Phylogeography

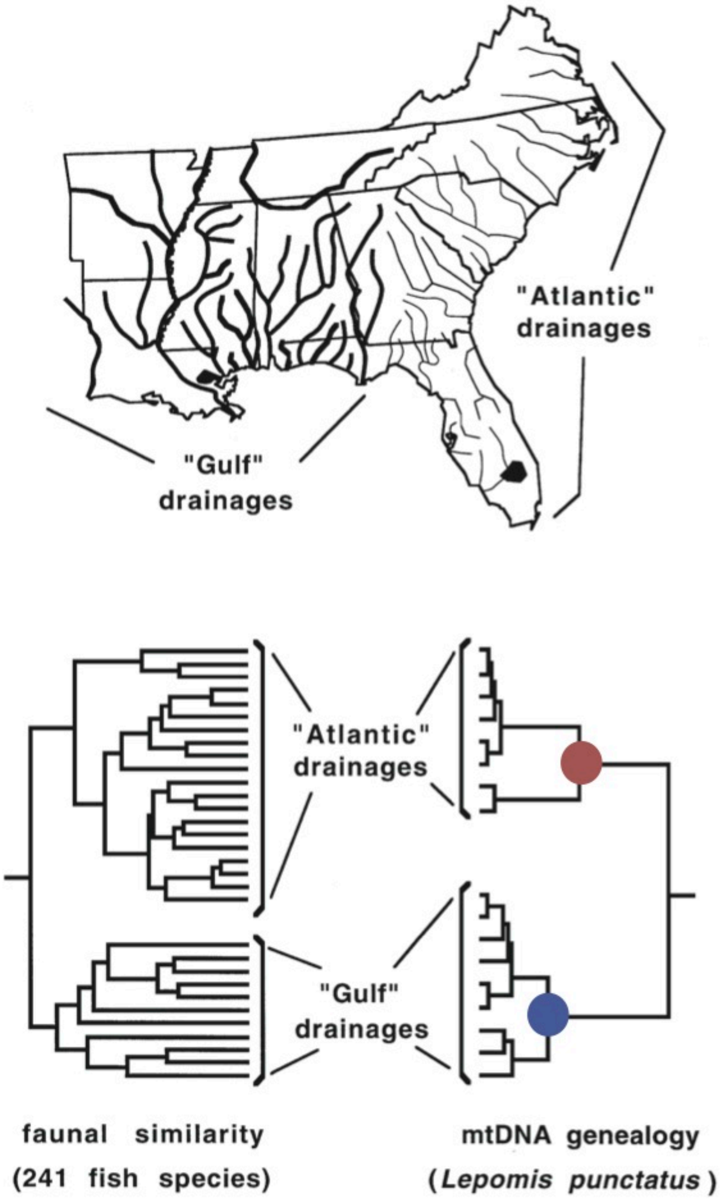
The field of phylogeography developed in the 1980s along with the increased availability of DNA sequence haplotype data. These data were analytically suitable for using phylogenetic analytical techniques, but were variable enough to address intraspecific questions. These questions include ones of taxonomic and/or demographic limits, migration patterns, and how abiotic environmental changes influence biotic communities. Although early phylogeography studies relied on qualitative analysis of the phylogenetic relationship of individuals, the field has become increasingly statistical in nature and in many ways has returned to the quantitative models developed by population geneticists.

John Avise coined the term “phylogeography” in 1987[1] to describe the use of phylogenetic approaches to evaluate genetic polymorphism data in natural populations and the distribution of this variation across a geographic landscape. The term derives from combining ‘phylogeny’ (the study of the evolutionary relationships of organisms) and ‘biogeography’, a field that typically studies the distribution of species across the global landscape. The field of phylogeography initially distinguished itself in some ways from population genetics through an increased reliance on inference from the distribution of individual alleles across geographic space, rather than solely the aggregate statistics made from pooled samples of individuals at distinct locations (i.e. comparisons of allele frequencies among sampled populations).

