38 Jander, G. *et al.* (2001) The *TASTY* locus on chromosome 1 of *Arabidopsis* affects feeding of the insect herbivore *Trichoplusia ni*. *Plant Physiol.* 126, 890–898
39 Hering, E.M. (1957) *Blattminen von Europa*, Uitgeverij Dr W. Junk
40 Kwiatowski, J. and Ayala, F.J. (1999) Phylogeny of *Drosophila* and related genera: Conflict between molecular and anatomical analyses. *Mol. Phylog. Evol.* 13, 319–328
41 Feeny, P. (1976) Plant apparency and chemical defense. In *Biochemical Interaction Between Plants and Insects* (Wallace, I. *et al.*, eds), pp. 1–40, Plenum Press
42 Schenk, P.M. *et al.* (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11655–11660
43 Stotz, H.U. *et al.* (2000) Induced plant defense responses against chewing insects. Ethylene signaling reduces resistance of *Arabidopsis* against Egyptian cotton worm but not diamondback moth. *Plant Physiol.* 124, 1007–1017
44 Reymond, P. *et al.* (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* 12, 707–720
45 Mauricio, R. (1998) Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *Am. Nat.* 151, 20–28
46 Siemens, D. and Mitchell-Olds, T. (1998) Evolution of pest induced defenses in *Brassica* plants: tests of theory. *Ecology* 79, 632–646
47 Stowe, K.A. (1998) Experimental evolution of resistance in *Brassica rapa*: Correlated response of tolerance in lines selected for glucosinolate content. *Evolution* 52, 703–712
48 Wittstock, U. and Halkier, B.A. (2000) Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. *J. Biol. Chem.* 275, 14659–14666
49 Kliebenstein, D.J. *et al.* (2001) Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. *Plant Cell* 13, 681–693
50 Kliman, R.M. *et al.* (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156, 1913–1931
51 McKay, J.K. *et al.* (2001) Local adaptation despite absence of marker diversity in the rare Sapphire Rockcress. *Proc. R. Soc. London B Biol. Sci.* 268, 1715–1721
52 Settles, A.M. and Byrne, M. (1998) Opportunities and challenges grow from *Arabidopsis* genome sequencing. *Genome Res.* 8, 83–85

Prospects for nuclear gene phylogeography

Matthew P. Hare

In phylogeography, an empirical focus on gene lineages enables the history of population processes to be inferred from the simultaneous analysis of temporal and spatial patterns. Rapidly evolving cytoplasmic DNA has been the empirical workhorse propelling the success of this nascent field. Now, as more sophisticated historical models are being tested, there is a growing need for phylogeography to expand from a largely marker-specific discipline to a more general analytical approach that can be applied across independent loci. Recent results using nuclear haplotypes to study phylogeography indicate that the anticipated technical and biological hurdles can be overcome in many taxa to achieve phylogeographical comparisons across unlinked loci. Although many challenges remain, a more complete understanding of the historical, demographic and selective processes shaping phylogeographical patterns is emerging.

Matthew P. Hare
Dept of Biology, University of Maryland, College Park, MD 20742, USA.
e-mail: matt.hare@umail.umd.edu

Over the past decade, phylogeography grew as a discipline because allelic phylogenies provided explicitly historical tools for the study of geographical subdivision among populations[1]. Analysing the relative ages and historical relationships of alleles in a geographical context could distinguish ongoing processes such as GENE FLOW (see Glossary) from previous events such as range expansion. The temporal resolution offered by genealogies was previously unavailable through classical population genetic analyses (e.g. hierarchical partitioning of variation using *F*-statistics)[2]. In addition, using alleles rather than populations as the basic unit in phylogenetic clustering promoted integrated analyses of history above and below the species level without the need to make *a priori* taxonomic distinctions. Animal mitochondrial (mt)DNA and chloroplast (cp)DNA have been the primary data sources making phylogeography such a productive empirical approach. Extending and generalizing phylogeography by using nuclear HAPLOTYPE data has been a desirable goal in efforts to test ever more sophisticated historical models. However, given the potential for technical and biological complications when analysing nuclear haplotype data at the intraspecific level, it has been uncertain whether a phylogeographical study design could generally be extended to noncytoplasmic markers[1]. An evaluation of achievements made to date in nuclear phylogeography arguably provides the most practical guide to its future potential.

