



HAL
open science

Phylogeography

Inessa Voet, Violaine Nicolas

► **To cite this version:**

Inessa Voet, Violaine Nicolas. Phylogeography. Eric Guilbert. Biogeography: an integrative approach of the evolution of living, 1, Wiley; ISTE Science Publishing, pp.59-76, 2021, 9781789450606. <10.1002/9781119882381.ch3>. <hal-03468408>

HAL Id: hal-03468408

<https://hal.science/hal-03468408v1>

Submitted on 3 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Phylogeography

Inessa VOET

Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle

Violaine NICOLAS

Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle

4.1. Introduction

Sometimes a novel word encapsulates a concept and becomes part of the working lexicon of a discipline. Such was the case for “phylogeography”. This word was introduced by Avise in 1987 in his work “Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics” (Avise *et al.* 1987). Phylogeography endeavors to understand the processes that underlie the geographical distribution of genetic variation within and among closely related species. Phylogeography takes a population genetics and phylogenetic perspective on biogeography. In the early days of phylogeography, scientists used mitochondrial (mt) DNA to address precisely how conspecific individuals were genealogically linked through shared ancestors. In genetic surveys of various species, striking patterns were being uncovered in the spatial arrangements of mtDNA lineages. In other words, genealogy and geography seemed to be connected. After 1987, ‘phylogeography’ became the convenient shorthand for referring to explicit genealogical inquiries (including those for nuclear genes) into the spatial and temporal dimensions of microevolution. The unwavering popularity of phylogeography is indisputable (Fig. 4.1).

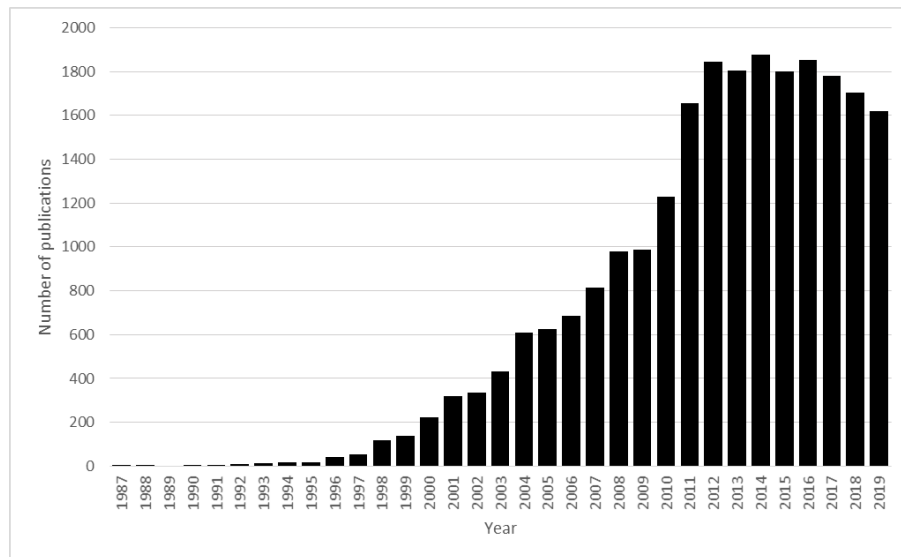


Figure 4.1. Number of publications with the word “phylogeography” in the “topic” per year, according to Web of Science

Phylogeography deals with historical, phylogenetic components of spatial distributions of gene lineages. It thus provides a powerful empirical and conceptual bridge between microevolutionary (population genetics) and macroevolutionary (phylogeny, historical biogeography) disciplines. Phylogeography describes how genetic variation, introduced by mutation, is geographically structured within and between populations by population genetic processes, such as drift, gene flow, selection and recombination. For populations that have been separated historically and have experienced little or no gene flow, genetic differences can accumulate by these evolutionary processes, potentially resulting in speciation.

Phylogeography is sometimes considered as a sub-discipline of biogeography. Phylogeographic studies provide detailed species-specific information on how geologic events, environmental influences, and geographic factors interact with aspects of a species ecology and natural history in shaping its evolution. Phylogeographers seek to interpret extent and mode by which historical processes in population demographics and landscape alterations may have left evolutionary footprints on the contemporary geographic distributions and phylogenetic signature of gene-based organismal traits. Thus, phylogeography serves to expand and to balance the perspectives of ecogeography (Avise *et al.* 1987; Avise 2000).

The aim of this chapter is to explain what phylogeography is and to present methodological and conceptual developments in this field.

4.2. The early days of phylogeography: cytoplasmic genomes and qualitative post hoc explanations of historical processes

During the first decades of phylogeography, most studies involved analyses of exclusively animal mtDNA. The wide use of this cytoplasmic genome in phylogeographic studies is explained by the fact that it has several phylogenetically favorable properties (Avice *et al.* 1987; Avice 2000, 2009): 1) it occurs in all animals, homologous comparisons can thus be made among a wide variety of organisms; 2) it is easy to isolate (most somatic cells contain hundreds or thousands of mtDNA molecules, and heteroplasmy is uncommon); 3) it has a simple genetic structure lacking complicated features such as repetitive DNA and introns, and with many genes being free of indels; 4) due to its maternal inheritance (mtDNA inheritance is both haploid and asexual) it exhibits a straightforward mode of genetic transmission, without recombination or other genetic rearrangements suffered by the nuclear DNA; 5) although some coding segments are highly conserved, many mitochondrial genes evolve at a rapid pace. Therefore, genetic disparity between diverging lineages accumulates rapidly, and can be detected on a relatively short temporal scale. MtDNA genotypes are referred to as haplotypes, which differ from one another by particular mutations accumulated since female ancestors were last shared. Because of the rapid pace of mtDNA evolution, many different haplotypes typically coexist within a species. These haplotype sequences can be used to estimate the matrilineal histories of individuals and populations.

Plants have two cytoplasmic genomes: mtDNA and chloroplastic DNA (cpDNA). Contrary to animals, plant mtDNA is poorly suited for phylogeographic studies (Avice 2009) within species because of its slow pace of sequence evolution (about 100X slower than in animals) and several other technical hurdles (e.g. inheritance is not always maternal, gene order rearrangements are common, and plant mtDNA is highly variable in size). In the early days of plants phylogeography, most studies were conducted on cpDNA because, like mtDNA in animals, it has several phylogenetically favorable properties (Avice 2000): 1) cpDNA typically occurs in many copies per cell, and most individuals display a single cpDNA haplotype sequence; 2) several chloroplastic genes evolve at a relatively rapid pace and various individuals within a species can show detectable differences in cpDNA

sequences, 3) it is haploid and non-recombining, even if the transmission can be matrilineal, patrilineal or a mixture of the two depending on the mode of genetic transmission of cpDNA in the species being investigated.