The focus of the phylogeographic approach has been on reconstructing the geographic history of a species or group of species. For example, many early studies in this field would assess whether geologic or other forms of large-scale global change were associated with spatial patterns of intraspecific diversity. In more recent years, these approaches have been used to assess the degree to which intraspecific diversity is associated with the distribution of particular habitats or ecological interactions through time. In particularly integrative efforts, this association is strengthened through coordinated analysis of genealogical data from regions that have been recognized as refuges (for example, during the last glacial maximum) based on reconstructions of past climates [2-4].

While the initial insight from phylogeography was that many taxa are comprised of spatially and genetically distinct subgroups, the temporal component of differentiation is clearly important. First of all, researchers recognized that differentiation takes time. The most straightforward phylogeographic inference is using genetic data to identify a pattern known as reciprocal monophyly[5]. ‘Monophyly’, in systematics terminology, is a phenomenon in which a group of organisms or lineages have a single common ancestor, and all descendants of that ancestor are in the group (thus the grouping is based on homology). Reciprocal monophyly results from situations where two or more distinct groups (whether the distinguishing feature is geographic or some other phenotypic trait) are each characterized by independent evolutionary histories. It is diagnosed when the distinct groups are separated by ‘most recent common ancestors’ that are unique to each group, and all descendants of that ancestral lineage are found in the group (Figure 1).

Figure 1. Early applications of phylogeography recognized that faunal patterns may be reflected in intraspecific divergence of alleles. Here, Atlantic and Gulf drainages of spotted sunfish exhibit reciprocal monophyly; the ancestral nodes that are distinct for the two regional populations are shown in red and blue. From Walker and Avise (1998).