A gene tree for a single cytoplasmic or nuclear (nDNA) LOCUS provides a slim and sometimes misleading representation of the population histories through which alleles were transmitted[3,4]. Because cytoplasmic loci are usually inherited uniparentally, they will not have genealogical patterns that are representative of the entire population history, especially when sex biases have affected fitness or dispersal behavior[5]. Loci under selection can also have genetic patterns that deviate from expectations based on population history and demography, yet the selected patterns often mimic alternative demographic histories. Testing alternative hypotheses about evolutionary forces must ultimately rely on the fact that selection acts locally within a genome, whereas population demography leaves a common signature across all neutral loci[1,6].

 0169-5347/01/$ – see front matter PII: S0169-5347(01)02326-6

Box 1. Intragenic recombination

If the intragenic recombination rate approaches the nucleotide substitution rate at a locus, the haplotypes will have more than one immediate ancestor and different segments within a haplotype will have independent histories. Ignoring recombination can ruin the molecular clock, undermine phylogeny reconstruction by introducing homoplasy, and can also bias the gene tree shape in a fashion similar to population expansion[a]. Methods are improving for recombination rate estimation and the identification of recombinants in an alignment of sequences[b,c]. Three main techniques have been used to reconstruct gene trees in spite of the occurrence of some recombination: (1) if detectable recombinants are few, they can be removed before[d] or in conjunction with phylogenetic analysis[e]; (2) if linked clusters of polymorphic sites separate apparent recombination break points within a locus, then a gene tree can be reconstructed from each partition[e]; and (3) a network can be constructed that depicts the reticulate haplotype relationships caused by recombination[f].

Two aspects of recombination suggest that genetic shuffling within populations will not obscure strong phylogeographical patterns. First, reduced gene flow between isolated populations could restrict the effects of recombination to within each population, allowing dichotomous phylogenetic structure to build between populations (e.g. Ref. g). Second, spatially heterogeneous recombination rates along chromosomes provide the opportunity to choose a level of recombination most desired for the goal at hand[h]. For phylogeography, chromosomal regions of low recombination could be desirable if independent selective sweeps accelerate differentiation between populations[i]. Alternatively, regions of high recombination can provide favorable conditions for the analysis of historical demography[j].

References

a Schierup, M.H. and Hein, J. (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156, 879–891

b Kuhner, M.K. *et al.* (2000) Maximum likelihood estimation of recombination rates from population data. *Genetics* 156, 1393–1401

c Crandall, K.A. and Templeton, A.R. (1999) Statistical methods for detecting recombination. In *The Evolution of HIV* (Crandall, K.A., ed.), pp. 153–176, Johns Hopkins University Press

d Harding, R.M. *et al.* (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* 60, 772–789

e Templeton, A.R. *et al.* (2000) Cladistic structure within the human *lipoprotein lipase* gene and its implications for phenotypic association studies. *Genetics* 156, 1259–1275

f Posada, D. and Crandall, K.A. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* 16, 37–45

g Bernardi, G. *et al.* (1993) Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. *Proc. Natl. Acad. Sci. U. S. A.* 90, 9271–9274

h Kaessmann, H. *et al.* (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat. Genet.* 22, 78–81

i Stephan, W. *et al.* (1998) A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5649–5654

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

The limitations of inferences made from single gene trees have been made more apparent by the increasingly complex phylogeographical hypotheses being tested in the context of conservation and evolutionary studies (e.g. Refs 7,8). Powerful advances in COALESCENT THEORY provide the framework for testing complex histories[6], but no matter how sophisticated the theory is, alternate explanations for data patterns often cannot be distinguished without comparisons across independent loci. The ultimate strength and broader applicability of a phylogeographical approach therefore depends on the ability to incorporate nuclear loci. In keeping with traditional methods employed in phylogeography, I focus here on haplotype-based genealogical approaches in spite of the growing phylogeographical utility reported for nuclear fragment data such as amplified fragment length polymorphisms and microsatellites[9].

Prospects for nuclear phylogeography depend on both the technical feasibility of acquiring nuclear haplotype data in different taxa, and on the usefulness of those data for inferring processes that structure the spatial distribution of gene lineages (Boxes 1,2). These prospects are assessed here with a survey of nDNA results in phylogeography. First, I examine nuclear gene tree applications where the focus is on questions and approaches borrowed by phylogeography from systematics. I then shift attention to a broader window of inquiry by examining empirical results from another parent discipline of phylogeography, population genetics, to consider how a focus on biogeography continues to sharpen the historical and demographic inferences made from genealogical comparisons across loci.

