The history and purview of phylogeography: a personal reflection

JOHN C. AVISE

Department of Genetics, University of Georgia, Athens, GA, 30602–7223, USA

Abstract

Last year marked the 10th anniversary of the birth of phylogeography as a formal discipline. However, the field’s gestation began in the mid-1970s with the introduction of mitochondrial (mt) DNA analyses to population genetics, and to the profound shift toward genealogical thought at the intraspecific level (now formalized as coalescent theory) that these methods prompted. This paper traces the early history and explosive growth of phylogeography, and closes with predictions about future challenges for the field that centre on several facets of genealogical concordance.

Keywords: allelic pathways, coalescent theory, gene trees, genealogical concordance, intraspecific phylogeny, mitochondrial DNA

Introduction

Phylogeography is a field of study concerned with the principles and processes governing the geographical distributions of genealogical lineages, especially those at the intraspecific level. The word itself was coined a decade ago (Avise et al. 1987a) and its use in the evolutionary genetics literature has grown exponentially since then (Fig. 1). As of the end of 1996, more than 130 papers had employed ‘phylogeography’ in the title or as an index word, and they represent only the tip of the iceberg because numerous additional studies have dealt with the topic implicitly although not by name. As a subdiscipline of biogeography (Fig. 2), phylogeography emphasizes historical aspects of the contemporary spatial distributions of gene lineages (Avise 1996a). The analysis and interpretation of lineage distributions usually requires input from molecular genetics, population genetics, phylogenetics, demography, ethology, and historical geography. Thus, phylogeography is an integrative discipline.

In purest form, empirical phylogeographic analyses deal with the spatial distributions within and among populations of alleles whose phylogenetic relationships are deduced. Because mitochondrial (mt) DNA evolves rapidly in populations of higher animals and usually is transmitted maternally without intermolecular recombination, it has been the workhorse of most (> 80%) of the phylogeographic studies conducted to date (Fig. 3). However, empirical or theoretical treatments that address phylogenetic aspects of the spatial distributions of any genetic traits (morphological, behavioural, or any other) also can qualify as phylogeographic under a broader definition of the term. Furthermore, a matrilineal phylogeny (or any other allelic transmission pathway) constitutes only a minuscule fraction of the composite genealogical information within a sexual pedigree (Fig. 4). A phylogeny for spatially structured populations can be conceptualized as a statistical distribution of partially bundled allelic pathways of descent (Fig. 4) each characterized by its own unique coalescent pattern (Maddison 1995; Avise & Wollenberg 1997). The many distinctions yet connections between notions of phylogeny at the levels of genes vs. populations have made phylogeography a rich point of contact between the traditionally distinct fields of population genetics and phylogenetic biology (Avise 1989a; Hey 1994).

To introduce this special issue of Molecular Ecology on phylogeography, I will recount briefly the history of the discipline from a personal, anecdotal (and no doubt biased) perspective. Phylogeographic efforts have been tied closely to analyses of animal mtDNA, so that is where the story will begin.

History of phylogeography: one researcher’s view

Science often is serendipitous, as the following stories well illustrate. Shortly after joining the University of Georgia as an Assistant Professor in 1975, I gave a departmental
seminar describing work on allozyme variation in fishes. Echoing a sentiment popular at the time, I concluded that regulatory rather than structural genes should be studied next because changes in gene regulation were perhaps at the heart of adaptive evolution. I queried the audience for suggestions on how I might examine regulatory genes, and one responder asked whether I had considered using restriction enzymes to assay repetitive nuclear DNA sequences, which at the time were viewed as prime candidates as regulatory modulators (Britten & Davidson 1969, 1971). I had never heard of restriction enzymes! However, the idea was intriguing so I soon approached several faculties at the University in an attempt to identify a collaborating laboratory where I might learn restriction digestion techniques. To my chagrin, the inquiries met with cool responses, except one: Dr Robert Lansman welcomed me to his laboratory, but noted with apology that he had limited experience with nuclear DNA and instead conducted research on the biochemistry and cellular biology of mitochondrial DNA. I barely had heard of mitochondrial DNA! However, left with few options, I accepted Bob’s offer in order to gain familiarity with DNA level assays.