Using mtDNA, Avise (Avise *et al.* 1987; Avise 2000) showed that intraspecific phylogenies overlaid on geographic maps can lead to five major categories (Fig. 4.2) of possibilities, depending on the combination of phylogenetic discontinuity (or genetic “breaks”: arrays of related genotypes differing from other such arrays by many mutational steps) and spatial separation levels. The authors provided interpretations for these five categories:

- I - Phylogenetic discontinuities with spatial separation: genetically divergent populations occupy separate geographic regions within the range of the species. This situation can be the result of long-term extrinsic (i.e. zoogeographic) barriers to gene flow, or of the extinction of intermediate genotypes in widely distributed species with limited dispersal and gene flow capabilities.
- II - Phylogenetic discontinuities with lack of spatial separation: genetically divergent populations co-occur geographically. This can be the result of recent secondary admixture zones or intrinsic (e.g. reproductive isolation) barriers acting among sympatric sibling species. Phylogeographic breaks can also arise in continuously distributed species if the average individual dispersal distances and/or population size of the species are low (Irwin 2002).
- III - Phylogenetic continuity with spatial separation, which involves historically limited gene flow between populations in a species not subdivided by firm long-term zoogeographic barriers to dispersal.
- IV - Phylogenetic continuity with lack of spatial separation which arise when there is very extensive gene flow in a species not subdivided by long-term zoogeographic barriers.
- V - Phylogenetic continuity with partial spatial separation: in this category some mtDNA haplotypes (the ancestral ones) are geographically widespread, while allied haplotypes (newly arisen mutations) are localized. This pattern can appear in taxa with historically intermediate levels of gene flow between geographic populations.

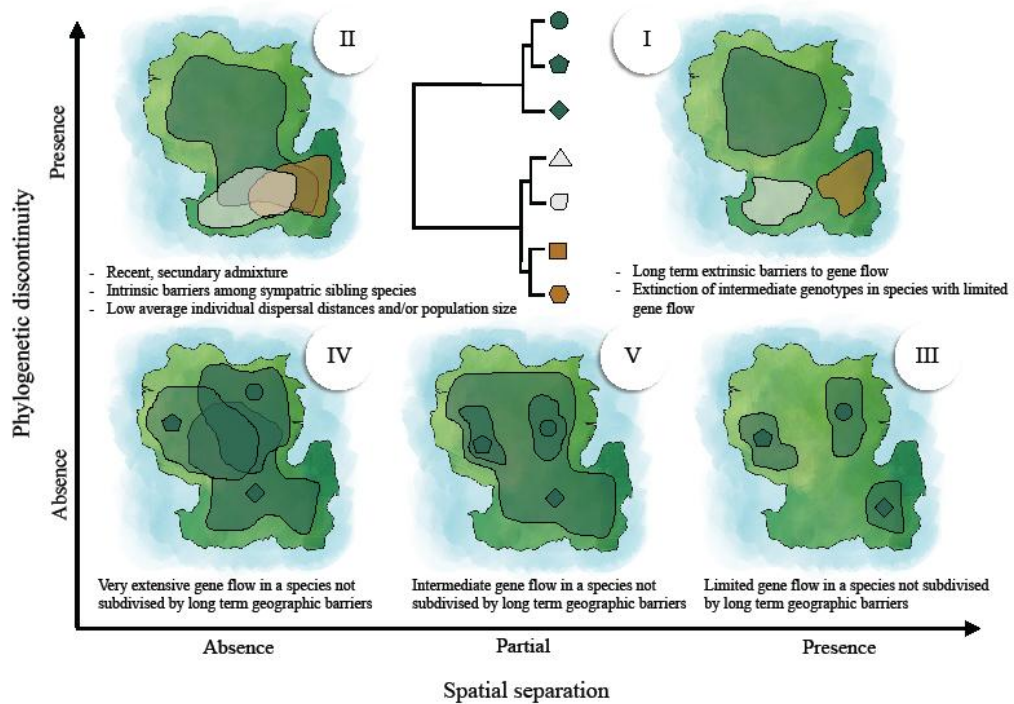


Figure 4.2. Schematic representation of the five categories of phylogeographic patterns theoretically observable in mtDNA surveys.

4.3. Statistical phylogeography

In the early days of phylogeography, mtDNA or cpDNA gene trees were used for demographic inference. However, there is an inherent contradiction in using gene trees as a literal interpretation of a species past, because gene trees contain historical signal, but also stochasticity: there is a random loss of gene lineages by genetic drift because of the different number of descendants individuals can or will produce. Phylogeography has undergone both marked methodological and philosophical transformations in how genetic data are used for demographic inference. Because many events may have occurred in the past (e.g. bottlenecks, population expansion, migration and vicariance), and gene lineages may be lost by chance, a species history might not be easily inferred from a gene genealogy. In the

early 2000s, several researchers criticized the narrative nature and lack of statistical rigor of phylogeography. They emphasized that it was important to consider the stochasticity of population genetic processes and to explicitly assess the confidence of various phylogeographic conclusions. There are several reasons why historical inferences derived directly from a gene genealogy may be inaccurate or misleading, which are not mutually exclusive (reviewed by Knowles 2004): 1) the stochastic process of lineage sorting can produce discord between the population's and gene's history; 2) the actual history can be obscured by deep coalescence of gene lineages; 3) the resolution of the genetic marker can be insufficient for recovering the population history from a gene tree estimate; 4) the gene genealogy can reflect the action of selection rather than the population's history; 5) alternative hypotheses can be indistinguishable because of high stochastic variance of trees of independent genes. A transition occurred from describing to testing the processes underlying patterns of genetic variation. This was concomitant with a shift in the way historical inferences were made, namely the explicit consideration of stochastic variance and reference to predictions based on models that are defined *a priori*. This was the rise of what is called "statistical phylogeography" (Knowles and Maddison 2002; Knowles 2004, 2009). Inferences about a species' history are now based on statistical tests of historical hypotheses and estimates of demographical parameters. A wide range of historical scenarios can be modelled, from very simple to very complex, including different geographical scenarios, timing of population divergence, and estimates of effective population sizes, migration, population admixture and growth.

Statistical phylogeographical methods generally fall into one of two categories: those that take a summary-statistic approach vs. likelihood or Bayesian analyses. In likelihood analyses the probability of observing the data under alternative models is calculated, while in the summary-statistic approach a simple statistic summarizes the data. Summary-statistic approaches are easy to implement, but their interpretations are sensitive to the defined models and different population histories can sometimes 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. Knowledge of the biological system under study is essential for deciding which approach to use for phylogeographic analyses. What might be an appropriate method for one study might not be for another, depending on the specifics of each species history and the type of data collected.

