

Phylogeographic studies in plants: problems and prospects

B. A. SCHAAL, D. A. HAYWORTH, K. M. OLSEN, J. T. RAUSCHER and W. A. SMITH

Department of Biology, Campus Box 1137, Washington University, St Louis, MO 63130, USA

Abstract

Genetic structuring of plant populations is strongly influenced by both common ancestry and current patterns of interpopulation genetic exchange. The interaction of these two forces is particularly confounding and hence interesting in plants. This complexity of plant genetic structures is due in part to a diversity of reproductive ecologies affecting genetic exchange and the fact that reproductive barriers are often weak between otherwise morphologically well-defined species. Phylogeographic methods provide a means of examining the history of genetic exchange among populations, with the potential to distinguish biogeographic patterns of genetic variation caused by gene flow from those caused by common ancestry. With regard to plants, phylogeography will be most useful when applied broadly across the entire spectrum of potential genetic exchange. Although current phylogeographic studies of plants show promise, widespread application of this approach has been hindered by a lack of appropriate molecular variation; this problem is discussed and possible solutions considered.

Keywords: cpDNA, nDNA, phylogenetics, phylogeography, plants, population genetics

Introduction

Plants demonstrate an astonishing diversity of morphology, adaptation, and ecology, the product of millions of years of lineage divergence and diversification. Characterizing this diversity and understanding the mechanisms through which it arises is the realm of population genetics and systematics alike. Avise *et al.* (1987) previously presented the concept of phylogeography, which seemed to promise an important step forward in our quest to understand these processes. During the past 10 years phylogeography has had a major impact on research in animal systems (see other articles in this issue), but has produced very few explicitly phylogeographic studies in plants.

We suggest that this paucity of plant studies exists not because phylogeography is less applicable or useful in plants. In fact, we believe phylogeography to be especially appropriate considering some of the unique characteristics of plant evolution. The problem has been a lack of useful genetic variation applicable to a phylogeographic analysis. In this paper, we discuss several issues in plant

evolution for which phylogeographic analyses will prove to be an important tool, with reference to current phylogeographic work. We conclude with a discussion of the search for appropriate genetic variation in plants and the specific technical limitations and practical issues of applying phylogeographic methodology to plants.

The determinants of population genetic structure

Patterns of genetic exchange

The genetic structuring of a population of organisms, and ultimately the establishment of independent evolutionary lineages, is strongly influenced by the pattern of genetic exchange (gene flow) within and between populations. Variation in factors influencing the reproductive ecology of plants can have profound effects on the nature of this exchange. One such factor is breeding system. At one extreme are taxa such as oaks (*Quercus* spp.), where a lack of reproductive barriers facilitates gene flow, not only between populations of the same species, but between well-diverged species. Consequently, in oaks intraspecific population structure continues to be influenced by inter-specific genetic exchange following the phenotypic divergence of species. At the other extreme are plants such as

Correspondence: B. A. Schaal. Fax: +1-(314)-935-4432; E-mail: schaal@biodec.wustl.edu

dandelions (*Taraxacum officinale*), which in the USA reproduce exclusively by apomixis (parthenogenesis). Each maternal lineage within a population is genetically isolated from every other lineage. Thus, in dandelions, separate phylogenetic lineages are established within populations, at a point where one usually expects free genetic exchange and recombination.

The actual mechanism by which genes migrate can also be an important determinant of genetic exchange. Gene flow occurs through the movement of seeds and/or pollen, and the variety of dispersal mechanisms observed in plants effects a diverse pattern of genetic exchange. Dispersal patterns are further influenced by spatial and temporal variation in systems of mating (Hamrick 1987; Hamrick & Godt 1990). For example, variation in rates of outcrossing vs. selfing may be dependent on geographical or seasonal variation in the distribution of pollinators, leading to different patterns of genetic exchange between populations within a species over time. Thus, patterns of genetic exchange in any plant system are influenced by a complex interaction of breeding system and patterns of dispersal.