Before long, we were generating agarose gels with mtDNA restriction profiles, initially from small mammals. Although I was still viewing the effort mainly as a training exercise, intriguing questions began to emerge. Why did each individual display only a few mtDNA bands on a gel, rather than a smear of fragments from the billions of mtDNA molecules that must be included in an assay? (It must be because each specimen had a specifiable mtDNA genotype with respect to the restriction sites assayed.) Why did different mice within local populations often display distinct RFLP patterns, such that observed mtDNA variation primarily was distributed among rather than within individuals? (With hindsight, it must be because mtDNA mutations arise...
frequently, and sometimes precipitate within a small number of animal generations a genotypic turnover in the population of mtDNAs in a germ-cell lineage from which the assayed soma were derived.) Why did mtDNA genotypes in organismal populations appear connectable to one another in phylogenetically intelligible ways? (Because intermolecular recombination must be rare or nonexistent in these maternally inherited molecules, such that the matrilineal histories of mutation events were recorded in extant mtDNA genotypes.) Why did members of sexually reproducing species usually group together by mtDNA genotypes when the evolutionary connecting agents of mating and genetic recombination seemed not to apply to these asexually transmitted genomes? (Because, as we now know, coalescent processes ensure phylogenetic links among genotypes via vertical pathways of ancestry even in the absence of inter-lineage genetic exchange mediated by mating events.) What ramifications might stem from the heretical practice made possible by mtDNA of viewing haplotypes as clones and individual animals as OTUs (operational taxonomic units), in population genetic analyses? (The list of responses is now long.)

In general, many unorthodox perspectives on evolution eventually were to emerge from studies of mtDNA (reviewed in Avise 1991), but years would pass before...
relatively clear answers to some of the questions listed above and others similar to them were to be forthcoming. The lag time reflected in part the difficulty experienced by many researchers (certainly by me) in reorienting thought away from the traditional Mendelian perspectives that applied so well, for example, to allozyme systems on which many of us had been trained.

My collaboration with the Lansman laboratory went well, and our first paper on mtDNA variation in a natural population soon appeared (Avise et al. 1979a), followed shortly thereafter by the first large-scale phylogeographic survey of any species based on mtDNA lineages (Avise et al. 1979b). The technical stage for these efforts had been set in the early 1970s through prior mtDNA research on several fronts. For example, Brown & Vinograd (1974) and Upholt & Dawid (1977) had demonstrated the feasibility of generating restriction enzyme cleavage maps for animal mtDNAs; Dawid & Blackler (1972), and Hutchinson et al. (1974) among others had documented predominant maternal inheritance for mtDNA in higher animals; and Upholt (1977) had developed a statistical procedure for estimating sequence divergence among mtDNA genotypes from comparisons of restriction digests. Furthermore, in the same year that our phylogeographic works appeared in print, Brown et al. (1979) published an extremely influential article highlighting the unexpected fast pace of mtDNA sequence evolution as gauged by interspecies comparisons of higher primates.

In the late 1970s, excitement generated by the new mtDNA discoveries ran high. I remember pondering the many research possibilities, of which two of anecdotal interest can be mentioned. Early on, it occurred to me that mtDNA might be a wonderful tool for analysing the evolution of parthenogenetic vertebrates, for at least two reasons. First, all such unisexual biotypes were thought to have arisen through hybridization between sexual species, such that by utilizing mtDNA data it should be possible to identify the maternal parent taxon in each case. Second, because parthenogenetic taxa reproduce asexually, the history of maternal lineages within them should in principle be one-and-the-same as the entire organismal phylogeny (unlike the case for a sexual species). I remember reasoning that it would be safe to shelve these ideas for the moment in the belief that many years would elapse before any molecular biologists might dream of this ‘obscure’ biological application for mtDNA. I could not have been more wrong. One of the first mtDNA analyses of natural populations dealt with precisely these evolutionary issues in parthenogenetic lizards (Brown & Wright 1975, 1979; see below)!

Eventually, my laboratory did examine evolutionary processes in gynogenetic and hybridogenetic fish complexes using mtDNA (reviewed in Avise et al. 1992), but only well after Wes Brown, Craig Moritz (Brown’s postdoctoral researcher at the time), and their associates had produced an important series of mtDNA papers on the origins and evolution of parthenogenetic reptiles and other unisexual vertebrates (e.g. Densmore et al. 1989; Echelle et al. 1989; Moritz 1991).

It also seemed evident to Bob Lansman and myself that mtDNA analyses of human populations would be of great interest. However, we elected not to pursue this topic. Personally, I was wary of the inevitable social and political fallout from whatever findings might be uncovered about the nature of genetic differences between human skin colour races, or between humans and great apes; and, in any event, it seemed likely that the necessary research would be accomplished by someone. Here, my crystal ball proved truer. An influential study on human mtDNA evolution soon appeared (Brown 1980), followed by a number of more extensive but also controversy generating mtDNA analyses of higher primate phylogeny (e.g. Ferris et al. 1981; Brown et al. 1982) and human geographical variation (notably by Cann et al. 1987; reviews in Nei & Roychoudhury 1993; Cavalli-Sforza et al. 1994; Takahata 1995).