Each gene has its own history, and gene trees can be different from the species tree. The analysis of multiple unlinked loci is critical for accommodating coalescent stochasticity and improving the accuracy of inferences about demographic history and estimates of divergence times (Knowles 2004). Single-copy nuclear genes can provide voluminous sequence data for phylogeographic assessments at the intraspecific level. However, in sexually reproducing organisms, meiosis and syngamy continually re-assort unlinked genes into new multi-locus combinations. These recombinational processes can also manifest within a locus whenever successful meiotic crossover events occur between different alleles. Each recombinant allele that emerges is then an amalgamated stretch of DNA, the subsets of which might have quite different evolutionary pasts. Thus, intragenic (or inter-allelic) recombination tends to complicate and garble what would otherwise be cleaner genealogical signatures from the nuclear genome. This contrasts with the historical clarity characteristic of the nonrecombining cytoplasmic genomes. Three technical and biological hurdles have for a long time impeded progress (Avice 2009): 1) the relatively slow pace of sequence evolution at many nuclear loci; 2) the difficulty of isolating nuclear haplotypes, one at a time, from diploid organisms; and 3) the phenomenon of intragenic recombination. The first complication was first partly overcome by sequencing single-copy nuclear sequences that evolve more rapidly than average (e.g. intron sequences instead of exon sequences). The second complication was overcome by sequencing genes that are present in haploid conditions in each member of the heterogametic sex (e.g. X- and Y- chromosomes of mammals, Z- and W- chromosomes of birds) or by resolving allele ambiguities in diploid conditions through cloning or via computation using software like PHASE (Stephens *et al.* 2001; Stephens and Donnelly 2003). Generating such multilocus data was time expensive: most practitioners of biogeography have spent a significant portion of their time developing and screening molecular markers suitable to their study system and appropriate to their evolutionary timescale of interest (i.e. with enough variability), amplifying and sequencing DNA for each sample at each locus, and phasing nuclear data.

The development of next-generation sequencing (NGS) was very appealing in the field of phylogeography as it could have the ability to condense the many steps of multilocus data generation for non-model organisms into a more time-efficient and cost-effective process. However, NGS has been slow to take root in phylogeography compared to other fields like metagenomics and disease genetics. According to McCormack *et al.* (2013) this lag has been caused by the predominant focus on non-model organisms and the need for sequencing large numbers of samples per species. Obtaining whole-genome data for hundreds of individuals is still too costly, thus a method of genome size reduction is required. Phylogeography requires homologous genomic regions from multiple individuals to infer gene

genealogies. In traditional Sanger sequencing this is straightforward through the design of PCR primers specific to a particular locus. With NGS, DNA is not necessarily enriched for a single locus via PCR-based amplification, although this is one possible application, but for many loci through a variety of methods involving reduction of the size of the genome. One of the main drawbacks of NGS methods is that there is little control over exactly what regions of the genome are sequenced. On the other hand, the number of targets (and the number of reads associated with each target) is increased by orders of magnitude. Using NGS for phylogeography is only cost-effective if many individuals can be combined (multiplexed) in the same sequencing run and the costs subdivided among many samples. To this aim short identifying DNA sequences (called “indexes”, “barcodes”, or “tags”) need to be incorporated into the DNA fragments either by PCR or ligation. These tags identify an individual prior to pooling with other tagged samples. Sequences are later sorted using bioinformatics. Several reviews of the different NGS sequencing platforms and chemistries already exist, and we invite interested readers to read these papers (Shendure and Ji 2008; Glenn 2011). Methods that reduce the genome by restriction digest and manual size selection are particularly appropriate for studies at the intraspecific level, whereas methods that target specific genomic regions (i.e. target enrichment or sequence capture) have wider applicability from the population level to deep-level phylogenomic (McCormack *et al.* 2013). An overview of how to analyze NGS data to arrive at data sets applicable to the standard toolkit of phylogeography, including initial data processing to alignment and genotype calling can be found in McCormack *et al.* (2013). Most existing NGS technologies produce short reads and therefore are best suited for generating single nucleotide polymorphisms -SNPs-, not whole loci featuring many linked SNPs. This as for a long time limited the application of NGS data with respect to the standard analytical tool kit of phylogeography that rely mostly on gene trees. Individual SNPs have low information content on a per-locus basis and have been predominately used with classification methods such as Structure (Pritchard *et al.* 2000), principal component analysis (e.g. Dufresnes *et al.* 2020) or K-means clustering (Myers *et al.* 2020). However the used of SNPs for phylogenetic inference is growing rapidly (reviewed by Leaché and Oaks 2017). A fundamental difference between genome-wide-SNP phylogenetics and traditional approaches that use gene sequences is that DNA sequences are linear strings of nucleic acids composed of constant and variable characters, while SNP data are composed exclusively of variable characters, either as a direct result of the data collection strategy, or after bioinformatic steps removing the constant characters. This can lead to extreme overestimation of evolutionary rates and inaccurate topologies, branching times, and support values (Leaché *et al.* 2015). Another important problem with the use of SNPs is that the amount of missing data increases with the number of loci and sample size, due to allelic dropout. More work is needed to evaluate and select missing data thresholds

that avoid negative consequences for phylogenetic analyses (Leaché and Oaks 2017). Several methods for estimating SNP phylogenies can be used. Data concatenation, in which SNPs are put together end-to-end to construct one exceptionally long super locus, is problematic because it ignores incomplete lineage sorting, and in doing so assumes that all SNPs share the same coalescent history. If concatenation is used, studies suggest it is better to apply this method to the original sequence alignments from which the SNPs were extracted: removing constant sites results in acquisition bias that can inflate branch length estimates and in extreme cases produce an inaccurate phylogeny (Leaché *et al.* 2015). Methods for estimating species trees in the presence of incomplete lineage sorting have been developed. The most popular of those methods are summary methods like ASTRAL (Mirarab *et al.* 2014) and ASTRAL-2 (Mirarab and Warnow 2015). Several other methods were recently developed, such as SNAPP (Bryant *et al.* 2012), PoMo (De Maio *et al.* 2015) and SVDquartets (Chifman and Kubatko 2014).

To conclude, phylogeography relied heavily on non-recombining and rapidly evolving mtDNA or cpDNA to match gene genealogies with geography, but the advent of genomic data in combination with the development of coalescent theory has revolutionized the field. In general, the application of genome-wide approaches uncovers detailed population structure that is often missed by traditional markers such as mtDNA. In addition, applying phylogeographic methods to genome-wide datasets has shown to what a great extent different genes often result in different gene trees. This phylogenetic incongruence can provide a more detailed picture of population history because different gene trees capture particular historical events and population genetic processes that have shaped the present patterns of genetic diversity. However, recent work has also uncovered high levels of reticulation due to recombination and gene flow (Ottenburghs *et al.* 2019). New statistical methods are being developed to deal with such reticulated scenarios.

