

Part V

Conclusions

Chapter XIV

Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity

Steven Weiss* and Nuno Ferrand**

E-mail: *steven.weiss@uni-graz.at; **nferrand@mail.icav.up.pt

Abstract.

This contribution briefly summarizes the current state of phylogeography and its significance in evolutionary research, particularly involving organisms existing in southern European refugia. These refugia, characterized by a highly heterogeneous landscape, harbor a large percentage of Europe's organismal diversity. They are also shown to offer model systems where the long-term dynamics of particular evolutionary phenomena can be fruitfully explored within well-defined phylogeographic contexts. It is suggested, for example, that the dynamics of gene-flow and natural selection in hybrid zones within long-term glacial refugia are much more complex than those in previously glaciated regions. Emphasis in this review is also given to the breadth of available analytical approaches to phylogeographic analysis, including current controversies and ongoing developments such as the aim to better integrate coalescent theory and both large(geographic) and small scale (landscape) environmental variables. Despite different schools of thought on how to approach phylogeographic analysis, a plea is made to maintain pluralism when dealing with such complex, multi-disciplinary and stochastically influenced data sets. In summary, phylogeography is shown to be a highly successful and popular field of inquiry with a high potential for growth, especially as cutting edge analytical and genomic-oriented techniques, often developed within the field of human genetics, are applied across broader taxonomic scales.

Keywords: conservation, coalescence, nested clade phylogeographic analysis, hybrid zones, statistical phylogeography, landscape genetics

Introduction

The phylogeographic approach has proven its utility in evolutionary research across a broad array of taxa and areas of inquiry. Its greatest contributions thus far have been its emphasis on non-equilibrium aspects of population genetics, the relationship between demography and historical genealogy, and the interfacing of formerly distinct fields of population genetics and phylo-

genetics (Avice 2000). The phylogeographic structure of an organism provides the biologist with a broad-scale evolutionary framework, which in itself can answer some straightforward questions, but moreover provides a historical foundation upon which more precise hypotheses can be efficiently formulated and tested. The universality of this perspective is underscored by the observation that nearly all species in nature exhibit some level of genetic structuring associated with geography. This structure can be highly complex, especially for species whose ranges encompass regions subject to high-amplitude paleoclimatic fluctuations. When the 'genealogical' dimension is added to geographic patterns, phylogeographic structure arises. Phylogeographic structure reflects the interaction between both demographic and genealogical processes and landscape level dynamics. The enlightening appeal of phylogeography stems in part from the puzzle-like challenge of uncovering the historical role of geographical processes in shaping current genetic structure. Sea- or lake-level fluctuations, land-bridge formations, tectonic shifts, glacial advance and retreat and river course dynamics all represent biologically independent phenomena that can affect an organism's demography. The coupling of historical events with geographic breaks in an organism's genealogy underscores the success of the phylogeographic perspective.

European phylogeography developed rather quickly with continent-wide studies, which were soon related to a variable model of how southern peninsular refugia contributed to the postglacial colonization of northern Europe (Hewitt 1996, 1999, 2000; Chapter 3). Implicit in this model is that southern refugia harbor higher levels of biodiversity, which serve both as a source for future demographic expansions as well as evolutionary radiations. While the southern peninsulas obviously experienced more climatic stability than regions frequented by glaciers or tundra-like conditions, they are also highly fragmented by mountain ranges and bordered by seas, creating highly heterogeneous landscapes that foster organismal diversification. In recent years, many researchers have built a strong argument for the role of additional refugia in central or unglaciated portions of both northern and eastern Europe for some groups of organisms (Stewart & Lister 2001). This has been particularly apparent for cold-tolerant fishes, such as *Cottus gobio*, brown trout *Salmo trutta* and European grayling *Thymallus thymallus*, for which more northern refugia may have been solely responsible for the recolonization of formerly glaciated regions in northern Europe after the Last Glacial Maximum (Volckaert *et al.* 2002; García-Marín *et al.* 1999; Weiss *et al.* 2000, 2002). This perspective has also been addressed for boreal-temperate plants such as in *Betula* (Palmé *et al.* 2003) and *Calluna* (Rendell & Ennos 2002). This growing appreciation for northern refugia calls into question the role of peninsular refugia in the recolonization of previously glaciated regions, implying that for some organisms, these southern populations may be better characterized as long-term relicts, rather than sources for future expansions or radiations. Nonetheless, the three peninsular refugia of southern Europe are centers of

biodiversity and due to both natural fragmentation as well as anthropogenic influences, contain numerous threatened or endangered taxa and populations. In fact, the Mediterranean region as a whole is considered one of the 25 biodiversity hotspots on earth (Myers *et al.* 2000). While the species richness of southern Europe is well documented, research assimilating phylogeographic approaches demonstrate an underappreciated genealogical complexity of organisms within large refugial habitats. This is exemplified herein by the 'refugia within refugia' model (Chapter 5) demonstrating multiple highly divergent lineages for a number of species within one peninsula, or, for example in the wall lizard *Podarcis* sp., numerous undescribed taxa or cryptic species diversity (Harris & Sá-Sousa 2002; Pinho *et al.* in press).