Traditional phylogeography applied to nuclear DNA: theory

The marker of first choice for phylogeography will continue to be cytoplasmic DNA for the initial characterization of population structure, testing population MONOPHYLY, or inferring maternal gene flow[10,11]. Studies with these traditional phylogeographical goals often use nuclear data to corroborate initial results based on cytoplasmic loci. In theory, however, phylogeographical structure is expected to be less pronounced at diploid nuclear loci compared with cytoplasmic loci because of their different effective population sizes (N_e)[11,12]. The N_e of a locus is related to the number of breeding adults, but might be smaller or larger than that number depending on ploidy and mode of inheritance. In dioecious populations where males and females are equally abundant and have equal variance in reproductive success, autosomal nuclear loci have an N_e four times larger than that of uniparentally inherited cytoplasmic markers (see Ref. 5 for exceptions). Thus, under a neutral model of evolution, genetic drift will cause divergence between isolated populations to occur four times more slowly at nuclear

Box 2. Target loci for nuclear phylogeography

In many taxa, a lower substitution rate in nuclear DNA (nDNA) minimizes the back and parallel mutations that can lower phylogenetic resolution in mitochondrial DNA (mtDNA) data. Thus, in data with low homoplasy, even a single fixed difference can provide a statistically strong result at the intraspecific level, regardless of bootstrap support[a]. Nonetheless, nuclear sequences will often need to be longer than those typically sampled from mtDNA to assure adequate sampling of phylogenetically informative characters.

Noncoding nDNA typically has more variation than does the adjacent coding sequence. Intron DNA has frequently been targeted by using GenBank database sequences to design primers in two relatively conserved exons, and performing Exon Primed – Intron Crossing (EPIC) PCR (Ref. b). Although the taxonomic universality of EPIC primers has often been wanting, their development has promoted the collection of nDNA data in invertebrates[b], vertebrates[c–f] and plants[g]. Methods have also been devised to isolate random segments of genomic DNA with unknown function[h], or to screen these 'anonymous' loci for segments with relatively high polymorphism[i]. Data are not sufficient to recommend one locus or method over another.

When using PCR to isolate a DNA segment, the potential exists to amplify paralogs related by a gene duplication event and located at more than one chromosomal locus, in addition to orthologous alleles from each locus. Whether multiple loci are co-amplified will sometimes be sensitive to reaction conditions in the PCR (Ref. j). If intron sizes vary among paralogous loci, initial EPIC products can be sorted by size and individually sequenced for the design of locus-specific primers. Ultimately, distinguishing between paralogous and allelic variants might require tests of Hardy–Weinberg ratios within populations[j] or pedigree analysis.

References

a Harris, E.E. and Hey, J. (1999) X chromosome evidence for ancient human histories. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3320–3324

b Palumbi, S.R. (1996) The polymerase chain reaction. In *Molecular Systematics* (Hillis, D. *et al.*, eds), pp. 205–247, Sinauer

c Friesen, V.L. *et al.* (2000) Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. *Mol. Ecol.* 8, 2147–2149

d Quattro, J.M. and Jones, W.J. (1999) Amplification primers that target locus-specific introns in actinopterygian fishes. *Copeia* 1999, 191–196

e Friesen, V.L. *et al.* (1997) Intron variation in marbled murrelets detected using analysis of single-stranded conformational polymorphisms. *Mol. Ecol.* 6, 1047–1058

f Lyons, L.A. *et al.* (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* 15, 47–56

g Strand, A.E. *et al.* (1997) Nuclear DNA-based markers for plant evolutionary biology. *Mol. Ecol.* 6, 113–118

h Karl, S.A. and Avise, J.C. (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* 256, 100–102

i Bagley, M.J. and Gall, G.A.E. (1998) Mitochondrial and nuclear DNA sequence variability among populations of rainbow trout (*Oncorhynchus mykiss*). *Mol. Ecol.* 7, 945–961

j Hare, M.P. *et al.* (1996) Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Mol. Biol. Evol.* 13, 334–345

than at cytoplasmic loci, all else being equal. Differential dispersal by gender can further affect the relative magnitude of phylogeographical structure in biparentally and uniparentally inherited markers[12,13], but the stochastic variation accompanying genealogical differentiation can make it difficult to resolve these population processes from cytonuclear comparisons[14].