Patterns of historical relationship

Genetic variation is structured not only by the contemporary forces of genetic exchange but also by historical patterns of relationship. For a given level of current genetic exchange, populations having recent common ancestry will be genetically more similar than those having more distant common ancestry. Ultimately, if genetic exchange between two populations or species ceases altogether, then shared common ancestry will be the sole determinant of genetic similarity between them. Therefore, historical relationships will contribute in some measure to the genetic structure of all plant species.

Traditional approaches to the study of genetic structure

Genetic differentiation within species

Within species, genetic exchange rather than historical relationship has traditionally been emphasized as the determinant of genetic structure. Classical models for describing this structure (e.g. *F* statistics, Wright 1951) do not distinguish historical effects from recurrent processes. Estimates of gene flow (*Nm*) derived from these models assume that current population structure reflects an equilibrium between genetic drift and gene flow. Such models serve well to document population genetic structure and its underlying causes for some plant species, specifically those with a stable population structure. However, in many groups genetic exchange across the species range is

severely restricted, either by the wide geographical distribution of populations or by limited pollen and seed dispersal. In these cases historical events such as range expansion, range fragmentation, and population bottlenecks will be strong determinants of population genetic structure. The observed genetic similarity between such populations owes more to recent common ancestry than to any ongoing process of genetic exchange. Thus, the underlying sources of genetic cohesion in many instances cannot be accurately portrayed with traditional population genetic models.

Genetic differentiation among species

Above the species boundary, historical relationship has been emphasized as the sole determinant of genetic similarity. In fact, the cladistic methodology used in phylogenetic inference assumes that taxonomic units represent nonreticulating lineages. Hennig (1966) was careful to draw the distinction between the reticulate genealogical or 'tokogenetic' relationships that characterize genetic exchange among individuals or populations, and the ordered hierarchical phylogenetic relationships characterizing divergent species or higher taxa. Plant species, however, are singular in their ability to repeatedly violate the assumption of nonreticulating lineages. Hybridization involving two (or more) ancestral species has long been recognized as an important mechanism of speciation in plants (Stebbins 1950; Grant 1981), and the hybrid origin of several species has been established empirically (e.g. Wolfe & Elisens 1994; Rieseberg *et al.* 1995). Similarly, current reproductive barriers vary from being incomplete to nonexistent among many well-established plant species, and hybrid zones and regions of introgression are frequently encountered. The widespread occurrence of chloroplast 'capture' illustrates the ease with which plant interspecific hybridization and introgression may occur (see below). Thus, just as the assumption of equilibrium in traditional population genetic models is often not appropriate, the assumption that interspecific relationships are nonreticulating is also often inappropriate in plants.

The phylogeographic approach

Because cladistic inference of phylogenetic relationships requires that genetic variants do not form reticulating lineages, phylogenetic approaches cannot be directly applied at the level of either the individual or the population. With few exceptions, both of these levels of biological organization are characterized by reticulating patterns of genetic exchange (i.e. sexual recombination and gene flow). However, if the genetic variants being considered are at the level of the gene instead of the individual, non-

recombining segments of DNA can be organized into hierarchically ordered networks of descent, and can provide historical information that individuals cannot. Gene genealogies can therefore form the basis of historical approaches to the study of intraspecific processes. Moreover, such cladistic gene genealogy methods can be applied to the study of reticulate evolution above the species level. This analysis of the spatial distribution of gene genealogies forms the basis of phylogeography.

While phylogeography uses the historical information inherent in gene trees, it is not merely an extension of phylogenetic principles to the intraspecific level. Rather, it characterizes population subdivision by recognizing geographical patterns of genealogical structure across the range of a species (Avice 1994). By synthesizing the influence of both history and current genetic exchange, phylogeography can potentially surmount the limitations inherent to classical population genetics and systematics. The analysis and interpretation of allele genealogies to infer population genetic processes has been further promoted by significant advances in coalescent theory (see reviews by Tavaré 1984; Ewens 1990; Hudson 1990). Recent theoretical and empirical work on coalescence has been applied to population structure (Avice *et al.* 1988; Slatkin 1989, 1991; Slatkin & Maddison 1989, 1990; Templeton *et al.* 1995 and references therein) and promises to open new avenues of investigation.