Simple interpretations of high genetic diversity within southern refugia, however, have been challenged. In particular, the results have been shown to depend on how genetic diversity is quantified (Comps *et al.* 2001; Petit *et al.* 2003). This stems from the fact that different measures of genetic diversity, such as allelic richness, heterozygosity and population divergence are differentially influenced by demographic factors. Heterozygosity, for example, can be very high in hybrid zones formed by expanding lineages north of refuge areas, as expected when two formerly divergent lineages come into secondary contact. However, secondary contact zones within southern refugia have received much less attention, and may be characterized by extremely high pairwise population divergences and much more complex evolutionary dynamics. For example, the hybrid zone of the European rabbit *Oryctolagus cuniculus* in Iberia contains two lineages thought to have arisen in allopatry approximately two million years ago. This hybrid zone has been used as a model system to demonstrate highly differential modes of introgression across markers (Chapters 7 & 8), saturation of mutation spectra in microsatellites (Queney *et al.* 2001), and to suggest balancing selection in several protein loci (Ferrand *et al.*, unpublished data). Within the same geographic region, a hybrid zone of the endemic Schreiber's green lizard *Lacerta schreiberi* has revealed not only differential introgression across markers, but also recombinant haplotypes across relatively short genomic regions (Godinho *et al.* in press). Such marker discordance in areas of secondary contact clearly reflects more complex and long-term interactions of fluctuating demography and selection than expected in postglacial hybrid zones. While the studies referenced above all derive from the Iberian Peninsula, it is expected that hybrid zones within refugia exist on other peninsulas, though they await future study.

Caveats, criticisms and challenges of phylogeographic approaches

Phylogeographic research continues to be dominated by studies utilizing single locus uni-parentally inherited markers, namely mtDNA or cpDNA

genes. There are numerous caveats to exclusively relying on organelle markers (reviewed in Ballard & Whitlock 2004), but their utility in many systems is unquestionable, and it is perhaps remarkable to note the number of initial inferences drawn from single marker phylogeographic studies that withstand extended analysis including more genetic markers as well as sampled individuals. In contrast, where sex-biased dispersal is strong, or interspecific hybridization and capture of an organelle genome has occurred, the single organelle gene may be completely misleading (Bernatchez *et al.* 1995; Melo-Ferreira *et al.* 2005). Regardless of the frequency of such events, the problem is that there is no way to predict the reliability of such single locus data sets. If theory is followed, many gene genealogies must be sampled in order to have some degree of confidence that the history of an organismal lineage is being reasonably recovered. A counter-intuitive illustration of this point, especially for those trained in classical population genetics or ecological studies, is that the number of individuals needed to be sampled in a population in order to uncover the genealogical diversity is quite small, with a sample size of 10 sufficing for an approximate 90% probability of revealing the most ancestral haplotype (Hein *et al.* 2005). In contrast, the variance in coalescence trees reveals that the most efficient way of increasing the accuracy of inferences drawn from gene genealogies is to increase the number of independent (i.e. unlinked) loci sampled. An additional aspect, however, is the number of distinct demes existing across a species range. The best phylogeographic studies sample the entire range of the organism, and using some prior knowledge of an organism's genetic or phenotypic diversity, or predictions of how landscape fragmentation may have molded extant genetic structure, an attempt is made to provide sample coverage of potential major demes. More suspect or less convincing are phylogeographic studies that cover only small portions of a species range, or highly fragmented or opportunistic sampling of predicted or known demes, and then attempt to draw large-scale inferences. Such efforts suffer both from poor sampling and the uncertainties of a single-locus data set, and thus should be highly discouraged.

Another aspect of phylogeography that needs improvement is the approach often taken to estimate divergence times among populations or sister taxa, although this problem is certainly not specific to phylogeography. While the theoretical understanding that the coalescence of two lineages as measured with DNA sequence data will always predate the splitting of the actual populations has been with us for some time, many researchers ignore this difference and overestimate divergence, by as much as 50% or even more in highly structured populations (Edwards & Beerli 2000; Arbogast *et al.* 2002). For study-specific calibrations based on geological evidence this caveat may not be too important (but see Arbogast *et al.* 2002), but numerous studies still rely on universal clock rates or apply a calibration in one study across many taxa or wholly different time scales. Again, this problem is exacerbated when only a single locus is screened.