The application of phylogeography to conservation has often carried an emphasis on testing for monophyletic groups because this evidence for long-term isolation argues strongly for evolutionary distinctiveness or 'significance' of a population[1]. Genetic drift will inevitably lead to neutral gene monophyly in any population isolated for long enough, but for the reasons stated above, the approach to monophyly at nuclear autosomal loci is too slow to commonly expect deep intraspecific partitions, such as those found in animal mtDNA (Refs 1,15). Populations or species that diverged in the Pleistocene should occupy, in many cases, a 'mixed-monophyly' zone of divergence where mtDNA will have a high probability of monophyly, but the average nuclear locus will not. Mixed-monophyly probabilities are calculable under a neutral model based on cytoplasmic gene tree patterns and the 'three-times rule' (Box 3). The abundant mtDNA and cpDNA data available from many taxa therefore provide an initial null expectation for the degree of gene tree concordance across loci[15].

These theoretical comparisons establish a low expectation of nuclear monophyly among recent population isolates, but this is not necessarily a drawback for nuclear phylogeography. The temporal depth over which population processes can be inferred is limited in a monophyletic population by the age of the most recent common ancestral allele (the coalescent). Therefore, for questions about more ancient populations, the relatively deep coalescent times at nuclear loci are essential[6,16].

In general, interpreting population history and demography from POLYPHYLETIC gene patterns requires analytical methods that make use of both genealogical allelic relationships and allele frequency information. Aided by the use of coalescent theory[6], two analytical advances in phylogeography have strengthened statistical testing of historical models (Box 4). First, maximum parsimony testing of spatial and temporal hypotheses has been made more rigorous with a nested-clade statistical framework[2,17]. Second, population subdivision and gene flow can be tested with maximum likelihood estimation of parameters under alternative hypotheses[18,19], or by comparison of observed patterns with those simulated under alternative models using the coalescent process[8]. Incorporation of these analytical tools into phylogeography has proceeded slowly, but methods such as likelihood offer the promise that, with strong data sets, multiple population processes can be examined in an integrated fashion[18–20].

Box 3. The three-times rule

The three-times rule is a simple metric suggested by the fourfold theoretical difference in average times to monophyly for nuclear versus mitochondrial loci[a]. The rule can be explained diagrammatically (Fig. I) with two gene trees for mitochondrial DNA (mtDNA), one depicted early and the other depicted late in the divergence of populations A and B. If clade A is monophyletic, the intraclade sequence diversity (as measured by average sequence divergence, for example) provides a population-specific estimate of the time (x) that was required for mtDNA to become monophyletic. At time x, the average autosomal nuclear locus will still be polyphyletic, but is expected to become monophyletic after an additional 3x time units. The lengths of phylogenetic branches connecting mtDNA sister clades A and B are an indication of how long ago monophyly evolved at mtDNA. If the interclade branches are much shorter than three times the intraclade diversity, as in Fig. Ia, then monophyly at mtDNA is recent and most nuclear loci are likely to be polyphyletic or paraphyletic. If the interclade branches are much longer than three times the intraclade diversity, as in Fig. Ib, monophyly is expected at the majority of nuclear loci. Thus, when the interclade divergence:intraclade diversity ratio is much greater or much less than 3:1 on a mtDNA tree, the pattern provides information about the likelihood of nuclear monophyly[a].

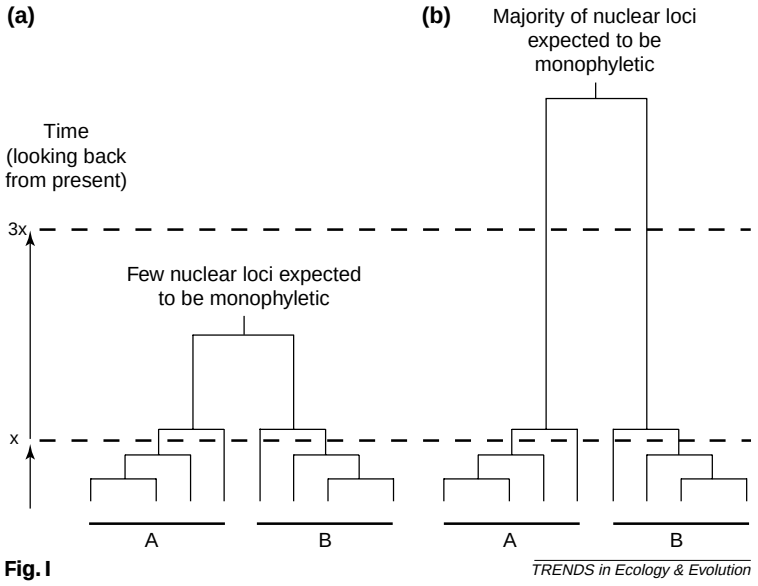

Fig. I

Reference

a Palumbi, S.R. *et al.* (2001) Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* 55, 859–868