4.3. Comparative phylogeography

Comparative phylogeography emerged three decades ago as an integrative approach to historical biogeography (Arbogast and Kenagy 2001). Comparative phylogeography can be defined as the study of the effects of evolutionary history and biogeography on the distribution of genetic variation in co-distributed species. If spatially consistent genealogical splits appear across multiple co-distributed species, the responsible evolutionary forces must have had widespread effects at the level of

biotic communities or ecosystems. Comparative phylogeographic studies can reveal ways in which entire communities or assemblages are structured by shared responses to a common past and unique species-specific historical events. Thus they can contribute to broader studies of ecology and evolution in a number of ways (Bermingham and Moritz 1998): 1) Phylogeographic analysis can identify historically and evolutionarily independent geographic regions that can be considered as natural replicates amongst which generalizations about specific processes can be tested statistically. For example, the evolutionary response to selection gradients can be compared across different historical isolates. 2) Phylogeography can provide an evolutionary and geographical context for the species comprising ecological communities, thus permitting determination of historical and spatial influences on patterns of species richness. 3) Understanding of historical responses to changes in the landscape and the identification of evolutionarily isolated areas can inform conservation strategies.

In general, a comparative phylogeographic study has two phases (Gutiérrez-García and Vázquez-Domínguez 2011). It begins with a descriptive survey that includes the collection of genetic data from the species and phylogenetic analyses of that data. The intraspecific data obtained at this phase provide insight into phylogenetic relationships of populations, genetic diversity and genetic distance between lineages, effective population size, gene flow, and the timing of diversification events. The second phase is the comparative component and includes diverse analyses to test for congruence between the evolutionary and distributional histories of each species; it also includes evaluation of the geographical, ecological, and biological hypotheses that could explain those histories. The aim at this phase is to evaluate whether the evolutionary histories of these species show a shared response to the same historical events.

Since its beginning comparative phylogeography uses concordance in patterns of genetic variation as a criterion for evaluating hypotheses. Phylogeographic congruence among several co-distributed taxa has provided key evidence of the impact of geological events, biogeographic barriers or past environmental change on today's distribution of genetic variation (Avice 2000; Hewitt 2004a, b). Comparative phylogeographic analyses have been used to address a broad array of ecological and evolutionary questions, from inferring Quaternary refugia and assigning conservation priorities (Marske *et al.* 2013; Dauby *et al.* 2014; Médail and Baumel 2018) to evaluating the stability of ecological associations between interacting

organisms (Stone *et al.* 2012). Commonly applied tests include the evaluation of spatially concordant phylogeographic breaks, concerted demographic expansion/contraction among members of an ecological community, or temporally synchronous population divergences (reviewed by Papadopoulou and Knowles 2016). Methodological advances of the last decade, especially coalescent-based tools for hypothesis testing and parameter estimation that include statistical assessment of concordance across taxa, such as the widely used hierarchical approximate Bayesian computation (Hickerson *et al.* 2006; Huang *et al.* 2011), also promoted the utility of the concordance criterion.

Rigorously testing the congruence of phylogeographical patterns among taxa (as a result of a common evolutionary history) is no easy task, considering that even sister species can differ in their ecological characteristics and life-history features. Species-specific demographic and ecological characteristics can result in different outcomes of the same historical and biogeographical events and, consequently, in the loss of concordant phylogeographical patterns between species. Therefore, in the planning of a comparative phylogeographic study, the information obtained through intraspecific phylogeographic evaluations is of fundamental value. It is also important to consider, among other things, the stochastic variation of loci; the degree of taxonomic resolution in the species of interest; their phenotypic characteristics (ecology, traits that limit or promote movement, traits having an impact on demography); together with fossil evidence and biogeographical, geological, and historical information relating to the geographical distributions being studied.

The term “Comparative phylogeography” was also recently used not only for comparing co-distributed species (i.e. in a spatial sense), but also in at least three other senses – (i) genomic – sharing phylogeographic patterns across multiple loci within a species, (ii) taxonomic - emphasizing how phylogeographic findings have impacted ecological and evolutionary thought in a diversity of organismal groups, and (iii) conceptual - addressing the place of phylogeography in relation to various allied disciplines in the biodiversity sciences (Avice *et al.* 2016).

4.4. Integrative studies

The need to incorporate other sources of data corroborating the genetic data used in phylogeographic tests was acknowledged from the beginning of the development of statistical phylogeography (Knowles 2004, 2009). Fossil data are commonly used as an independent source of data (e.g. Gavin *et al.* 2014; Lalis *et al.* 2016; Lalis *et al.* 2019). However, such data are not available for all species.

4.4.1. Integration of Ecological Niche Modeling in phylogeographic studies

Species distribution modelling techniques (SDMs; also known as ecological niche models or ENMs) generate information about abiotic preferences and tolerances of species, and hence estimates of the current, past and potential future distribution of species. ENMs employ independent data sources that can be used to evaluate or develop phylogeographic hypotheses about the processes generating patterns of genetic variation in disparate taxa (Carstens and Richards 2007; Richards *et al.* 2007; Alvarado-Serrano and Knowles 2014). For generating alternative biogeographical hypotheses, the necessary data consist of 1) a set of georeferenced localities that describe where the species has been documented to occur, 2) a set of GIS layers containing information about the pertinent aspects of the current environment for the geographic area and species of interest, and 3) for the case of palaeodistributions, a second set of GIS layers describing an estimate of the environment at a particular time period of interest in the past (Fig. 4.3). Using these inputs and a species distribution modelling algorithm, both the current and past distributions of the focal species can be estimated. These estimates of a species' past distributions, or palaeodistribution models, can then be used directly to explain the obtained genetic structure in a correlative (or corroborative) manner, or to guide the generation of alternative biogeographical hypotheses (or predictions) that can be tested with genetic data (Fig. 4.3).

Increased availability of species distribution data on the Internet (e.g. GBIF, VERTNET), as well as increased availability of climatic or environmental variables for both current (e.g. Worldclim, SRTM 90m DEMs, HYDRO1k) and past data (e.g. WorldClim data for the last glacial maximum or last interglacial); and development of easy-use algorithms (e.g. Maxent) explain the rapid development of ecological niche modeling approaches and their use in phylogeographic studies. A rapid search in the Web of Science database with the words “phylogeography” and “Ecological Niche Modeling” in the “topic” shows that the first papers combining these two approaches were published in the early 2000s, and increased considerably after 2010. During

the last five years this topic combination has been represented by almost 100 papers per year.