A central area of much debate is the application of statistical tools for making inferences in an explicit phylogeographic context. Phylogeographic analysis began with a fusion of monumental technical advances in the laboratory combined with hand-drawn intuitively satisfying genealogical networks and purely *ad hoc* inferences (Avice *et al.* 1979). Subsequently, various theoretical developments and statistical tools were applied to the analysis of genes in a phylogeographic context, such as pairwise mismatch analysis (Slatkin & Hudson 1991; Rogers & Harpending 1992); analysis of molecular variance, AMOVA (Excoffier *et al.* 1992), and several approaches to reconstructing haplotype networks, recently reviewed in Posada & Crandall (2001). However, the most comprehensive attempt to develop a systematic approach to phylogeographic analysis has been described in a series of publications, now best summarized as Nested Clade Phylogeographic Analysis (NCPA) (Templeton 1998, 2004). Some studies have pointed out deficiencies in the inferences drawn from NCPA involving specific historical scenarios or demographic events that lead to false inferences (Alexandrino *et al.* 2002; Paulo *et al.* 2002; Masta *et al.* 2003). These criticisms have led to either changes in the methodology, or explicit caveats of when the approach is more likely to fail (Templeton 2004). For example, range expansion is often not inferred even in systems where common sense or other information dictates that range expansion has occurred. However, no analysis can support range expansion when a genetic marker lacks the resolution to do so, as is often the case for mtDNA and recent (i.e. postglacial) range expansion. A much more sweeping criticism is that NCPA is *ad hoc*, non-statistical and not amenable to falsification (Knowles & Maddison 2002; Knowles 2004). This emerging debate has resulted in the coining of the term '*Statistical Phylogeography*' for alternative model-based approaches.

NCPA, however, does involve several explicitly statistical steps, the first of which is the reconstruction of a 95% parsimony network, and the second the construction of a nested design for testing genetic/geographic associations based on this network. This design is three-dimensional, incorporating nested spatial distances, genetic variation, and a qualitative temporal component based on the genealogy (the level of nesting). The most common alternative to such a design is a simple Mantel test on the global data set, an approach that produces only a single statistic, but moreover lacks a temporal component. The clear weakness of the nested statistical design in NCPA is that it rests wholly on the reliability of the haplotype network. Petit & Grivet (2002) further question the appropriateness of the permutational procedure that determines statistical significance, arguing that populations rather than individuals should be permuted (but see Templeton 2002a). There have also been other attempts to capture the temporal component of allelic data without using trees or networks (Pons & Petit 1996; Grivet & Petit 2002).

The most unorthodox component of NCPA is the use of a key to draw inferences from the correlation statistics. Such an approach can be considered a

posteriori as opposed to *a priori* hypothesis testing, the common mode of natural scientists. However, the use of an inference key is not *ad hoc* and combined with the frequent conclusion that a data set lacks genetic resolution or adequate geographic sampling to draw an inference, NCPA inhibits rather than promotes *ad hoc* inferences. Templeton (2004) discusses in detail fundamental considerations, limitations and complementarity of '*a priori* and *a posteriori*' approaches to phylogeographic analysis. More broadly, any gene-tree based methodology, including NCPA, can be viewed as a graphical approach as opposed to a more traditional mathematical approach utilizing summary statistics, whereby both approaches clearly have their limitations (Hey & Machado 2003). Smouse (1998) concludes that both trees (bi-furcating) and networks (multi-furcating) have serious drawbacks, and researchers should rather spend more time analyzing the haplotype data with more straightforward statistical approaches.

'Statistical Phylogeography', as presented by Knowles (2004), proposes to steer the discipline into a model-driven, hypothesis-testing framework, which nonetheless considers some of the additional, at times non-biological information that may be important for recovering or drawing additional meaning to a phylogeographic scenario. A major concern of Knowles (2004) is an underappreciation of the stochastic variance inherent in gene genealogies (not reflected in a single haplotype network), which will lead to inaccurate or misleading interpretations and thus must be incorporated into the analytical approach. The challenge is to define *a priori* model-based predictions, which can be falsified in a statistical framework that incorporates the stochasticity of gene coalescence. Whereas recognition of coalescent theory has been implicitly integrated into phylogeographic interpretation since its onset (Avise *et al.* 1979) and used to justify inference frameworks in NCPA (e.g. Castelleo & Templeton 1994), it is not often explicitly or computationally integrated into hypothesis testing and the generation of statistical confidence. Thus, the advocacy for so-called 'Statistical Phylogeography' reflects several of the discipline's most common caveats and further parallels the prognosis that better integration of coalescent theory into phylogeographic analysis will not only improve our confidence in phylogeographic inferences, but also aid in the advancement of population genetics as a whole (Wakeley 2003).

Current trends and state-of-the-art (2001-present)

Despite the theoretical advantages of multi-locus approaches, the majority of current manuscripts in phylogeography is still based on single locus uniparentally inherited markers. Viewing the last 12 issues of *Molecular Ecology* (Oct. 2005-Nov. 2004), there were 42 manuscripts with the term 'phylogeography' in the title, among which 34 (81%) were based exclusively on

organelle genes (mtDNA or cpDNA). Nonetheless, these single-locus studies improve upon past efforts in that more nucleotides are screened and sample coverage often includes continent-wide or circumpolar coverage (Alsos *et al.* 2005; Van Houdt *et al.* 2005). Additionally, due to concerns about heterogeneous substitution rates, many studies routinely use more than one organelle gene (11 of the 34 noted above). Coalescent simulations and the so-called 'gene-tree/population tree' approach are also applied to answer more discrete phylogeographic scenarios (e.g. one, two or several refugia) still using mtDNA data sets (Knowles 2001; Carstens *et al.* 2005) with particular subroutines within the modular software package Mesquite (Maddison & Maddison 2004).

