al*., 2009). Statistical and computational issues remain challenging (Nielsen and Beaumont, 2009), but phylogeography has a bright future.

References

Avise JC (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7**: 371–379.

Avise JC (2000) *Phylogeography. The history and formation of species*. Cambridge, MA: Harvard University Press.

Avise JC (2007) Twenty-five key evolutionary insights from the phylogeographic revolution in population genetics. In: Weiss S and Ferrand N (eds) *Phylogeography of Southern European Refugia*, pp. 7–21. Dordrecht: Springer.

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography* **36**: 3–15.

Avise JC, Arnold J, Ball RM *et al*. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**: 489–522.

Avise JC, Giblin-Davidson C, Laerm J, Patton JC and Lansman RA (1979) Mitochondrial DNA clones within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences of the USA* **76**: 4350–4354.

Beheregaray L (2008) Twenty years of phylogeography: the state of the field and challenges for the Southern Hemisphere. *Molecular Ecology* **17**: 3754–3774.

Bermingham E and Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* **113**: 939–965.

Bermingham E and Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology* **7**: 367–369.

Brito PH and Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**: 439–455.

Brumsfield R, Liu L, Lum D and Edwards SV (2008) Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae: *Manacus*) from multilocus sequence data. *Systematic Biology* **57**: 719–731.

Carnaval A, Hickerson MJ, Haddad CFB, Rodrigues MT and Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science* **323**: 785–789.

Carstens BC and Knowles LL (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Systematic Biology* **56**: 1–12.

Davis EB, Koo MS, Conroy C, Patton JL and Moritz C (2008) The California hotspots project: identifying regions of rapid diversification of mammals. *Molecular Ecology* **17**: 120–138.

Edwards SV (2008) Is a new and genera theory of molecular systematics emerging? *Evolution* **63**: 1–19.

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

Edwards SV, Liu L and Pearl DK (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the USA* **104**: 5936–5941.

Ewens W (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**: 87–112.

Garrick RC, Dyer RJ, Beheregaray LB and Sunnucks P (2008) Babies and bathwater: a comment on the premature obituary for nested clade phylogeographic analysis. *Molecular Ecology* **17**: 1401–1403.

Hey J and Machado CA (2003) The study of structured populations – new hope for a difficult and divided science. *Nature Reviews. Genetics* **4**: 535–543.

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

Hudson RR and Turelli M (2003) Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescent times for nuclear loci versus mitochondrial DNA. *Evolution* **57**: 182–190.

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

Kingman JFC (1982a) On the genealogy of large populations. *Journal of Applied Probability A* **19**: 27–43.

Kingman JFC (1982b) The coalescent. *Stochastic Processes and Their Applications* **13**: 235–248.

Knowles LL (2004) The burgeoning field of stastical phylogeography. *Journal of Evolutionary Biology* **17**: 1–10.

Knowles LL (2008) Why does a method that fails continue to be used? *Evolution* **62**: 2713–2717.

Knowles LL and Carstens BC (2007) Estimating a geographically explicit model of population divergence. *Evolution* **61**: 477–493.

Knowles LL and Maddison WP (2002) Statistical phylogeography. *Molecular Ecology* **11**: 2623–2635.

Lemmon AR and Lemmon EM (2008) A likelihood framework for estimating phylogeographic history on a continuous landscape. *Systematic Biology* **57**: 544–561.

Liu L and Pearl DK (2007) Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology* **56**: 504–514.

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

Maddison WP and Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55**: 21–30.

Nielsen R and Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology* **18**: 1034–1047.

Palmer JD (1990) Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet* **6**: 115–120.

Panchal M and Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution* **61**: 1466–1480.

Petit J (2008) The coup de grâce for nested clade phylogeographic analysis? *Molecular Ecology* **17**: 516–518.

Richards CL, Carstens BC and Knowles LL (2007) Distribution modeling and statistical phylogeography: an integrative framework for generating and testing alternative hypotheses. *Journal of Biogeography* **34**: 1833–1845.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.

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

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**: 789–809.

Templeton AR (2008) Nested clade analyses: an extensively validated method for strong phylogeographic inference. *Molecular Ecology* **17**: 1877–1880.

Templeton AR (2009a) Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographic analysis vs. approximate Bayesian computation. *Molecular Ecology* **18**: 319–331.

Templeton AR (2009b) Why does a method that fails continue to be used? The answer. *Evolution* **63**: 807–812.

Templeton AR, Routman E and Phillips CA (1995) Separating population structure from population history – a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**: 767–782.

Wakeley J (2008) *Coalescent Theory*. Greenwood Village, CO: Roberts and Company. 326 pp.

Wakeley J and Hey J (1997) Estimating ancestral population parameters. *Genetics* **145**: 847–855.

Further Reading

Avise JC and Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* **7**: 45–67.

Degnan JH and Kubatko LS (2008) Discordance of species trees with their most likely gene trees. *PLoS Genetics* **2**: 762–768.

Hewitt GF (2001) Speciation, hybrid zones, and phylogeography – or seeing genes in time and space. *Molecular Ecology* **10**: 537–549.

Kidd DM and Ritchie MG (2006) Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography* **33**: 1851–1865.

Moritz C, Hoskin CJ, MacKenzie JB *et al*. (2009) Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of the Royal Society of London. Series B* **276**: 1235–1244.

Riddle BR, Dawson MN, Hadley EA *et al*. (2008) The role of molecular genetics in sculpting the future of integrative biogeography. *Program: Physical Geography* **32**: 173–202.

Rissler LJ, Hijmans RJ, Graham CH, Moritz C, Wake DB (2006) Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *American Naturalist* **167**: 655–666.

Swenson NG (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic, and speciation research. *Journal of Evolutionary Biology* **21**: 421–434.

Zink RM and Barraclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**: 2107–2121.