Traditional phylogeography applied to nuclear DNA: results

Two practical concerns have been raised about using nuclear haplotypes for phylogeography[1]. First, phylogenetic methods of gene tree reconstruction could be impeded by recombination (Box 1). Second, poor resolution could result from low mutation rates, leading to too few informative polymorphisms (Box 2). The nuclear phylogeography studies summarized in Table 1 suggest that these complications will not necessarily be pervasive or debilitating. Some studies report low levels of inferred recombination[21,22], even when haplotypes longer than one kb were sampled[16,23,24]. Furthermore, both coding and noncoding sequences have contained ample polymorphism for robust gene trees in a broad array of taxa (Table 1).

The quantitative magnitude of population subdivision evidenced by nuclear gene trees has varied, as one would expect (Table 1). In exceptional cases, monophyletic nuclear clades indicated long-standing isolation of populations[22,25]. Genetic differentiation detected at higher intraspecific levels, that is, deep in an intraspecific gene tree, was often associated with minimal phylogeographical structure among the shallow (more recent) gene tree branches[7,26,27]. Human population structure between Africa and Asia was observed in some β-globin haplotypes estimated to be as young as 150 000 yrs and others as old as 800 000 yrs (Ref. 16). Among Alaska populations of marbled murrelet *Brachyramphus marmoratus* that diverged in the late Pleistocene, maximum likelihood analysis of intron variation at nine nuclear loci demonstrated a history of asymmetric gene flow between Aleutian Island and mainland populations[7]. By contrast, intron variation in recently colonized populations of the Medfly *Ceratitis capitata* showed no phylogeographical sorting of ancestral African lineages[28]. In some cases, the resolution of nuclear phylogeography has been extended towards more recent evolutionary processes by combining the analysis of sequence haplotypes with other more rapidly evolving markers[29,30].

A polyphyletic relationship among recently isolated populations is often caused by the persistence of ancestral polymorphisms, but similar genealogical patterns can be caused by moderate gene flow over longer divergence times[31]. Likelihood methods to distinguish these mechanisms are improving[20] (Box 4), but are still somewhat restrictive in their assumptions. In some cases, classical phylogeographical inference can distinguish gene flow from ancestral polymorphism. In a study of genetically divergent oyster *Crassostrea virginica* populations, nuclear restriction fragment length polymorphisms (RFLP) were deeply introgressed in both directions across a steep genetic cline in Florida (USA)[32]. The two populations, sampled far away from the cline, were found to be polyphyletic at three independent nuclear loci[33]. Recent introgression was rejected as an explanation for nuclear polyphyly, because identical haplotypes were observed at distant localities within each marine basin, consistent with long-distance gene flow, but no haplotypes were shared between the basins. Another example involves an important root crop species, cassava *Manihot esculenta esculenta*. The restricted geographical distribution of *G3pdh* alleles shared between cassava and its potential *Manihot* progenitors localized the center of origin for domestication to the southern border of the Amazon basin[24]. The strength of

Box 4. Analytical advances for testing phylogeographical hypotheses

The potential power of phylogeographical analysis comes from the explicitly temporal information content of gene trees; however, the analytical and statistical tools needed to interpret genealogical data are still being developed. By combining spatial and temporal tests of association, nested clade analysis detects geographical structure with more power than that achieved with traditional fixation indices[a]. For example, in spite of a lack of evidence for population structure based on F_{st} analysis of alcohol dehydrogenase (*Adh*) haplotypes in *Drosophila melanogaster* from the eastern USA, nested analysis of the gene tree demonstrated significant associations between allele clades and sampling localities[a]. More importantly, nested clade analysis can distinguish ongoing population processes from historical events, such as colonization or range expansion. These single-locus phylogeographic statistics are based on random permutations of contingency data tables at various nested levels. The availability of computer programs has recently increased the accessibility of these important procedures[b] (see http://bioag.byu.edu/zoology/crandall_lab/).