The more important advance to true multi-locus phylogeographic studies, however, has been rather slow despite the optimistic prospects outlined in Hare (2001). Relevant theoretical issues include an underappreciation of the stochasticity of the coalescent process (Hudson & Turelli 2003), the statistical treatment of recombinants (Templeton 2004), and both the computational complexity and lack of empirical data on the among locus variance in substitution rates (Yang 1997; Arbogast *et al.* 2002). Practical hurdles involve locating nDNA regions with adequate variation, and resolving haplotypes from diploid genotypes. This latter problem can be approached with a variety of either statistical or experimental methods (reviewed in Zhang & Hewitt 2003). For a typical phylogeographic study reasonable throughput can be achieved through the combination of PCR-SSCP and subsample sequencing or initial PCR and allele-specific PCR (ARMS, Newton *et al.* 1989) to resolve the potential ambiguities of heterozygotes, which can be extreme when intron indels are present. For large data sets, probable haplotypes at a level of confidence adjustable to the study's demands can be estimated using maximum likelihood or Bayesian approaches (Excoffier and Slatkin 1995; Stephens *et al.* 2001)

Templeton (2002b) provides an example of a multi-locus phylogeographic study using 10 genes to argue for a more complex out of Africa colonization by humans, based on both NCPA and some coalescent simulations, though his conclusions are somewhat controversial. We must emphasize that the use of a single nuclear gene fragment as a token consideration of the nuclear genome for a broad-scale phylogeographic study may achieve very little. Furthermore, the application of multi-locus microsatellite surveys to phylogeographic studies may serve as a poor surrogate to nDNA haplotype data. The combined use of microsatellites and mtDNA in an animal study can control for extreme sex-biased dispersal or interspecific introgression, events that can make inferences based solely on mtDNA data extremely misleading. However, it is difficult to analyze microsatellite allele variation in a strictly phylogeographic context. This is because the available analytical techniques do not explicitly integrate a temporal perspective, and more importantly the high and variable substitution rate of repeat loci results in an unpredictable and

non-linear relation between genetic divergence and time. This problem is magnified across deeper phylogeographic breaks, where homoplasy will lead to vastly underestimated divergence estimates among some, but not all pairwise population comparisons. One of the best examples of this is the demonstration of stationary distributions of microsatellite allele frequencies in a phylogeographic context with the European rabbit (Queney *et al.* 2001), an observation that has been further substantiated when it was shown that sex-biased dispersal is not prevalent in the system (Geraldes & Ferrand in press). Thus, while theorists bemoan the lack of true multi-locus phylogeographic studies, it is apparent that both theoretical and logistic constraints continue to limit such efforts.

One relatively recent approach, however, combines nDNA single-nucleotide polymorphisms (SNPs) with a linked microsatellite. Thus, one can take advantage of the high mutation rate of a microsatellite incorporated into haplotype data to age alleles in a phylogeographic context (Tishkoff *et al.* 2001). Such data can also be used with new software developed for more flexible coalescent simulations allowing incorporation of differing mutation rates and effective population sizes in estimating the divergence of populations (Hey & Nielsen 2004). This approach was recently used to evaluate the number of founders in the colonization of America by humans (Hey 2005) as well as to investigate the divergence times among Lake Malawi cichlids (Won *et al.* 2005). There are numerous opportunities to use such approaches to answer explicit questions within a phylogeographic framework.

An additional avenue of inquiry that is directly applicable to phylogeographic analysis involves attempts to incorporate geography into coalescent simulations. The development of the software DANCING TREES will give birth to an inference tool that explicitly incorporates the spatial component in the coalescent (S. Baird, personal communication). The incorporation of climatic and other landscape variables into simulated phylogeographic scenarios across, for example, postglacial time scales, has also been accomplished with the development of the software SPLATCHE (Curat *et al.* 2004), which has been applied to questions concerning the geographic origin of early modern humans (Ray *et al.* 2005). At finer scales, the link between environmental factors and population genetics is developing into a field of its own – landscape genetics – whereby the newest software allowing for such analysis is GENELAND (Guillot *et al.* 2005a,b).