I should digress from this personal account for a moment to relate the history of Wes Brown’s involvement with mtDNA, because this traces the other major root of evolutionary interest in the molecule. The story began in 1968 when Brown went to Caltech as a graduate student and was introduced to mtDNA in the laboratories of Giuseppe Attardi and Jerome Vinograd, where mtDNA transcription and physical chemistry, respectively, were being studied. In 1971, Brown went to an exhibition of Max Escher paintings at the Los Angeles County Museum, where he happened to meet John Wright, the curator of the herpetology department. Wright was probably the most knowledgeable person in the world on Cnemidophorus lizards, and Brown’s chance meeting with him that day was to lead to their collaborative studies on the evolutionary origins of parthenogenetic taxa from a genealogical perspective. Brown gathered the mtDNA data at Caltech from 1971 to 1973 but, as mentioned above, the first papers did not appear until several years later. After a postdoctoral stint at the University of California at San Francisco, Brown moved across the Bay in 1978 to join the Allan Wilson group at Berkeley. There he restructured and equipped the laboratory for studies of animal mtDNA, and among other efforts initiated the important research mentioned above on human genealogical evolution.

Returning to the developing story at the University of Georgia, in those early years another important event for me personally stemmed from a casual conversation over lunch. I was explaining to my colleague Dr Michael Clegg our recent findings on modes of inheritance and patterns of geographical variation in mtDNA for small mammals, and he mentioned that the issues raised...
seemed analogous to those for surname evolution in many human societies. This simple comment struck home, and helped greatly in my otherwise tortuous transition from Mendelian to phylogenetic thinking at the intraspecific level. The surname analogy does indeed hold well (Avise 1989b). Just as sons and daughters ‘inherit’ their father’s nonrecombined surname (before recent rule changes in some families), so too do progeny normally receive nonrecombined mtDNA from their mothers. Furthermore, much the way that mutations sometimes arise in surnames (my own name was a 19th century misspelling of ‘Avis’), point mutations occasionally arise and cumulatively differentiate related mtDNA genotypes. Thus, mtDNA molecules record matrilineal histories much as surnames record patrilines, except that the matrilineal records extend much further back in time (surnames were invented de novo only within the past few centuries).

These insights were new to me, but not completely so to the field. Beginning much earlier in the century, statistical demographers had studied the dynamics of surname turnover in human populations (Lotka 1931) using models that now could be applied often with little modification to gene lineages such as those provided by mtDNA (Schaffer 1970). Such models stimulated my own and my students’ efforts to examine the theoretical ties between population demography and phylogeographic patterns within (Chapman et al. 1982; Avise et al. 1984, 1988; Avise 1995) and among (Neigel & Avise 1986) populations and species, and to address these expectations in a series of empirical mtDNA studies on a wide variety of organisms in nature (reviewed in Avise 1994). ‘Coalescent theory’ is the term now applied to the formal mathematical and statistical properties of gene genealogies (Kingman 1982; Watterson 1984; Donnelly & Tavaré 1986; Hudson 1990), and results from this discipline are highly relevant to molecular phylogeographic interpretations.

Several other important developments in the history of phylogeography are summarized in Table 1. In addition to these signal events, throughout the 1980s and 1990s there has been a burgeoning growth in the application of both genealogical theory and molecular data to phylogeographic analyses. This has included extensions and refinements of coalescent theory for populations of varying demographies (Hudson 1990; Slatkin & Hudson 1991; Rogers & Harpending 1992; Nee et al. 1995; Eller & Harpending 1996; see also several articles in this issue of Molecular Ecology), improvements in statistical and cladistic procedures for extracting phylogeographic information from empirical data on gene genealogies (e.g. Slatkin 1989; Neigel et al. 1991; Templeton et al. 1995; Templeton & Georgiadis 1996; this issue), and a great plethora of empirical applications primarily involving mtDNA (Fig. 3). Of course, progress in several related areas, not the least of which are molecular and computer technologies, have contributed significantly to the general scientific climate.