Phylogeographic analysis of plant population genetic structure

Elucidating the factors that determine genetic structure has been of longstanding interest to plant population geneticists. The widespread acceptance of the biological species concept in the past led to the commonly held ideal that species were defined by their mating relationships and that gene flow maintains the genetic cohesion of species. Ehrlich & Raven (1969), however, questioned the notion of widespread gene migration and suggested that shared common ancestry and similar selective regimes could account for the observed genetic cohesion of plants.

Empirical evidence on the extent and influence of gene flow in plants has been inconclusive. Direct estimates of pollen and seed dispersal in numerous taxa indicate that the vast majority of such movement is locally restricted and reveal that plant populations are far from panmictic (Levin & Kerster 1974; Schaal 1980). However, direct estimates may critically underestimate the importance of rare, long-distance gene flow in species cohesion, as these studies are necessarily limited in both time and space. Studies of plant paternity have suggested that long-distance gene dispersal between populations is not as infrequent as previously thought (Ellstrand *et al.* 1989; Dow & Ashley 1996; Dawson *et al.* 1997). Even if rare, such gene-

flow events may potentially have great evolutionary significance. In addition to direct estimates, gene flow has also been estimated indirectly from studies of genetic differentiation. Allozyme studies often reveal little population divergence over a wide geographical area (Hamrick 1987). In the past it was common to attribute this pattern of geographical variation to the homogenizing effects of recurrent gene flow, and historical explanations for the observed patterns (e.g. persistent ancestral polymorphisms) were typically not considered. Because of these incomplete and often contradictory estimates, the issue of the intensity and importance of gene flow among plant populations remains unresolved.

Phylogeography affords new insight into the role of gene flow in structuring plant populations. Specifically, it provides a way of detecting historic and recurrent gene-flow events and can potentially discriminate between gene flow and patterns caused by ancestral polymorphism. Phylogeographic analysis relies on interpreting patterns of congruence or lack of congruence between the geographical distribution of haplotypes and their genealogical relationships (Fig. 1). A pattern of congruence is seen if clades of closely related haplotypes are geographically restricted and occur in proximity to each other (Fig. 1b). Such congruence indicates a longstanding pattern of highly restricted gene flow. Assuming that populations are not panmictic, this pattern arises because novel mutations remain localized within the geographical context of their origins (Neigel *et al.* 1991; Slatkin 1991, 1993; Templeton *et al.* 1995). As a corollary, the most ancient haplotypes should be located at the centre of the gene tree and be geographically widespread, whereas the most recent haplotypes should be at the tips of the gene tree and be localized geographically (Golding 1987; Castelletto & Templeton 1994; Templeton *et al.* 1995). This basic pattern of congruence will be violated if polymorphisms persist and are differentially sorted following the divergence of populations (Fig. 1c), or if interpopulation gene flow occurs (Fig. 1d).

Among the limited number of studies in plants that might be considered phylogeographic, so far there has been little evidence for long-distance gene flow. For example, in our own laboratory, Matos (1992) examined gene flow among geographically isolated mountain populations of *Pinus hartwegii* occurring in Mexico. Using a gene tree based on chloroplast DNA (cpDNA) restriction site analysis, she found that the most closely related haplotypes were restricted to a given mountain top, indicating that these populations have been genetically isolated for a long time, despite the potential for long-distance gene flow (this example is discussed in detail in Templeton 1994). Similar geographical structuring of cpDNA haplotype diversity has also been revealed in other species, such as wild yam (*Dioscorea bulbifera*;