Yet another theme that has been until most recently ignored in phylogeographic research is the effects of natural selection on the markers we choose to use, as well as its large-scale influence on genetic architecture and the adaptive landscape. One view is that estimates of ancient demography are best done with selectively neutral markers, and thus effort is expended to detect markers under selection in order that they be removed from particular analyses (Vitalis *et al.* 2001). More problematic is mtDNA, so often used as a putatively neutral marker in phylogeographic studies, but difficult to replace with

a nuclear gene region with the same level of genetic structure. Recently, natural selection has been hypothesized to have played a major role in shaping the present-day diversity of mtDNA in humans (Elson *et al.* 2004). Thus, the concern that selection on the mtDNA molecule may have affected the phylogeographic structure in other organisms is increasing (Ballard & Whitlock 2004). For nuclear gene loci, methods to detect markers under selection have been developing (see reviews of Beaumont 2005 & Storz 2005) and under particular population models the detection of F_{st} outlier loci (Beaumont & Nichols 1996) is currently gaining use. In contrast to viewing natural selection as problematic, phylogeographers can also use markers under selection in order to gain a better understanding of the evolutionary mechanisms that have accompanied or perhaps even directly influenced demographic declines and expansions at a large scale.

Consideration of a broader array of evolutionary mechanisms, and incorporation of selection, recombination and linkage disequilibrium into phylogeographic studies is easily feasible only for model organisms. In humans, the use of compound haplotypes such as SNPSTRs (Mountain *et al.* 2002) or HapSTRs (Hey *et al.* 2004) has been used to describe the signature of natural selection across broad geographic scales in the glucose-6-phosphate dehydrogenase and lactase loci (Tishkoff *et al.* 2001; Coelho *et al.* 2005). In *Drosophila melanogaster* screening of whole genomic regions across large geographic scales not only suggests that selection has played a major role in global demographic expansion, but also underscores the effect that both selection and genomically heterogeneous recombination rates have on the behavior of nuclear gene markers across phylogeographically relevant scales (Kauer *et al.* 2002, 2003; Catania *et al.* 2004). As the genetic knowledge base of non-model organisms continues to increase, approaches that incorporate genome-wide information, and/or markers under selection will be more broadly applied to questions carried out within a phylogeographic context. It is at this point that a synergy between population genetics and phylogeography can be recognized that can pave the way for many future evolutionary research programs.

In fact, population geneticists who appreciate phylogeography recognize its utility in providing a framework within which more focused, evolutionary genetic questions can be pursued. Thus, the study system of the European rabbit in the Iberian Peninsula presents two distinct phylogroups that meet in a contact zone within a long-term refugial environment, providing a setting to study differential rates of introgression, selection, drift and migration across multiple genetic markers over significant periods of time (Chapter 8; Queney *et al.* 2001; Geraldes & Ferrand *in press*). Research on the evolutionary history of the golden-striped salamander *Chioglossa lusitanica* began with the study of allozyme and mtDNA variation across the entire species range but has more recently focused on the contact zone of two major phylogroups, revealing seemingly discordant patterns among loci, fine-scaled gene flow

and vicariance (Chapter 6; Sequeira *et al.* 2005). These data have been coupled to climatic modeling of suitable habitat across relevant phylogeographic times scales, supporting the hypothesized vicariance scenario as well as allowing future predictions (Teixeira *et al.* 2001; Teixeira & Arntzen 2002). While using only a single nDNA locus, Godinho *et al.* (in press) demonstrated that a nuclear genealogy involving recombinant haplotypes can aid in the investigation of ancient demography and admixture dynamics within the framework of an mtDNA-defined phylogeographic structure. On a larger scale, the combination of comparative phylogeography and paleoclimatic modeling was used to verify the location of refugia and predict species responses to future environmental changes (Hugall *et al.* 2002). Drawing on one of the largest sample collections of a wild organism, Petit *et al.* (2002a,b) integrate fossil tree pollen data to verify the location of glacial refugia and increase both the resolution and confidence of prior phylogeographic inferences that were made based on organelle genetics alone. This perspective of considering how biota respond to climatic change, embedded in a broad-scale phylogeographic framework has recently led to the consideration of the conservation relevance of the rear (i.e. southern) edge of species distributions, in contrast to the frequent focus on the postglacial expanding edge (Hampe & Petit 2005), further illuminating the importance of understanding the role of southern peninsulas for the conservation of European biodiversity (Taberlet & Cheddadi 2002). These broad-scale environmental perspectives invariably fuel our attention to the interaction of genetics, landscape dynamics and habitat requirements at the organismic level, forming the basis of landscape genetic approaches that integrate GIS based modeling and the estimation of traditional population genetic parameters (Manel *et al.* 2003; Spear *et al.* 2005). Thus, extending on the simplistic dichotomous characterization of postglacial expansion stemming from dispersal from refugia into formerly uninhabited environments, landscape genetics seeks to understand both dispersal and population differentiation in terms of the landscape variables that promote or inhibit individual movement and colonization (Geffen *et al.* 2004; Arnaud 2003).