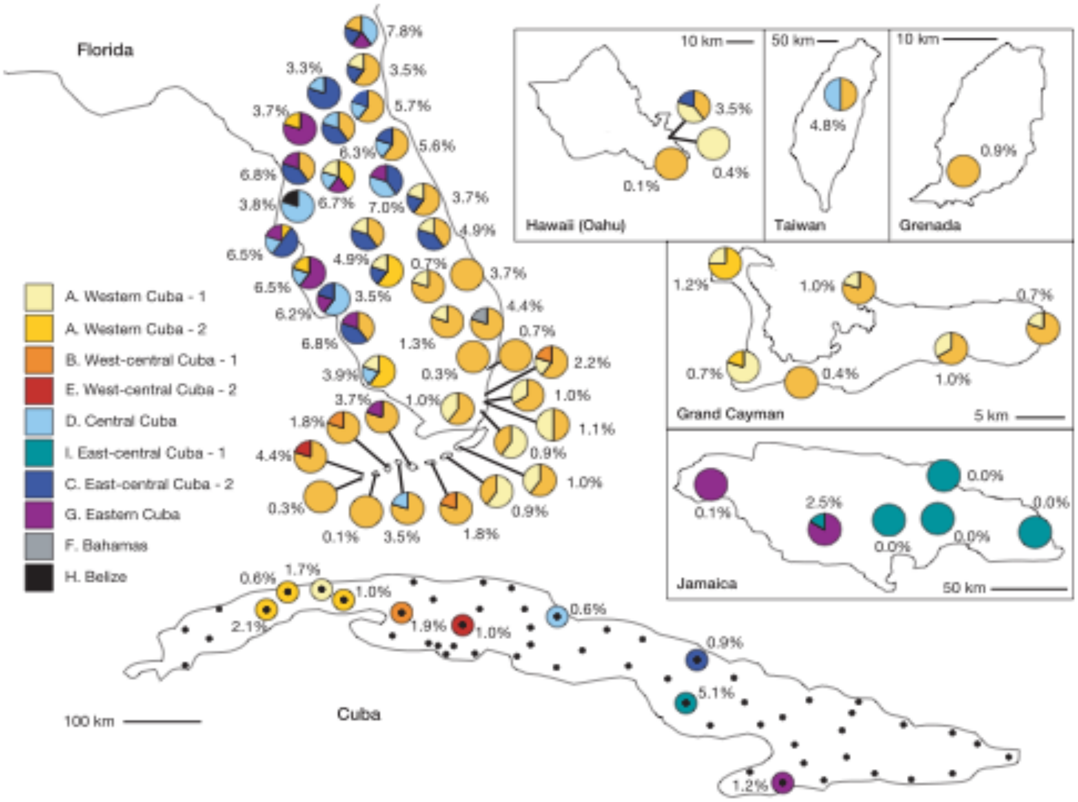


Theory and modeling work in the 1980s showed that reciprocal monophyly for two groups requires significant demographic isolation - limited migration and gene flow, over a number of generations that is proportional to the *effective population size* of the groups. The effective population size (abbreviated Ne) is the number of individuals that, fitting certain assumptions about reproduction (random mating, equal gender ratio, unbiased reproductive output, etc.), experiences the same level of genetic drift (changes in allele frequency from generation to generation) as the observed level in an empirical population. Thus, if the effective population size is large, the time required before genetic drift leads to a pattern of reciprocal monophyly must be large; theory and simulations show that, on average, the time it takes for two populations, isolated from a single ancestral population, to reach reciprocal monophyly is approximately 4Ne generations.

It is not sufficient to use reciprocal monophyly to illustrate that significant time has passed since there was demographic contact between the populations, particularly if the concordant separation of whole communities, or the congruence of this separation with a particular environmental event, is of interest. The genetic variation within each population under comparison can be used to estimate Ne for each population as well as for the ancestral population from which they are derived. A hallmark of the statistical approach that has been taken in the field of phylogeography in the last decade is the use of these diversity measures within and across "taxon pairs" (the reciprocally monophyletic groups for which the divergence time is in question[6]) to simultaneously recognize Ne and what that diversity says about the likely actual divergence time. For example, many North American birds are separated into eastern and western species or subspecies (e.g., the red-shafted and yellow-shafted northern flickers). Modern statistical approaches have shown that the variance in depth of gene trees (across loci and across species) is too great for a single environmental change to be responsible for the initial isolation of eastern and western populations[7].

This increased focus on combining inference from the spatial pattern of the gene tree as well as the statistical properties of that gene tree is called 'statistical phylogeography'[8]. Remarkably, this means that the gene tree itself, as reconstructed from what is often a minute subsample of an organism's genome, is considered to be a 'nuisance parameter'; in other words, phylogeography now relies less on the algorithms and criteria used for visualizing organismal relationships, and more on the statistical models that apply to intraspecific reproduction, migration, mutation, and other evolutionary processes that describe the likely demographic history of a population, given these empirical data. This statistical focus has also enabled researchers to pry into more subtle ecological and distributional shifts. Given the temporal constraints of reciprocal monophyly, this means that more recent transitions in the ancestry of a population or populations can be examined, and even circumstances for which an equilibrium is maintained between population size fluctuations, biased migration, and other phenomena. One of the more intriguing applications of this approach in recent years includes high-resolution reconstruction of how humans migrated to the Americas. This work showed that an extremely small migrant population was probably responsible for all New World diversity prior to migration from Europe[9].

Figure 2. Phylogeographic methods may be used to identify differentiable parts of the native range of a species, and then use this information to 'assign' individuals of unknown source. Here, individuals of *Anolis sagrei* in Florida were evaluated for the likely pathways of introduction from Caribbean populations. From Kolbe et al. (2004).



Work on lizards in the genus *Anolis* can illustrate the diversity of phylogeographic research. Sequence data from a mitochondrial gene show that there is tremendous phylogenetic diversity among populations of *A. sagrei* throughout the Caribbean[10]. In some cases, populations from different parts of the same island can be distinguished based on the mitochondrial sequence diversity recovered. When individuals from a species are introduced to a new location, or when dispersing individuals are found outside of their known habitat, these phylogeographic approaches can be used as a form of *assignment test* to indicate the likely source population for those individuals. With *A. sagrei*, this information was used to indicate that the invasive populations of this anole in Florida are the result of introductions from numerous geographic sources (Figure 1).