Maximum likelihood has the potential to improve tests of population structure and estimates of gene flow by using all of the information in the data in a coalescent framework. Currently available programs for use with sequence haplotypes (and other types of data) differ mostly in the underlying mutation model, finite sites[c] (see MIGRATE at http://evolution.genetics.washington.edu/lamarc.html) or infinite sites[d] (GENETREE at http://ftp.monash.edu.au/pub/gtree/). Data should have no homoplasy to be compatible with the infinite sites model. Parameter estimates made with MIGRATE integrate over all possible genealogies and migration events among two or more populations[e], whereas results from GENETREE are conditional on the unique gene tree described by DNA sequences. Data subsets inconsistent with the infinite sites model are typically removed before analysis with GENETREE. The infinite sites model has also been used in a Bayesian and likelihood framework to estimate jointly divergence time and migration rate for two populations[f]. These joint estimates can distinguish low gene flow connecting recent populations from moderate gene flow over a longer divergence. Care must be taken with these methods, because they rely on computationally demanding simulations and it can be difficult to evaluate how many simulations are required for reliable results. The strengths of nested clade analysis and coalescent simulations can be mutually reinforcing and can combine to strengthen inferences about population processes in some cases[g].

References

a Templeton, A.R. (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* 7, 381–397

b Posada, D. *et al.* (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9, 487–488

c Beerli, P. and Felsenstein, J. (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152, 763–773

d Bahlo, M. and Griffiths, R.C. (2000) Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* 57, 79–95

e Beerli, P. and Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations using a coalescent approach. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4563–4568

f Nielsen, R. and Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158, 885–896

g Carbone, I. and Kohn, L.M. (2001) A microbial population-species interface: nested cladistic and coalescent inference with multilocus data. *Mol. Ecol.* 10, 947–964

phylogeographical inferences such as these depends on adequate geographical sampling[2].

Intense efforts have been made to obtain nuclear phylogeographical data bearing on human origins and migrations (e.g. Refs 16,23,34). These phylogeographical studies have been successful at demonstrating subdivision among human populations, and are also exemplary in utilizing the rapid accumulation of data from independent loci to distinguish between patterns caused by historical demography from those caused by selection. For example, the fixed nucleotide difference found at the X-linked pyruvate dehydrogenase *E1* locus between African and non-African humans would, by itself, suggest severe limits to gene flow[23]. Comparison of variation at this locus with other X-linked loci, however, suggested that selection influenced pyruvate dehydrogenase patterns outside of Africa[23]. As phylogeography expands beyond cytoplasmic markers, its ability to explore the interactions of demographic and selective forces shaping spatial patterns of diversity also expands.

Selection, historical gene flow and speciation

Dichotomizing the genome into nuclear and cytoplasmic expectations (e.g. with the three-times rule) is a crude simplification justified only by the present state of data available from most taxa. Ideally, when inferences about population history and demography are supported by multiple independent loci, they provide an historical context within which the effects of locus-specific selection become more recognizable and interpretable. In addition, the broad time frame over which nuclear genealogies retain information about population demography means that, in principle, population processes that occurred before, during and after speciation can be studied[35,36].

Under the biological species concept, gene flow is expected among populations but not between species. This view has made intraspecific patterns the primary focus of phylogeography[1]. However, species boundaries are poorly known in most taxa, are blurry during speciation and can be semipermeable to gene flow after speciation. The study of differential gene flow among loci is one of the exciting opportunities made possible by the increased availability of nuclear genealogies. If locus-specific patterns of differentiation are demonstrated to be selection mediated, they can help explicate speciation mechanisms.

Differential gene flow was implicated by the variation in genealogical patterns observed among

Table 1. Selected studies using nuclear haplotype genealogies to study phylogeography

Taxon	Type of locus	Method of haplotype determination	Intragenic recombination?	Gene tree?	Population monophyly?	Comments on nDNA patterns	Refs
Cassava *Manihot esculenta*	Exons and introns	Direct sequence and inferred haplotypes	None	Yes	No	Gene tree indicated single progenitor species; location of shared alleles suggested site of domestication	24
Ascomycete fungus *Coccidioides immitis*	Five loci, exons and introns	Haploid, direct sequence	None	Yes (five)	Yes	All loci showed concordant phylogeographical partition of California and non-California isolates	22
Human *Homo sapiens*	Exons and introns	X-linked, sampled in males, direct sequence	None	Yes	Yes	Non-African samples were monophyletic, but selection appeared to be important	23
Tidepool copepod *Tigriopus californicus*	Exon and mtDNA	Isofemale lines, direct sequence	Not reported	Yes	Yes	nDNA and mtDNA agreement on deep phylogeographic partitions	25
Humpback whale *Megaptera novaengliae*	Intron	Multiple clones sequenced	Not reported	Yes	No	Two major clades, but no geographical structure, unlike mtDNA	13
Rainbow trout *Oncorhynchus mykiss*	Six loci, four anon. and two introns, mtDNA	SSCP-sequence or direct sequence and inferred haplotypes	Low, one locus only	Yes	No	Ancestral and/or introgressed variation limited resolving power, although contrast with mtDNA was informative	21
Black tiger prawn *Penaeus monodon*	Intron	Multiple clones sequenced	Not reported	Yes	Yes	Western Pacific monophyletic relative to western Indian Ocean	26