In summary, current trends directly or indirectly related to phylogeography seek to expand upon the integration of an array of data external to genetic architecture, in order to understand the true nature and importance of spatial genetic structure to both the evolutionary legacy and future of an organism. At the genomic level, developments are almost all directed towards improving upon the legitimacy of the genetic signal in reflecting the true history of populations, by increasing genomic coverage and incorporating coalescent theory and statistical inference.

Prospects and predictions for European phylogeography

Phylogeography is not a discipline without detractors. Recently, the entire paradigm of the importance of phylogeographic structure in Europe was challenged. Based on a study of ancient DNA, which revealed a lack of phylogeographic structure in mammals of the last interglacial, Hofreiter *et al.* (2004) suggest that the current structure of extant animals is in a transitory phase. Despite the continued reliance on initial patterns seen at organelle genes, for some organisms, like forest trees, these genes may reveal a very limited view on the geographic structure of the genome due to limited dispersal, especially for seeds, as well as insufficient mutation rates. Considering the consistent discordance between organelle and nDNA markers in animal studies in contact zones (particularly older hybrid zones) this contention may be relevant for other taxa as well. Nonetheless, the caveats are known, and the phylogeographer's tool chest has grown significantly so there is little doubt that our understanding of the pattern and importance of genealogical distributions in time and space within species will also expand. And, if current literature is any indication, the phylogeographic perspective will remain an important baseline component of evolutionary research.

While obvious to some, young researchers engaging in phylogeographic research should be aware that a large percentage of the approaches applied in the field were first developed, and continue to be developed in the highly competitive field of human genetics. If this field is any indication of what will become of our understanding of phylogeographic structure and inferences on paleodemography for non-human organisms, then much controversy lies ahead. Increasingly sophisticated models of the out-of-Africa expansion of modern humans continue to fuel debate on the level of ancient population size, substructure, selection and migration and the relevance of these factors to the observed genomic architecture (Harding & McVean 2004). Nonetheless, human studies are fueling the development of coalescent based approaches, integration of more complex demographic scenarios such as metapopulation dynamics, and examination of gene expression and selection regimes in time and space for phenotypic traits of interest, approaches that are paving the way for the evolutionary biologist interested in applying cutting edge technology to the evolutionary study of non-human species.

As noted by John Avise at the first European conference on Phylogeography in Vairão, Portugal (2002), European phylogeographers have done a remarkable job in putting together large-scale data sets encompassing whole species ranges and most notably, have often resolved questions that would have been difficult or impossible to answer without a phylogeographic perspective. The description of the phylogeographic structure of an organismal lineage is often just the starting point before delving into more specific questions or seeking statistical support for a particular inference or historical sce-

nario. Nevertheless, obtaining an evolutionary framework of a species, in space and time, is an irreplaceable milestone in understanding that can affect both the direction and interpretation of genetic data at any scale of inquiry.

Acknowledgements

Nuno Ferrand was supported by a sabbatical grant (SFRH/BSAB/318/2002) from Fundação para a Ciência e a Tecnologia (FCT). Much of the results described in this chapter were supported, in part, by FCT research projects, including POCTI/BIA/11174/98, POCTI/BSE/40987/2001 and POCTI/BSE/34151/2000. Nuno Ferrand and Steven Weiss want to acknowledge all the members of CIBIO for their enthusiasm and dedication, as well as for providing an excellent and inspiring atmosphere. We want to thank Rémy Petit for his comments on an earlier draft of this paper.

References

- Alexandrino J, Arntzen JW, Ferrand N (2002) Nested clade analysis and the genetic evidence for population expansion in the phylogeography of the golden-striped salamander, *Chioglossa lusitanica* (Amphibia : Urodela). *Heredity* **88**, 66-74.
- Alsos IG, Engelskjøn T, Gielly L, Taberlet P, Brochman C (2005) Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. *Molecular Ecology* **14**, 2739-2754.
- Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review in Ecology and Systematics* **33**, 707-740.
- Arnaud JF (2003) Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. *Landscape Ecology* **18**, 333-346.
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge.
- Avise JC, Lansman RA, Shade RO (1979) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution of the genus *Peromyscus*. *Genetics* **92**, 279-295.
- Ballard JWO, Whitlock M (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.
- Beaumont M (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology and Evolution* **20**, 435-440.
- Beaumont M, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society London, B* **263**, 1619-1626.
- Bernatchez L, Glémet H, Wilson C, Danzmann RG (1995) Fixation of introgressed mitochondrial genome of arctic charr (*Salvelinus alpinus* L.) in an allopatric population of brook charr (*Salvelinus fontinalis* Mitchell). *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 179-185.