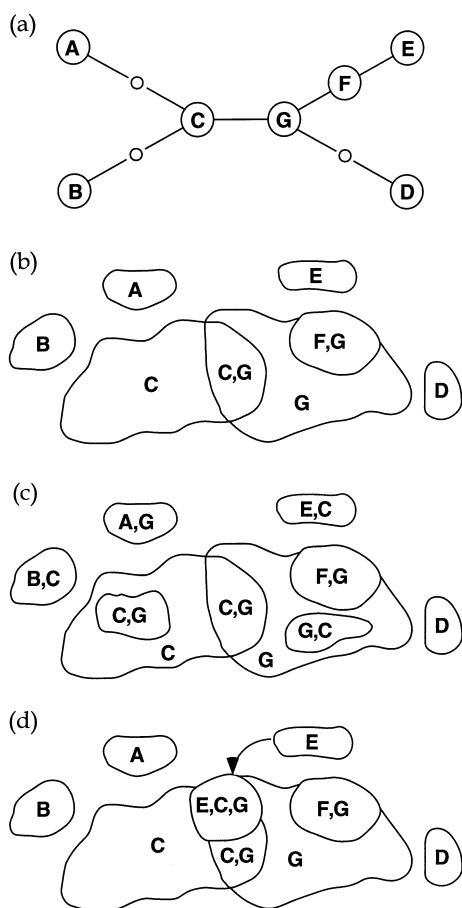


Fig. 1 Hypothetical single-mutational step tree of haplotypes A–G and example of geographic distributions expected to result from different phylogeographic histories. (a) One-step gene tree of haplotypes A–G. Each branch represents one mutational step. Internal nodes (inferred or occupied by extant haplotypes) are considered ancestral to terminal nodes (tips). (b) Congruence of geography and gene tree as expected to result from highly restricted gene flow. Ancient haplotypes have more widespread distributions. (c) Incongruence of geography and gene tree resulting from persistence and differential sorting of ancestral polymorphism (haplotypes C and G) among descendant populations. (d) Incongruence of geography and gene tree resulting from interpopulational gene flow. Where such gene flow involves ‘tip’ haplotypes (as illustrated here), it may be distinguished from the incongruence illustrated in (c).

Terauchi *et al.* 1991), closed-cone pines (Hong *et al.* 1993), and *Streptanthus glandulosus* (Mayer & Soltis 1994). While these results appear to support the conclusion of restricted gene flow from seed and pollen dispersal studies, the resolution of cpDNA variation in many cases has not been sufficient to assess more-local patterns of gene flow within species.

In some studies, the pattern of haplotype distribution reveals that a sizeable portion of a species range contains

little or no genetic variation relative to the rest of the species range. Such a pattern is consistent with a rapid range expansion in that region. Because of the profound effect of glaciation on plant distributions, this pattern might be expected in many native temperate and tropical species. The widespread distribution of individual haplotypes is clearly concordant with rapid range expansion from known refugia following Pleistocene glaciation in both Europe (e.g. *Fagus sylvatica* (Demesure *et al.* 1996), *Quercus* (Ferris *et al.* 1993)) and North America (e.g. north-western USA species in the *Heuchera* group of the Saxifragaceae (Soltis *et al.* 1989, 1991a, 1992a; reviewed in Soltis *et al.* 1997)). In other cases, wide geographical distributions of a single haplotype may reflect even more rapid historical dispersal by humans, as in the legume *Gliricidia* (Lavin *et al.* 1991) and wild yams (Terauchi *et al.* 1991).

In cases where incongruities exist between haplotype tree and geographical pattern, it is desirable to distinguish between gene flow and the sorting of persistent ancestral polymorphism as alternative causes. One potential way to accomplish this is to examine the position of the geographically incongruent haplotypes within the gene tree. Haplotypes that are ancient will represent internal nodes of the gene tree. If these haplotypes predate the divergence of populations, then we may infer that the distribution of these haplotypes owes more to the persistence of ancestral polymorphisms than to recent gene flow (Fig. 1c). Alternatively, if the geographically incongruent haplotypes are derived (represented as ‘tips’ in a gene tree), gene flow is the more likely explanation for their distribution (Fig. 1d) (Templeton 1994). This method for discriminating between the two alternatives has proved useful in several phylogeographic studies of animals (Templeton 1993, 1994; Templeton *et al.* 1995).