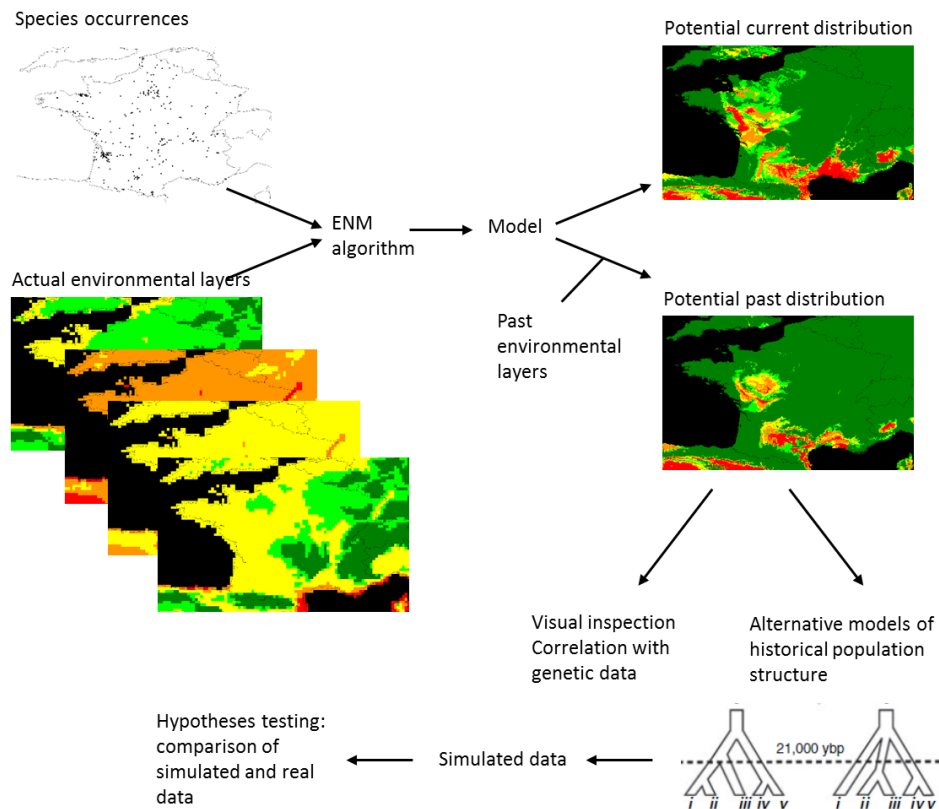


Figure 4.3. Schematic drawing explaining how Ecological niche modeling (ENM) can be used in an integrative framework with phylogeographic data

One of the most common applications of ENMs in phylogeography is to interpret patterns of genetic variation based on the post hoc concordance between patterns of genetic divergence and projections of the species distribution. For example, a deep phylogenetic split between geographically close populations may be suggestive of long-term isolation. Support for this hypothesis might be obtained by visual inspection of projections of past species distributions (i.e. from ENMs based on paleoclimatic variables), where the distributional disjunctions correspond to the

genetic differentiation observed between regions (e.g. Allal *et al.* 2011). Alternatively actually allopatric populations can be genetically similar due to past connectivity, and projections of past species distributions will show spatial continuity between populations in the past (e.g. Nicolas *et al.* 2018). Because these applications are correlative, ad hoc interpretations might be misleading given that other processes might have generated the observed genetic patterns (Alvarado-Serrano and Knowles 2014).

Habitat suitability scores derived from ENMs can be used in a correlative manner to test the impact of the landscape on population genetic connectivity (e.g. Row *et al.* 2010). However it should be emphasized that once again, the correlation between the landscape, or habitat suitability and genetic data may not be causal, because genetic patterns are not exclusively linked to the present landscape configuration but can reflect past configuration of habitat (Alvarado-Serrano and Knowles 2014).

Rather than seeking visual corroboration or correlations between information from ENMs and patterns of genetic variation, ENMs can instead be used to generate hypotheses that are subsequently tested with genetic data. Statistical phylogeographic inferences rely on explicit models of historical scenarios. For any set of multiple populations there is an extremely large set of possible histories with different combinations of connections or routes of colonization among contemporary populations, as well as varying numbers or spatial configurations of populations in the past. The pertinent question is how to decide what model would be an appropriate representation? ENMs, coupled with palaeoclimate estimates, can provide the information necessary for generating alternative models in cases for which no external information on past distributions has previously been available (Carstens and Richards 2007; Richards *et al.* 2007). Estimates of a species' past distributions, or palaeodistribution models, can be used to guide the generation of alternative biogeographical hypotheses. The alternative population structures suggested by palaeodistribution models can be evaluated by constructing null distributions for expected patterns of genetic variation (or a summary statistic that is used to characterize the data) from data simulated by a neutral coalescent process under a specific population model. This approach provides a statistical framework for determining whether the empirical data differed significantly from theoretical expectations under a given model, while taking into account the inherent stochasticity of genetic processes. For example, Portik *et al.* (2017) by combining genome-wide variation and ENM in a model-testing framework showed that although forest refugia played a prominent role in the intraspecific diversification of the amphibian *Scotobleps gabonicus*, there was also evidence for potential

interactions between landscape features and historical refugia, including major rivers and elevational barriers.

Predictions for species distributions at different time periods from ENMs can be used to identify regions of environmental stability where a species may (in principle) have persisted over time (i.e. refugia), in contrast to unstable areas (i.e. areas where climatic changes would have made the region uninhabitable during particular periods). For example, Faye *et al.* (2016) show that, in the palm genus *Podococcus*, high and unique genetic diversity is strongly correlated with inferred areas of climatic stability since the Last Glacial Maxima in Central Africa. Carnaval *et al.* (2009) using frogs as indicator species, ecological niche models under paleoclimates, and simultaneous Bayesian analyses of multispecies molecular data, compared alternative hypotheses of assemblage-scale response to late Quaternary climate change. They show that the southern Atlantic forest of South America was climatically unstable relative to the central region, which served as a large climatic refugium for neotropical species in the late Pleistocene.

An essential aspect of phylogeography is to determine the relationship between the diversification of lineages and variation in ecological niche. By using climatic data to estimate potential geographic distributions, ENMs make it possible to explicitly test if discrete lineages have diverged into different climatic environments or alternatively have retained a particular niche despite population divergence (niche conservatism, Wiens *et al.* 2010; Peterson 2011). For example, ecological divergence would show phylogeographic estimates supporting divergence with gene flow and ENM comparisons would suggest that sister lineages occur in ecologically distinct environments with no evidence of a biogeographic barrier. Alternatively, if biogeographic barriers are promoting divergence, phylogeographic models should show reduced gene flow while ENM demonstrate little or no ecological divergence. Wan *et al.* (2018) showed that in Asian shrew-like moles, much genetic diversification has occurred without evident niche divergence; and that topographical diversity has provided strong geographical isolation.