The dispersal and introduction of *A. sagrei* may not be wholly human-mediated. Another study on this species used the 'statistical phylogeography' approach for reconstructing likely trends of gene flow throughout the Caribbean. This approach involves collecting empirical data (in this case, microsatellites - one key feature of phylogeographic analysis being the use of data that have clear mutational models and the opportunity to polarize mutational events on gene trees) and searching a range of gene trees with varying parameters of effective population size and migration rates among sample locations for the set of parameters that best describes the empirical data. Recent work showed that dispersal of lizards among these islands is strongly mediated by storms and prevailing ocean currents between islands[11]; the opportunity to evaluate asymmetric dispersal is a particularly interesting advance in the phylogeographic exploration of natural populations.

Certainly one of the more important aspects of identifying distinct demographic units within a species (*i.e.*, monophyletic clades) is the use of these data in conservation. Although phylogeographic criteria are not applied equally across all taxa under the Endangered Species Act or other protective legislation, 'distinct population segments' of vertebrate species can be identified by the presence of monophyletic clades within a species' range. Reciprocal monophyly at mitochondria loci alone is generally sufficient to define 'management units', while 'evolutionarily significant units' are generally recognized when there is corroborating information from the nuclear genome as well[12]. Some authors have explored the use of phylogenetic diversity within a species (or among closely related taxa) as one criterion for prioritizing populations when management resources are limited.

Overall, the discipline of phylogeography has transformed our understanding of biodiversity. It has spawned new ways of recognizing and naming biodiversity[13], associating this diversity with global environmental change, and elevating new expectations for what baseline information is necessary to evaluate variation in ecological interactions and the potential for local adaptation. This field has been transformed once already, from a qualitative assessment of phylogenetic patterns toward a statistical assessment involving model fitting and an understanding of the inherent stochasticity of demographic processes. It is currently being transformed again by ready availability of inexpensive next-generation sequence data, which allows almost any species to have a thorough assessment of genetic variation among geographic samples. It is at this point that the demarcation between genomics, population genetics, and phylogeography begins to be less useful, and we see that information from all of these fields are necessary for full exploration of the mechanisms that generate and maintain biodiversity.

1 Avise, J. C. *et al.* Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst* **18**, 489-522 (1987).

2 Waltari, E. *et al.* Locating Pleistocene Refugia: Comparing Phylogeographic and Ecological Niche Model Predictions. *PLos One* **2**, e563 (2007).

3 Sumner, J., Jessop, T., Paetkau, D. & Moritz, C. Limited effect of anthropogenic habitat fragmentation on molecular diversity in a rain forest skink, Gnypetoscincus queenslandiae. *Mol. Ecol.* **13**, 259-269 (2004).

4 Petit, R. J. *et al.* Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**, 1563-1565 (2007).

5 Avise, J. C. *Phylogeography*. (Harvard University Press, 2000).

6 Hickerson, M. J. & Meyer, C. P. Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *Bmc Evolutionary Biology* **8**, -, doi:Artn 322 Doi 10.1186/1471-2148-8-322 (2008).

7 Edwards, S. V. & Beerli, P. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839-1854 (2000).

8 Knowles, L. L. The burgeoning field of statistical phylogeography. *J. Evol. Biol.* **17**, 1-10 (2004).

9 Hey, J. On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biology* **3**, e193 (2005).

10 Kolbe, J. J. *et al.* Genetic variation increases during biological invasion by a Cuban lizard. *Nature* **431**, 177-181 (2004).

11 Calsbeek, R. & Smith, T. B. Ocean currents mediate evolution in island lizards. *Nature* **426**, 552-556 (2003).

12 Moritz, C., Brown, Smith, M. & Nee. Vol. New Uses for New Phylogenies (ed Harvey) 203-214 (1996).

13 Pringle, A., Barron, E., Sartor, K. & Wares, J. P. Fungi and the Anthropocene: Biodiversity discovery in an epoch of loss. *Fungal Ecology* **4**, 121-123 (2011).