Patterns of incongruence between gene trees and their geographical distributions have been observed in several plant studies. In *Eucalyptus nitens*, 13 chloroplast haplotypes formed two clades that were not segregated by geography (Byrne & Moran 1994). A similar pattern was observed for *Phacelia dubia*, where one haplotype occurred in combination with other haplotypes in two disjunct regions (Levy *et al.* 1996). In this latter case, the haplotype responsible for the incongruence is a ‘tip’ in the gene tree which, by the methods described above, would implicate gene flow as the more plausible explanation for its distribution. However, the authors favour ancestral polymorphism based upon the geographical distribution of the haplotypes and improbability of gene migration between the populations in question. In such cases, distinguishing between these hypotheses is difficult. Ambiguity results because these gene trees are derived from loci with relatively low mutation rates (discussed below), so that the distinction between ‘tip’ and ‘internal’ nodes is unclear. In

addition, even a well-resolved phylogeny provides but a single assessment of genetic structure, which may or may not be congruent with gene trees based on other loci (see below).

Genetic vs. morphological/taxonomic differentiation in plants

A second issue relevant to the discussion of phylogeography in plants is the relationship between genetic and morphological differentiation. Although taxonomy has traditionally been based on morphology, botanists have long recognized that significant morphological variation can be accounted for by a relatively small number of genes (reviewed by Hilu 1983; Gottlieb 1984). This observation is supported by more recent molecular genetic studies examining the basis of morphology and development. For example, large phenotypic differences, such as determinate vs. indeterminate flowering (Bradley *et al.* 1997) or flower colour and shape (Bradshaw *et al.* 1995), may be due to variation in few genes.

This genetic architecture has important implications for evolutionary studies in plants. Because morphological divergence may have little relationship to the degree of genetic differentiation between lineages, it can be difficult to predict the genetic cohesiveness of a group based on its morphological differentiation or taxonomic status alone. The phenomenon of adaptive radiation exemplifies this pattern well. Groups such as the Hawaiian silversword alliance (*Argyroxiphium* spp.) (Carr 1987; Baldwin *et al.* 1990) and the *Espeletia* radiation of northern South America (Cuatrecasas 1986; Monasterio & Sarmiento 1991) demonstrate unusually rapid speciation and morphological evolution given their suspected time of origin, and yet may show little genetic divergence between species (e.g. Witter & Carr 1988) and few reproductive barriers to hybridization (Carr & Kyhos 1981, 1986; Berry & Calvo 1994). For example, in the case of the *Espeletia* radiation, species vary from tiny herbaceous rosettes to branched trees over 20 m tall, yet hybrids are known between many of these morphologically divergent species (Berry & Calvo 1994).

As a result of this lack of genetic differentiation and the subsequent potential for interspecific hybridization, genetic exchange may be occurring between what appear to be independent lineages. Introgression has been observed in many groups (e.g. maize (Doebly 1989) and irises (Anderson 1949; Arnold *et al.* 1992); reviewed by Rieseberg and Wendell 1993). It has been most clearly documented in the phenomenon known as 'chloroplast capture,' whereby the chloroplast genome of one species introgresses into populations of another following interspecific hybridization; plants are morphologically identifiable to one species but have the chloroplast genome

of another species. Several outstanding examples of chloroplast capture have been documented, including sunflowers (Rieseberg *et al.* 1990), poplars (Smith & Sytsma 1990), species in the *Heuchera* group of the Saxifragaceae (Soltis *et al.* 1991b), and oaks (Whittemore & Schaal 1991). In the study by Whittemore and Schaal, chloroplast capture was so common in eastern USA species of oaks that the geographical location of a given plant was a better predictor of chloroplast haplotype than was taxonomic designation.

Phylogeography, when applied across a broad taxonomic range, provides a theoretical framework with which one can test ideas of genetic isolation without being restricted to taxonomic preconceptions based on morphological divergence. Several studies have used phylogeography to lend insight into the relationship between phenotypic differentiation and genetic isolation. In a study of *Dioscorea bulbifera*, chloroplast haplotypes were much more strongly correlated with geographical position than with morphologically based subspecies designation (Terauchi *et al.* 1991). Other studies in species such as *Streptanthus glandulosus* (Mayer & Soltis 1994) have shown similar patterns at the varietal or subspecific level.