Table 1 Brief chronology of some of the important developments in the history of phylogeography*

<table>
<thead>
<tr>
<th>Date</th>
<th>Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974</td>
<td>Brown &amp; Vinograd demonstrate how to generate restriction site maps for animal mtDNAs</td>
</tr>
<tr>
<td>1975</td>
<td>Watterson describes some basic properties of gene genealogies, marking the beginnings of modern coalescent theory</td>
</tr>
<tr>
<td>1977</td>
<td>Upholt develops the first statistical method to estimate mtDNA sequence divergence from restriction digest data</td>
</tr>
<tr>
<td>1979</td>
<td>Brown, George &amp; Wilson document rapid mtDNA evolution</td>
</tr>
<tr>
<td>1980</td>
<td>Brown provides an initial report on human mtDNA variation</td>
</tr>
<tr>
<td>1983</td>
<td>Tajima and also Hudson initiate statistical treatments of the distinction between a gene tree and a population tree</td>
</tr>
<tr>
<td>1986</td>
<td>Bermingham &amp; Avise initiate comparative phylogeographic appraisals of mtDNA for multiple codistributed species</td>
</tr>
<tr>
<td>1987</td>
<td>Avise &amp; colleagues coin the word ‘phylogeography’, define the field, and introduce several phylogeographic hypotheses</td>
</tr>
<tr>
<td>1989</td>
<td>Slatkin &amp; Maddison introduce a method for estimating interpopulation gene flow from the phylogenies of alleles</td>
</tr>
<tr>
<td>1990</td>
<td>Avise &amp; Ball introduce principles of genealogical concordance as a component of phylogeographic assessment</td>
</tr>
<tr>
<td>1992</td>
<td>Avise summarizes the first extensive compilation, involving multiple species and genetic assays, of phylogeographic patterns for a regional fauna</td>
</tr>
<tr>
<td>1994</td>
<td>Moritz promotes the conceptual distinction between ‘shallow’ vs. ‘deep’ intraspecific phylogenies by introducing the terms ‘management units’ and ‘evolutionarily significant units’ (see also Ryder 1986; Riddle 1996).</td>
</tr>
<tr>
<td>1996</td>
<td>Volumes edited by Avise &amp; Hamrick, and by Smith &amp; Wayne, summarize the many roles for molecular phylogeographic analysis in conservation biology</td>
</tr>
<tr>
<td>1998</td>
<td>Interest in phylogeography continues to flower, as evidenced, for example, by this special issue of Molecular Ecology</td>
</tr>
</tbody>
</table>

*Particularly with regard to molecular and statistical sides of the field, and with due apologies to numerous other contributors whose works were important but due to space limitations cannot be included here.
that permitted the flowering of phylogeographic studies during the last two decades.

The future of phylogeography

What does the future hold for phylogeography? I suspect that the field is still in a rapid phase of growth (Fig. 1) and, as presaged by this special issue of *Molecular Ecology*, that many more empirical studies on diverse organisms can be anticipated. There will also be a further expansion of interest in the utility of coalescent theory as a formal conceptual thread for tying together more coherently the micro- and macroevolutionary disciplines of population genetics and phylogenetics.

More specifically, I see ample room for the expansion of phylogeography in three areas, each tied to a distinct aspect of ‘genealogical concordance’ (Avise 1996b). These three facets of concordance and their phylogeographic relevances are listed next, with brief descriptions of how I envision each as an exciting frontier for further research efforts.

Genealogical concordance, aspect ‘i’

Concordance in significant genealogical partitions across multiple unlinked loci within a species.

Relevance: helps to establish that the phylogenetic partitions in gene trees register deep as opposed to shallow historical partitions in an organismal phylogeny.

As already noted, the great majority of empirical genetic research into phylogeography has involved mtDNA (Fig. 3), yet the matrilineal pathways of ancestry registered by this molecule represent only a minuscule fraction of the total historical record within a sexual organismal pedigree (Fig. 4). In principle, much of the remainder of that history should be enconced in autosomal gene trees through which alleles have been transmitted via both genders. However, few attempts have been made to estimate nuclear gene genealogies in a phylogeographic context (Aquadro *et al.* 1991; Bernardi *et al.* 1993; Burton & Lee 1994; Palumbi & Baker 1994).

At least two complications, one technical and one biological, typically arise in attempts to recover genealogical information from nuclear genes. The technical difficulty is in isolating DNA haplotypes, one at a time, from diploid organisms at single-copy loci. Only with nuclear haplotypes cleanly separated can molecular assays such as DNA sequencing or restriction site mapping then be used to recover the phase (coupling vs. repulsion) of multiple DNA sequence variants, and thereby facilitate treatments of the alleles by phylogenetic procedures analogous to those conducted routinely for mtDNA haplotypes (which nature purifies). Avise (1994; p. 134) lists several experimental approaches for isolating haplotypes at particular nuclear loci. To this list can be added two recent PCR-based approaches (Hillis *et al.* 1996) that involve physical separation of nuclear haplotypes either by ‘DGGE’ (denaturing gradient gel electrophoresis; Lessa 1993) or by ‘SSCP’ (single-strand conformational polymorphism; Ortí *et al.* 1997).