Recently, the temporal focus of phylogeographic and niche studies has broadened, exploring both historical and current geographic patterns to produce ecological projections of past and future distributions. For instance, Martínez-Méndez *et al.* (2015) integrated analyses of evolutionary history, contemporary geographic distribution, and future distribution projections for the lizard *Sceloporus serrifer*, and concluded that populations inhabiting the lowest lands will become extinct before the year 2070.

To conclude, the rapid methodological advances in the construction of ENMs, together with the increasing availability of geospatial environmental and species distribution data, will no doubt expand the ways in which ENMs might be used to address phylogeographic questions, continuing the impressive trajectory of their applications in phylogeography to date. However, understanding the relationship between the distribution of genetic variation of species and the environmental conditions around them remains challenging, especially given the errors and biases associated with the data, the relative uncertainty regarding the use of algorithms, the discrepancies in method application and interpretation, and the current gaps in the theory regarding the integration of different information sources. Thus, there are significant challenges that limit progress in the application of ENM to phylogeographic studies, as well as various approaches currently under development that seek to solve them (reviewed by Luna-Arangur  and V zquez-Dom nguez 2020).

4.4.2. Integration of life-history traits in phylogeographic studies

Genetic structure of neutral genes primarily reflects demographic processes (e.g., drift, expansion, changes in effective population size) that are a consequence of historical biotic and abiotic conditions during a species' evolutionary history. Phenotypes are targets of selection and affect the performance of organisms in variable environments. Furthermore, different classes of phenotypes vary in how they impact processes such as dispersal, colonization, and persistence, thereby providing a window into the importance of various evolutionary processes in current and historical selective environments. For example, phenotypes related to locomotor efficiency, physiological tolerance or body size influence migration and gene flow among subdivided populations. Recruitment rate, life span and time to maturity affect population size and turnover, and thus the amount of genetic variation in subdivided populations. Given that species-specific phenotypes can dictate spatial variation in population responses to environmental change, phylogeography would benefit from a more integrative and inclusive framework, one that incorporates predictions based on those phenotypes, an approach that has been termed "trait-based phylogeography" (Paz *et al.* 2015). This is important because variation in life-history traits may explain variation in phylogeographic structure among species, which is important to take into account in comparative phylogeography. According to Zamudio *et al.* (2016) a new conceptual framework that explicitly integrates quantitative analyses of phenotypic variation within a phylogeographic framework can greatly enhance our knowledge of how genetic and phenotypic divergence arise,

how they are linked, and how they respond to changing ecological and evolutionary contexts.

Lack of concordance in temporal and spatial clustering in co-distributed taxa may not mean that taxa are not responding to a common landscape or climatic barrier; rather, discrepancies may reflect variation in ecological traits and dispersal capabilities of taxa sampled across the presumed barrier (Papadopoulou and Knowles 2016). These efforts refine expectations for clustered divergences by explicitly including geography and trait-based responses for each species (Massatti and Knowles 2014).

All lineages harbor phenotypic and genetic variation among individuals, and this variation can be geographically partitioned in different ways, as shown in Fig. 4.4 (reviewed by Zamudio *et al.* 2016). Integrative approaches that combine high-throughput sequencing, experimental manipulations and high-quality phenotypic datasets should allow different biological mechanisms underlying phenotype-genotype concordance to be distinguished. Quantifying functional genetic variation within the context of the phylogeographic history of a species and across the range of environments it inhabits can reveal how regional variation in selective regimes and demographic processes drive the evolution of adaptive phenotypes. An analytical framework that advocates genealogical and spatially explicit analyses of intraspecific functional genetic and phenotypic variation will bridge microevolutionary processes acting on individual populations and macroevolutionary patterns at larger spatial and temporal scales. These approaches have been possible only in relatively few systems to date, such as model organisms with extensive genomic resources like stickleback (Deagle *et al.* 2013), in well-known genetic pathways causing shifts in coloration (Hoekstra *et al.* 2004) or in physiological adaptation to high-altitude (Bulgarella *et al.* 2012). However, identifying the underlying genetic basis of phenotypic variation within species is becoming increasingly tractable. Examining functional phenotypic variation in a phylogeographic framework holds great promise for exploring links between genotypic and phenotypic diversity and adaptation across variable environments.