- Carstens BC, Stevenson AL, Degenhardt JD, Sullivan J (2005) Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology* 53, 781-792.
- Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Evolution and Phylogenetics* 3, 102-113.
- Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlötterer C (2004) World-wide survey of an Accord insertion and its association with DDT resistance in *Drosophila melanogaster*. *Molecular Ecology* 13, 2491-2504.
- Coelho M, Luiselli D, Bertorelle G, Lopes AI, Seixas S, Destro-Bisol G, Rocha J (2005). Microsatellite variation and evolution of human lactase persistence. *Human Genetics* 117, 329-339.
- Comps B, Gömöry D, Letouzey J, Thiebaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics* 157, 389-397.
- Curat M, Ray N, Excoffier L (2004) SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes* 4, 139-142.
- Edwards SV, Beerli P (2000) Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54, 1839-1854.
- Elson JL, Turnbull DM, Howell N (2004) Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *American Journal of Human Genetics* 74, 229-238.
- Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Molecular Biology and Evolution* 12, 921-927.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479-491.
- Garcia-Marin JL, Utter FM, Pla C (1999) Postglacial colonization of brown trout in Europe based on distribution of allozyme variants. *Heredity* 82, 46-56.
- Geffen E, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology* 13, 2481-2490.
- Geraldes A, Ferrand N (in press) A 7-bp insertion 3' to the HGM box suggests the duplication and concerted evolution of the rabbit SRY gene. *Genetics Selection Evolution*.
- Godinho R, Mendonça B, Crespo EG, Ferrand N (in press) Genealogy of the nuclear β -fibrinogen locus in a highly structured lizard species: comparison with mtDNA and evidence for intragenic recombination in the hybrid zone. *Heredity*.
- Grivet D, Petit RJ (2002) Phylogeography of the common ivy (*Hedera* sp.) in Europe: genetic differentiation through space and time. *Molecular Ecology* 11, 1351-1362.
- Guillot G, Mortier F, Estoup A (2005a) Geneland: a program for landscape genetics. *Molecular Ecology Notes* 5, 703-711.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005b) A spatial statistical model for landscape genetics. *Genetics* 170, 1261-1280.
- Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* 8, 461-467.
- Harding RM, McVean G (2004) A structured ancestral population for the evolution of modern humans. *Current Opinion in Genetics and Development* 14, 667-674.

- Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* **16**, 700-706.
- Harris DJ, Sá-Sousa P (2002) Molecular Phylogenetics of Iberian Wall lizards (*Podarcis*): is *Podarcis hispanica* a species complex? *Molecular Phylogenetics and Evolution* **23**, 75-81.
- Hein J, Schierup MH, Wiuf C (2005) *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press, Oxford.
- Hewitt, GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.
- Hewitt, GM (1999) Post-glacial recolonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.
- Hewitt, GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.
- Hey J (2005) On the number of New world founders: a population genetic portrait of the peopling of the Americas. *PloS Biology* **3**, 1-11.
- Hey J, Machado CA (2003) The study of structured populations--new hope for a difficult and divided science. *Nature Reviews Genetics* **4**, 535-543.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence time of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.
- Hey J, Won Y-J, Sivasundar A, Nielson R, Markert JA (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated populations. *Molecular Ecology* **13**, 909-919.
- Hofreiter M, Serre D, Rohland N, Rabeder G, Nagel D, Conrad N, Münzel S, Pääbo S. (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences, USA* **101**, 12963-12968.
- Hudson RR, Turelli M (2003) Stochasticity overrules the 'Three-Times Rule': genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* **57**, 182-190.
- Hugall A, Moritz C, Moussalli A, Stanisic J (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropic rainforest land snail *Gnarosiphia belendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences, USA* **99**, 6112-6117.
- Kauer M, Zangerl B, Dieringer D, Schlötterer C (2002) Background selection and hitchhiking - an explanation for the contrasting chromosomal patterns of microsatellite variability in African and non-African populations of *D. melanogaster*. *Genetics* **160**, 247-256.
- Kauer MO, Dieringer D, Schlötterer C (2003) A microsatellite variability screen for positive selection associated with the 'out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics* **165**, 1137-1148.
- Knowles LL (2001) Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. *Proceedings of the Royal Society London, B* **268**, 1-6.
- Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology* **17**, 1-10.
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology* **11**, 2623-2635.