The second complication in the recovery of nuclear gene trees at the intraspecific level is biological, and necessitates the presence of genomic regions that accumulate mutations rapidly yet are nearly free of intragenic recombination over the ecological or evolutionary timescales of interest. It remains to be seen how common such gene regions are, and how readily they can be identified and studied, but this certainly is an open frontier for meaningful inquiry.

On the conceptual front, a ‘multilocus coalescent theory’ is needed – one that considers the means, variances, and frequency distributions of genealogical information across multiple unlinked gene trees within an organismal pedigree. The development of such a theory might begin with consideration of the composite genealogical properties expected for independent neutral loci in random mating populations with specified demographic histories. Useful phylogeographic extensions then might involve populations spatially structured in various ways, and also might consider epistatic or otherwise non-neutral genes, or those that display partial linkage. The broader challenge will be to integrate the multilocus coalescent theory with empirical data to be gathered from multiple nuclear (and mitochondrial) gene genealogies within and among spatially arrayed natural populations.

Genealogical concordance, aspect ‘ii’

Concordance in the geographical positions of significant gene-tree partitions across multiple codistributed species.

Relevance: strongly implicates shared historical biogeographic factors in shaping the deeper intraspecific phylogenies, often on a regional scale.

Only a few empirical molecular studies have attempted comparative phylogeographic assessments within each of multiple codistributed species (Bermingham & Avise 1986; Avise 1992; da Silva & Patton 1993; Wenink *et al.* 1994; Turner *et al.* 1996; Zink 1996; Patton *et al.* 1997; see also several articles in this issue). Yet, such comparative assessments offer perhaps the greatest hope for significant advances in understanding how organismal behaviour, and the demographic and natural histories of populations, can influence intraspecific phylogeographic patterns. In this comparative light, findings of genealogical concordance will be of interest as reflections of shared community histories (Cracraft 1988). But findings of non-congruent genetic patterns will be valuable also because...
they may illuminate historical differences among species in levels of gene flow, responses to geographical barriers or selective gradients, rates of molecular evolution, effective population sizes, or other such molecular, ecological, and demographic factors (Avise et al. 1987b; Bowen & Avise 1990; Lamb et al. 1992; Zink 1996).

Genealogical concordance, aspect ‘iii’

Concordance of molecular gene-tree partitions with geographical boundaries between traditionally recognized biogeographic provinces.

Relevance: strongly implicates shared historical biogeographic factors as shapers of intraspecific organisinal phylogenies and species distributional patterns.

Study of the third aspect of genealogical concordance inevitably will draw molecular phylogeography into closer contact with other disciplines such as ecology, historical geography, and phylogenetic biology written large. Preliminary results with several faunas in the southeastern USA suggest that significant phylogeographic ‘breaks’ within species may tend to align geographically with traditionally recognized boundaries between biogeographic provinces as inferred from historical geological data, or from concentrations in the distributional limits of species (Avise 1996b). This suggests that historical factors influence not only species compositions in regional communities, but also the spatial distributions of genealogically distinct populations within species. Quite apart from the conceptual challenges motivated by such observations, the phylogeographic data themselves can be of great utility as summaries of the population genetic resources that conservation biology seeks to preserve (Avise & Hamrick 1996). Phylogeography as a recognizable discipline grew from recent historical roots in molecular genetic analyses of mtDNA, and in mathematical studies of coalescent processes that seemed necessary to capitalize upon this new class of genealogical information within species. However, the full impact of phylogeographic thought remains to be realized in the broader biological and geographical sciences. Phylogeography has had an auspicious start. The greatest benefits and opportunities for the field will continue to arise, as they have in the past, from phylogeography’s central, integrative position within the evolutionary and ecological sciences.

Acknowledgements

Work from the Avise laboratory has been supported by the NSF. I wish to thank Drs Wes Brown, Richard Hudson, Warren Ewens, James Crow for sharing with me their thoughts on matters of history.

References


Avise JC (1992b) Molecular population structure and the biogeographical provinces.

References


Work in the Avise laboratory, conducted by students and postdoctoral researchers, involves the application of molecular genetic markers to questions in natural history, ecology, and evolution. Avise himself spends most of his time dreaming up verbage to make it all seem more important.