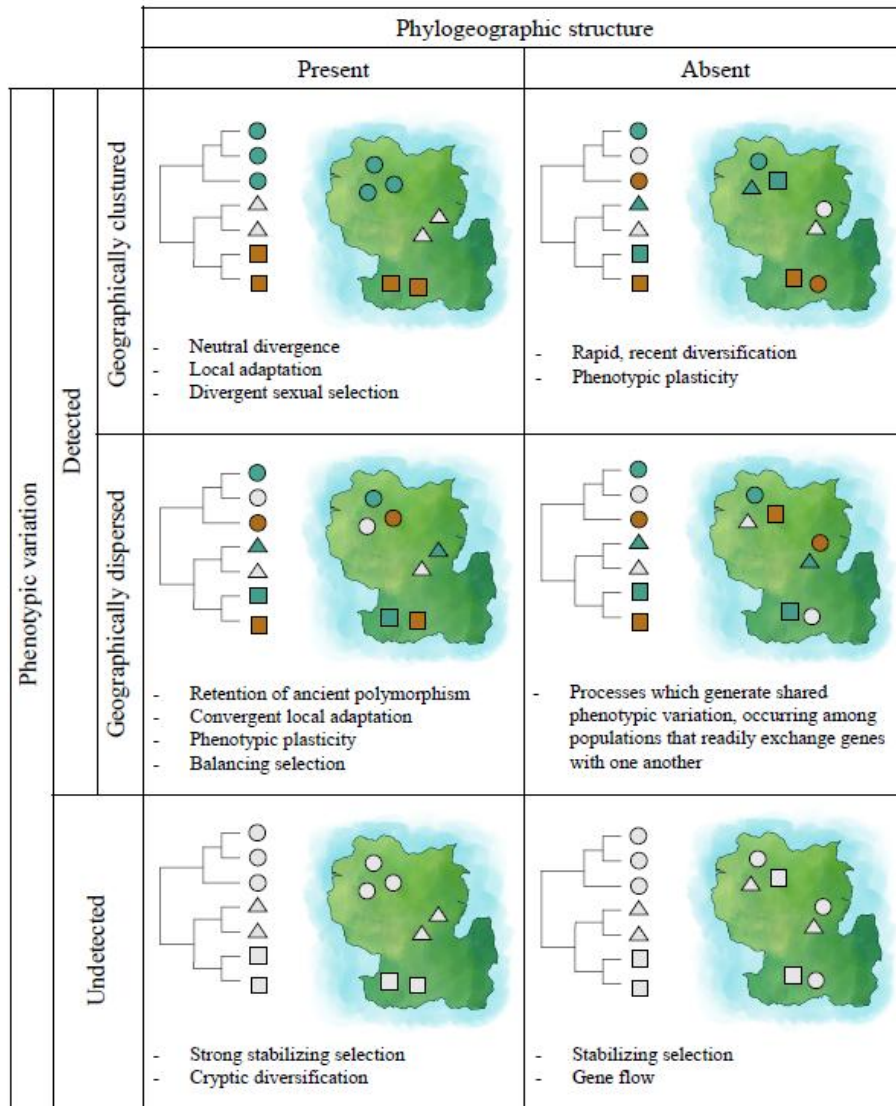


Figure 4.4. Six potential patterns of phylogeographic structure and phenotypic diversity among populations. The three colors represent different phenotypes. Although discrete morphs are depicted here, similar patterns can arise for phenotypes with continuous variation.

4.5. Conclusion

To conclude, phylogeographic studies have yielded hundreds of individual and comparative case studies which, together, provide invaluable insights into the roles of isolation, niche conservatism and environmental stability in generating patterns of alpha- and beta-diversity; and help identify key processes leading to and resulting from extinction events, including the population dynamics of species range reduction and its effects on the strength and temporal flexibility of networks of species interactions. At the same time, studies uniting evolutionary patterns of diversification with ecological processes have shed new light on classic questions such as the evolutionary factors controlling composition of local species assemblages, large-scale patterns of species richness, or the susceptibility of whole clades to extinction under environmental change (Marske *et al.* 2013). Other applications of phylogeographic analyses include defining species boundaries and assigning and assessing conservation priorities (Richards *et al.* 2007).

Owing to the development of statistical phylogeography and dense genomic sampling it is now possible to reconstruct complex demographic histories that involve multiple processes, including population growth and decline, lineage splitting, and post-divergence gene flow. It also allows a stronger focus on the relative importance of neutral vs. selective forces in driving microevolutionary change over time. Phylogeography is clearly at a turning point (Garrick *et al.* 2015): data set size and information content are improving dramatically, due to the large numbers of independent autosomal loci being assayed, and this increase in genomic sampling seems not to have come at the expense of geographic sampling.

4.6. References

- Allal, F., Sanou, H., Millet, L., Vaillant, A., Camus-Kulandaivelu, L., Logossa, Z.A., Lefevre, F., Bouvet, J.M. (2011). Past climate changes explain the phylogeography of *Vitellaria paradoxa* over Africa. *Heredity*, 107(2), 174-186.
- Alvarado-Serrano, D.F., Knowles, L.L. (2014). Ecological niche models in phylogeographic studies: applications, advances and precautions. *Mol Ecol Resour*, 14(2), 233-248.
- Arbogast, B.S., Kenagy, G.J. (2001). Comparative phylogeography as an integrative approach to historical biogeography. *J Biogeogr*, 28(7), 819-825.
- Avice, J.C. (2000). *Phylogeography: the history and formation of species*. Harvard University press, Cambridge.

- Avice, J.C. (2009). Phylogeography: retrospect and prospect. *J Biogeogr*, 36, 3-15.
- Avice, J.C., Bowen, B.W., Ayala, F.J. (2016). In the light of evolution X: Comparative phylogeography. *Proc Natl Acad Sci U S A*, 113(29), 7957-7961.
- Avice, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann Rev Ecol Syst*, 18, 489-522.
- Bermingham, E., Moritz, C. (1998). Comparative phylogeography: concepts and applications. *Mol Ecol*, 7(4), 367-369.
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol*, 29(8), 1917-1932.
- Bulgarella, M., Peters, J.L., Kopuchian, C., Valqui, T., Wilson, R.E., McCracken, K.G. (2012). Multilocus coalescent analysis of haemoglobin differentiation between low- and high-altitude populations of crested ducks (*Lophonetta specularioides*). *Mol Ecol*, 21(2), 350-368.
- Carnaval, A.C., Hickerson, M.J., Haddad, C.F., Rodrigues, M.T., Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323(5915), 785-789.
- Carstens, B.C., Richards, C.L. (2007). Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, 61(6), 1439-1454.
- Chifman, J., Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317-3324.
- Dauby, G., Duminil, J., Heuertz, M., Koffi, G.K., Stevart, T., Hardy, O.J. (2014). Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. *Mol Ecol*, 23(9), 2299-2312.
- De Maio, N., Schrempf, D., Kosiol, C. (2015). PoMo: An Allele Frequency-Based Approach for Species Tree Estimation. *Syst Biol*, 64(6), 1018-1031.
- Deagle, B.E., Jones, F.C., Absher, D.M., Kingsley, D.M., Reimchen, T.E. (2013). Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Mol Ecol*, 22(7), 1917-1932.
- Dufresnes, C., Nicieza, A.G., Litvinchuk, S.N., Rodrigues, N., Jeffries, D.L., Vences, M., Perrin, N., Martinez-Solano, I. (2020). Are glacial refugia hotspots of speciation and cytonuclear discordances? Answers from the genomic phylogeography of Spanish common frogs. *Mol Ecol*, 29(5), 986-1000.
- Faye, A., Deblauwe, V., Mariac, C., Richard, D., Sonke, B., Vigouroux, Y., Couvreur, T.L.P. (2016). Phylogeography of the genus *Podococcus*

(Palmae/Arecaceae) in Central African rain forests: Climate stability predicts unique genetic diversity. *Mol Phylogenet Evol*, 105, 126-138.

Garrick, R.C., Bonatelli, I.A., Hyseni, C., Morales, A., Pelletier, T.A., Perez, M.F., Rice, E., Satler, J.D., Symula, R.E., Thome, M.T., Carstens, B.C. (2015). The evolution of phylogeographic data sets. *Mol Ecol*, 24(6), 1164-1171.

Gavin, D.G., Fitzpatrick, M.C., Gugger, P.F., Heath, K.D., Rodriguez-Sanchez, F., Dobrowski, S.Z., Hampe, A., Hu, F.S., Ashcroft, M.B., Bartlein, P.J., Blois, J.L., Carstens, B.C., Davis, E.B., de Lafontaine, G., Edwards, M.E., Fernandez, M., Henne, P.D., Herring, E.M., Holden, Z.A., Kong, W.S., Liu, J., Magri, D., Matzke, N.J., McGlone, M.S., Salter, F., Stigall, A.L., Tsai, Y.H., Williams, J.W. (2014). Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytol*, 204(1), 37-54.

Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. *Mol Ecol Resour*, 11(5), 759-769.

Gutiérrez-García, T.A., Vázquez-Domínguez, E. (2011). Comparative Phylogeography: designing studies while surviving the process. *BioScience*.

Hewitt, G.M. (2004a). Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc B*, 359(1442), 183-195.

Hewitt, G.M. (2004b). The structure of biodiversity - insights from molecular phylogeography. *Front Zool*, 1(1), 4.

Hickerson, M.J., Stahl, E.A., Lessios, H.A. (2006). Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, 60(12), 2435-2453.

Hoekstra, H.E., Drumm, K.E., Nachman, M.W. (2004). Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, 58(6), 1329-1341.

Huang, W., Takebayashi, N., Qi, Y., Hickerson, M.J. (2011). MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics*, 12, 1.

Irwin, D.E. (2002). Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, 56(12), 2383-2394.

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

Knowles, L.L. (2009). Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, 40, 593-612.

Knowles, L.L., Maddison, W.P. (2002). Statistical phylogeography. *Mol Ecol*, 11, 2623-2635.

Lalis, A., Leblois, R., Liefried, S., Ouarour, A., Beeravolu, C., Michaux, J., Hamani, A., Denys, C., V., N. (2016). New molecular data favour an anthropogenic

introduction of the wood mouse (*Apodemus sylvaticus*) in North Africa. *J Zool Syst Evol Res*, 54(1), 1-15.

Lalis, A., Mona, S., Stoetzel, E., Bonhomme, F., Souttou, K., Ouarour, A., Aulagnier, S., Denys, C., Nicolas, V. (2019). Out of Africa: demographic and colonization history of the Algerian mouse (*Mus spretus* Lataste). *Heredity*, 122(2), 150-171.

Leaché, A.D., Oaks, J.R. (2017). The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. *Ann Rev Ecol Evol Systematics*, 48(1), 69-84.

Leaché, A.D., Banbury, B.L., Felsenstein, J., de Oca, A.N., Stamatakis, A. (2015). Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Syst Biol*, 64(6), 1032-1047.

Luna-Aranguré, C., Vázquez-Domínguez, E. (2020). Analysis of the application of ecological niche modeling in phylogeographic studies: contributions, challenges, and future. *Therya*, 11(1), 47-55.

Marske, K.A., Rahbek, C., Nogués-Bravo, D. (2013). Phylogeography: spanning the ecology-evolution continuum. *Ecography*, 36, 1169-1181.

Martínez-Méndez, N., Mejía, O., Méndez de la Cruz, F.R. (2015). The past, present and future of a lizard: The phylogeography and extinction risk of *Sceloporus serrifer* (Squamata: Phrynosomatidae) under a global warming scenario. *Zool Anzeiger*, 254, 86-98.

Massatti, R., Knowles, L.L. (2014). Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, 68(10), 2833-2846.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol*, 66(2), 526-538.

Médail, F., Baumel, A. (2018). Using phylogeography to define conservation priorities: the case of narrow endemic plants in the Mediterranean Basin hotspot. *Biol Conserv*, 224, 258-266.

Mirarab, S., Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31(12), i44-52.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, 30(17), i541-548.

Myers, E.A., McKelvy, A.D., Burbrink, F.T. (2020). Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). *Mol Ecol*, 29(4), 797-811.

- Nicolas, V., Mataame, A., Crochet, P.A., Geniez, P., Fahd, S., Ohler, A. (2018). Phylogeography and ecological niche modelling unravel the evolutionary history of the African green toad, *Bufotes boulengeri boulengeri* (Amphibia: Bufonidae), through the Quaternary. *J Zool Syst Evol Res*, 56(1), 102-116.
- Ottenburghs, J., Lavretsky, P., Peters, J.L., Kawakami, T., Kraus, R.H.S. (2019). Population Genomics and Phylogeography. In *Avian Genomics in Ecology and Evolution*, ed. Kraus R. Springer, Cham.
- Papadopoulou, A., Knowles, L.L. (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proc Natl Acad Sci U S A*, 113(29), 8018-8024.
- Paz, A., Ibanez, R., Lips, K.R., Crawford, A.J. (2015). Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. *Mol Ecol*, 24(14), 3723-3737.
- Peterson, A.T. (2011). Ecological niche conservatism: a time-structured review of evidence. *J Biogeog*, 38, 817-827.
- Portik, D.M., Leache, A.D., Rivera, D., Barej, M.F., Burger, M., Hirschfeld, M., Rodel, M.O., Blackburn, D.C., Fujita, M.K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol Ecol*, 26(19), 5245-5263.
- Pritchard, J.K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.
- Richards, C.L., Carstens, B.C., Knowles, L.L. (2007). Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *J Biogeogr*, 34, 1833-1845.
- Row, J.R., Blouin-Demers, G., Loughheed, S.C. (2010). Habitat distribution influences dispersal and fine-scale genetic population structure of eastern foxsnakes (*Mintonius gloydi*) across a fragmented landscape. *Mol Ecol*, 19(23), 5157-5171.
- Shendure, J., Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol*, 26(10), 1135-1145.
- Stephens, M., Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*, 73(5), 1162-1169.
- Stephens, M., Smith, N.J., Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, 68(4), 978-989.
- Stone, G.N., Lohse, K., Nicholls, J.A., Fuentes-Utrilla, P., Sinclair, F., Schonrogge, K., Csoka, G., Melika, G., Nieves-Aldrey, J.L., Pujade-Villar, J., Tavakoli, M., Askew, R.R., Hickerson, M.J. (2012). Reconstructing community assembly in time and space reveals enemy escape in a Western Palearctic insect community. *Curr Biol*, 22(6), 532-537.

Wan, T., He, K., Jin, W., Liu, S.-Y., Chen, Z.-Z., Zhang, B., Murphy, R.W., Jiang, X.-L. (2018). Climate niche conservatism and complex topography illuminate the cryptic diversification of Asian shrew-like moles. *J Biogeogr*, 45, 2400–2414.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V., Damschen, E.I., Jonathan Davies, T., Grytnes, J.A., Harrison, S.P., Hawkins, B.A., Holt, R.D., McCain, C.M., Stephens, P.R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol Lett*, 13(10), 1310-1324.

Zamudio, K.R., Bell, R.C., Mason, N.A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proc Natl Acad Sci U S A*, 113(29), 8041-8048.