loci in the chimps *Pan troglodytes* and *P. paniscus*[27] and in *Drosophila*[37]. In *Drosophila*, the abundance of polymorphisms shared by *D. pseudoobscura* and *D. persimilis* at the *Adh* locus was incompatible with the number of fixed differences observed at two X-linked loci under a neutral model with no gene flow after speciation. With the benefit of additional information on prezygotic reproductive barriers within the overlapping ranges of these two species, selection was hypothesized to limit gene flow at the X-linked loci in a manner consistent with a sympatric speciation model[37]. Thus, the phylogeographical patterns observed within and among species, compared across loci, provided important clues to the underlying evolutionary processes.

Perhaps cases of interspecies gene flow simply beg the question of what represents a species boundary. Defining the interface between populations connected by gene flow and species that are genetically discrete is a goal that cannot realistically be attained with single marker studies. In the plant parasitic ascomycete fungus *Sclerotinia sclerotiorum* and its close relatives, nuclear phylogeographical patterns of DNA sequence haplotypes from several loci were used to test for intra- and interlocus recombination and for population processes of fragmentation and dispersal within and between species[38]. A low level of recombination permitted the removal of its effects before nested cladistic and coalescent analyses. Both analyses supported fragmentation of *S. sclerotiorum* populations in association with ecological conditions and some host specificity, followed by dispersal among locations and among hosts. Gene flow was also inferred between species, but was confined to taxa in similar geographical and ecological habitats. Fragmented populations within *S. sclerotiorum* had all the marks of distinct species (e.g. lack of ongoing gene flow), but the time depth of fragmentation was considered too shallow for species status compared with the most recent common ancestor between species[38]. Regardless of whether this relational criterion adequately defines species boundaries, this study exemplifies a valuable approach to studying the temporal and geographical occurrence of population processes free from *a priori* assumptions based on taxonomy.

Conclusions and prospects

Based on the results and insights provided by intraspecific nuclear gene trees published to date, it is clear that the extension of phylogeographical analyses to diploid nuclear loci is both feasible and productive. As methods are developed to ameliorate potential complications with nuclear genealogies, phylogeography will grow from a largely marker-specific discipline to a broader arena within evolutionary genetics, where the importance of spatial patterns is emphasized in studies of historical population processes.

Practically speaking, the utility of nDNA for phylogeography in individual cases will need to be evaluated empirically, aided by two general rules of thumb. First, the influence of selection on phylogeographical patterns is to be expected.

Glossary

Coalescent theory: a branch of population genetic theory that estimates population parameters based on the distribution of allele coalescence, going back in time from a sample of alleles to their most recent common ancestor.
Gene flow: the transfer of genetic material from one population to another by dispersal and successful reproduction.
Haplotype: a unique haploid sequence of nucleotides at multiple polymorphic sites.
Locus: a segment of DNA that evolves independently from other loci because of recombination or independent assortment.
Monophyly: the phylogenetic sharing of a common ancestor by a group of haplotypes, defining a clade that is exclusive of other haplotypes. A population is referred to as being monophyletic if all of its sampled haplotypes are more closely related to each other than to any haplotype from another population.
Polyphyletic: when haplotypes from one population are genealogically more closely related to haplotypes in another population than they are to haplotypes in the same population, and vice versa.

Therefore, strong demographic and historical inferences require character concordance across independent loci. Second, with a few exceptions[5], phylogeographical sorting of nDNA polymorphisms will be slower than for mtDNA.

Nuclear phylogeography offers an opportunity to understand more completely the mosaic of genealogical patterns evolving in genomes as a response to the history and environment that populations have experienced. In many cases, the historical clarity apparent in mtDNA patterns will have told only part of the story, belying the complex amalgam of processes that shaped genetic diversity. The high resolution of mtDNA data will continue to be an asset and a priority in phylogeographical studies, but the genomic landscape within which mtDNA must be interpreted will be obscure until the phylogeography of nuclear variation is also studied.