At the interspecific level, an example of the use of a phylogeographic approach is a study by Matos (1992; see Templeton 1994) of hybridization and introgression in species of the *Pinus montezumae* complex of Mexico. In this case, several chloroplast haplotypes were found to be shared between species in sympatric populations. A reconstruction of the gene tree for the chloroplast haplotypes indicated that the shared haplotypes were not internal in the gene tree (Templeton 1994). Based on coalescent theory, one can infer that these shared haplotypes are not the most ancestral haplotypes within the populations. Rather, shared haplotypes tended to be near the tips of the gene tree, suggesting that introgression was the most likely explanation for shared polymorphisms. Such phylogeographic studies hold great promise for understanding hybridization and introgression and their role in plant evolution.

As the above examples indicate, the point at which plants establish isolated lineages can occur across a wide range of evolutionary divergence. Isolation is found well below the species level in some species. On the other hand, genetic exchange can also occur between well-established, morphologically distinct species. This variation in the tempo and mode of evolution in plants has hindered the study of plant genetic structure. Since phylogeography is applicable both above and below the species boundary, it offers for the first time a single method of analysis that will cover the range of plant evolutionary patterns and not be confounded by the taxonomic, spatial and temporal variability of plant populations.

The search for variable loci for phylogeographic analysis

It is clear that gene genealogies offer great promise for furthering our understanding of plant evolution. In order to construct gene trees, however, significant genetic variation must occur at the appropriate level (i.e. among the populations or taxonomic units being investigated). The detection of phylogenetically informative intraspecific variation is probably the single most difficult problem facing plant population biologists interested in using the techniques of phylogeography. While phylogeographic studies in animal systems rely heavily on the mitochondrial genome (mtDNA), plant mtDNA genes exhibit low rates of nucleotide substitution, such that specific loci do not contain adequate variation for generating intraspecific gene phylogenies. Moreover, plant mtDNAs are prone to extensive intramolecular recombination, causing heteroplasmy and hindering investigation of variation at the whole genome level (reviewed by Palmer 1992). We must therefore look towards alternative genomes in the search for informative variation.

Chloroplast genome

Although levels of genetic variation in plant cpDNA are lower than in animal mtDNA, intraspecific variation has been reported in a growing number of species (reviewed by Soltis *et al.* 1992b; see also Levy *et al.* 1996 and references therein). Virtually all published plant phylogeography studies have relied on the chloroplast genome as their sole source of genetic variation. Similar to mtDNA, the chloroplast genome may be considered a single, non-recombining unit of inheritance. The mutation rate of cpDNA varies for different regions of the genome, with most variation apparently occurring within the large single-copy regions and not in the inverted repeats. Taxa lacking the inverted repeats, such as gymnosperms, appear to be particularly variable (e.g. Wagner *et al.* 1991).

In species where sufficient genetic variation has been detected to permit phylogeographic analysis, this variation has been revealed by restriction enzyme digests of cpDNA. Most studies have used traditional restriction fragment analysis of the entire chloroplast genome, in which genetic variants reflect the gain or loss of restriction sites or length variation (e.g. Soltis *et al.* 1989; Lavin *et al.* 1991; Matos 1992; Byrne & Moran 1994; Mayer & Soltis 1994; Levy *et al.* 1996; Terauchi *et al.* 1991; Van Dijk & Bakx-Schotman 1997).

A more recent restriction enzyme-based approach involves digestion of particular PCR-amplified chloroplast loci to reveal fragment length polymorphisms within the amplified fragment (e.g. Demesure *et al.* 1996; El Mousadik & Petit 1996). Large portions of the chloro-

plast genome may be evaluated in numerous individuals using these readily accessible laboratory techniques. Furthermore, as at least 50% of all cpDNA variation may be attributable to small insertion/deletion mutations (Gaut *et al.* 1993; Gielly & Taberlet 1994), this method is more likely than restriction site analysis to reveal variation. However, concerns about the homology of length variants associated with simple-sequence repeat polymorphisms need to be addressed before this technique can be widely applied to construct gene trees.