- Maddison WP, Maddison DR (2004) *Mesquite: a Modular System for Evolutionary Analysis, version 1.01*. available at <http://mesquiteproject.org>.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.
- Masta SE, Laurent NM, Routman EJ (2003) Population genetic structure of the toad *Bufo woodhousii*: an empirical assessment of the effects of haplotype extinction on nested cladistic analysis. *Molecular Ecology* **12**, 1541-1454.
- Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology* **14**, 2459-2464.
- Mountain JL, Knight A, Jobin M, Gignoux C, Miller A, Lin AA, Underhill PA (2002) SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Research* **12**, 1766-1772.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Jennifer K (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.
- Newton CR, Graham A, Heptinsall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA: the amplification refractory mutation system (ARMS). *Nucleic Acids Research* **17**, 2503-2516.
- Palmé AE, Su Q, Rautenberg A, Manni F, Lascoux M (2003) Postglacial recolonization and cpDNA variation of silver birch *Betula pendula*. *Molecular Ecology* **12**, 201-212
- Paulo OS, Jordan WC, Bruford MW, Nichols RA (2002) Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian Lizard. *Molecular Ecology* **11**, 809-820.
- Petit RJ, Grivet D (2002) Optimal randomization strategies when testing the existence of a phylogeographic structure. *Genetics* **161**, 469-471.
- Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Czaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, deVries SGM, Ziegenhagen B, de Beaulieu J-L, Kremer A (2002a) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**, 49-74.
- Petit RJ, Czaikl UM, Bordács S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD, Dumolin-Lapègue S, Fineschi S, Finkeldey R, Gillies A, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Olalde M, Pemonge M, Popescu F, Slade D, Tabbener H, Tauchini D, Vries SGM, Ziegenhagen B, Kremer A (2002b) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**, 5-26.
- Petit JR, Aguinagalde I, Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palme A, Martin JP, Rendell S, Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563-1565.
- Pinho C, Ferrand N, Harris DJ (in press) Reexamination of the Iberian and North African Podarcis (Squamata: Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*.

- Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**, 1237-1245.
- Posada D and Crandall KA (2001) Intraspecific phylogenetics: trees grafting into networks. *Trends in Ecology and Evolution* **16**, 37-45.
- Queney G, Ferrand N, Weiss S, Mougél F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution* **18**, 2169-2178.
- Ray N, Currat M, Berthier P, Excoffier L (2005) Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Research* **15**, 1161-1167.
- Rendell S, Ennos RA (2002) Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. *Molecular Ecology* **11**, 69-78.
- Rogers J, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.
- Sequeira F, Alexandrino J, Rocha R, Arntzen JW, Ferrand N (2005) Genetic exchange across a hybrid zone within the Iberian endemic golden-striped salamander, *Chioglossa lusitanica*. *Molecular Ecology* **14**, 245-254.
- Slatkin M, Hudson RR (1991) Pairwise comparison for mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555-562.
- Smouse PE (1998) To tree or not to tree. *Molecular Ecology* **7**, 399-412.
- Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology* **14**, 2553-2564.
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal Human Genetics* **68**, 978-989.
- Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution* **16**, 608-613.
- Storz (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.
- Taberlet P, Cheddadi R (2002) Quaternary refugia and persistence of biodiversity. *Science* **297**, 2009-2010.
- Teixeira J, Arntzen JW (2002) Potential impact of climate warming on the distribution of the Golden-striped salamander, *Chioglossa lusitanica*, on the Iberian Peninsula. *Biodiversity and Conservation* **11**, 2167-2176.
- Teixeira J, Ferrand N, Arntzen JW (2001) Biogeography of the golden-striped salamander, *Chioglossa lusitanica*: a field survey and spatial modelling approach. *Ecography* **24**, 618-624.
- Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**, 381-397.
- Templeton AR (2002a) 'Optimal' randomization strategies when testing the existence of a phylogeographic structure: a reply to Petit and Grivet. *Genetics* **161**, 473-475
- Templeton AR (2002b) Out of Africa again and again. *Nature* **416**, 45-51.
- Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**, 789-809.

- Tishkoff SA, Varkonyi R, Cahinhian N, Abbas S, Argyropoulos G, Destro-Bisol G, Drousioutou A, Danbergfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Touma EH, Williams SM, Clark AG (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455-462.
- Van Houdt JKJ, Cleyn DE, Perretti A, Volckaert AM (2005) A mitogenic view on the evolutionary history for the Holarctic freshwater gadoid, burbot (*Lota lota*). *Molecular Ecology* **14**, 2445-2457.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811-1823.
- Volckaert FAM, Hänfling B, Hellmans B, Carvalho GR (2002) Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene. *Journal of Evolutionary Biology* **15**, 930-944.
- Wakeley J (2004) Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. In: *The Evolution of Population Biology* (eds. Singh RS, Uyenoyama MK), pp. 193-215. Cambridge University Press, Cambridge.
- Weiss S, Antunes A, Schlötterer C, Alexandrino P (2000) Mitochondrial haplotype diversity among Portuguese brown trout *Salmo trutta L.* populations: relevance to the post-Pleistocene recolonization of northern Europe. *Molecular Ecology* **9**, 691-698.
- Weiss S, Persat H, Eppe R, Schlötterer C, Ublein F (2002) Complex patterns of colonization and refugia revealed for European grayling *Thymallus thymallus*, based on complete sequencing of the mtDNA control region. *Molecular Ecology* **11**, 1393-1407.
- Won Y-J, Sivasundar A, Wang Y, Hey J (2005) On the origin of Lake Malawi cichlids: a population genetic analysis of divergence. *Proceedings of the National Academy of Sciences, USA* **102**, 6581-6586.
- Yang Z (1997) On the estimation of ancestral population sizes of modern humans. *Genetical Research* **69**, 111-116.
- Zhang DX, Hewitt GM (2003) Nuclear DNA analysis in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**, 563-584.