Acknowledgements
Work on this article was made possible in part by support from the Hrdy Fellowship at Harvard University. I thank J. Avise, P. Barber, S. Cohen, S. Lavery, R. Nielsen, K. Shaw and three anonymous reviewers for valuable comments. My apologies to those who have advanced nuclear phylogeography but who could not be cited here because of space constraints.

References

1 Avise, J.C. (2000) *Phylogeography: the History and Formation of Species*, Harvard University Press
2 Templeton, A.R. (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* 7, 381–397
3 Cronin, M.A. (1993) Mitochondrial DNA in wildlife taxonomy and conservation biology: cautionary notes. *Wildl. Soc. Bull.* 21, 339–348
4 Harpending, H.C. *et al.* (1998) Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. U. S. A.* 95, 1961–1967
5 Hoelzer, G.A. (1997) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution* 51, 622–626
6 Fu, Y-X. and Li, W-H. (1999) Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theor. Popul. Biol.* 56, 1–10
7 Congdon, B.C. *et al.* (2000) Mechanisms of population differentiation in marbled murrelets: historical versus contemporary processes. *Evolution* 54, 974–986
8 Ford, M.J. (1998) Testing models of migration and isolation among populations of chinook salmon (*Oncorhynchus tschawytscha*). *Evolution* 52, 539–557
9 Sunnucks, P. (2000) Efficient genetic markers for population biology. *Trends Ecol. Evol.* 15, 199–203
10 Avise, J.C. (1995) Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conserv. Biol.* 9, 686–690
11 Moore, W.S. (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49, 718–726
12 McCauley, D.E. (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends Ecol. Evol.* 10, 198–202
13 Palumbi, S.R. and Baker, C.S. (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Mol. Biol. Evol.* 11, 426–435
14 Buonaccorsi, V.P. *et al.* (2001) Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Mol. Ecol.* 10, 1179–1196
15 Palumbi, S.R. *et al.* (2001) Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* 55, 859–868
16 Harding, R.M. *et al.* (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* 60, 772–789
17 Posada, D. *et al.* (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9, 487–488
18 Beerli, P. and Felsenstein, J. (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152, 763–773
19 Bahlo, M. and Griffiths, R.C. (2000) Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* 57, 79–95
20 Nielsen, R. and Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158, 885–896
21 Bagley, M.J. and Gall, G.A.E. (1998) Mitochondrial and nuclear DNA sequence variability among populations of rainbow trout (*Oncorhynchus mykiss*). *Mol. Ecol.* 7, 945–961
22 Koufopanou, V. *et al.* (1997) Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5478–5482
23 Harris, E.E. and Hey, J. (1999) X chromosome evidence for ancient human histories. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3320–3324
24 Olsen, K.M. and Schaal, B.A. (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5586–5591
25 Burton, R.S. (1998) Intraspecific phylogeography across the Point Conception biogeographic boundary. *Evolution* 52, 734–745
26 Duda, T.F. and Palumbi, S.R. (1999) Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. *Mar. Biol.* 134, 705–710
27 Kaessmann, H. *et al.* (1999) Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286, 1159–1162
28 Villablanca, F.X. *et al.* (1998) Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Mol. Ecol.* 7, 547–560
29 Carbone, I. *et al.* (1999) Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. *Evolution* 53, 11–21
30 Tishkoff, S. *et al.* (2001) Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* 293, 455–462
31 Wakeley, J. (1995) Distinguishing migration from isolation using the variance of pairwise differences. *Theor. Popul. Biol.* 49, 369–386
32 Karl, S.A. and Avise, J.C. (1992) Balancing selection at allozyme loci in oysters: Implications from nuclear RFLPs. *Science* 256, 100–102
33 Hare, M.P. and Avise, J.C. (1998) Population structure in the American oyster as inferred by nuclear gene genealogies. *Mol. Biol. Evol.* 15, 119–128
34 Kaessmann, H. *et al.* (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat. Genet.* 22, 78–81
35 Wakeley, J. and Hey, J. (1997) Estimating ancestral population parameters. *Genetics* 145, 847–855
36 Nichols, R. (2001) Gene trees and species trees are not the same. *Trends Ecol. Evol.* 16, 358–364
37 Wang, R.L. *et al.* (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* 147, 1091–1106
38 Carbone, I. and Kohn, L.M. (2001) A microbial population–species interface: nested cladistic and coalescent inference with multilocus data. *Mol. Ecol.* 10, 947–964