Ultimately, direct knowledge of the sequences underlying cpDNA restriction site variation would be most desirable for gene tree construction. However, unlike restriction enzyme analyses, direct sequencing of cpDNA loci so far has not yielded optimal levels of variation for phylogeographic analysis. Although several small regions of the chloroplast genome (such as some intergenic spacers) show potential for phylogeographic analysis (Taberlet *et al.* 1991; Demesure *et al.* 1995; Dumolin-Lapegue *et al.* 1997), attempts by our laboratory as well as others indicate that single cpDNA loci are only occasionally useful at the intraspecific level. With present techniques, it is difficult to sequence more than small portion of the chloroplast genome, which limits the probability of detecting variation. As technology progresses and the sequencing of larger fragments of DNA becomes feasible, it seems likely that studies will utilize more of the variation potentially available in the chloroplast, allowing finer phylogeographic resolution.

Nuclear genome

The remaining alternative is the nuclear genome, which is largely unexplored in plants but which offers a potentially inexhaustible source of phylogenetically informative genetic variation. Many investigators are developing techniques and strategies for locating and efficiently sampling appropriate variation in nuclear DNA (nDNA). The internal transcribed spacer (ITS) region of ribosomal DNA, useful for plant systematics, is generally not conducive to phylogeographic study. First, for most species examined, intraspecific variation has not been detected in this region. Furthermore, as part of a multicopy gene family, the ITS region is subject to the poorly understood processes of concerted evolution, confounding interpretation of sequence polymorphism at the intraspecific level. A more promising source of nDNA variation may be provided by single-copy (or low-copy number) nuclear genes (Palumbi & Baker 1994; Slade *et al.* 1993, 1994; Strand *et al.* 1997). Nevertheless, at this early stage no single locus appears to be universally useful in all species of plants.

Use of nDNA for phylogeographic analyses requires that additional features of this genome be taken into

account. As the nuclear genome is diploid, complications involving interallelic recombination and heterozygosity can arise. Recombinant alleles arise from crossover events among alleles of a locus, resulting in chimeric haplotypes. If relatively few in number, these can be detected by their lack of phylogenetic congruence with other haplotypes. Several algorithms have been developed for their detection during cladogram construction (Templeton *et al.* 1992; Hein 1993). For heterozygous individuals, the two alternative alleles (haplotypes) must be analysed individually as operational taxonomic units (OTUs). In cases where the two alleles differ appreciably in size, they may be separated by gel electrophoresis prior to sequencing. Alleles of similar size can be isolated either by separation by chemical properties (Lessa 1992; Lessa & Applebaum 1993) or by cloning. Alternatively, if identical in size (i.e. no indels), the two alleles may be sequenced together and double bands at variable sites assigned to haplotypes previously identified from homozygous individuals (Clark 1990; Slade *et al.* 1993).

Another concern when using the nuclear genome involves the homology of loci. Some (and perhaps many) 'single-copy' nuclear genes exist as part of small gene families consisting of two to 10 expressed loci and possibly additional pseudogenes (Palumbi 1996). When a locus is part of a multicopy gene or multigene family, PCR amplification with conserved primers may produce a suite of fragments, including duplicated gene copies, pseudogenes, and even recombinant PCR artifacts (Palumbi 1996; Hillis *et al.* 1996). Care must therefore be taken to avoid comparing paralogous loci, which may be especially difficult to detect in cases where there has been differential homogenization of gene copies among populations (a phenomenon analogous to lineage sorting of alleles). If the different gene copies amplified by a set of primers are not consistently distinguishable and separable by size, the safest way to assure orthologous comparisons is to design locus-specific primers (Hillis *et al.* 1996). Despite these potential problems the nuclear genome represents perhaps the most fruitful avenue for study of plant phylogeography.

Inferences from loci differing in modes of inheritance

Ultimately, the potential of phylogeography will be fully realized when multiple loci are considered. As has been shown by both theoretical and empirical work, independent gene trees often do not reveal congruent phylogeographic patterns because polymorphisms present in independent loci will be sorted differently. Therefore, it would be ideal to consider the multiple gene trees which may be provided by both the chloroplast and multiple loci within the nuclear genome. In order to interpret comparisons between the chloroplast and nuclear loci, certain

characteristics that will influence the structure of their gene trees must be taken into account.

The mode of inheritance of the genome has important implications for the structure of gene trees. In most plants, a migrant seed carries with it complete chloroplast and nuclear genomes; in contrast, pollen carries only half a nuclear genome, and it may or may not transmit a chloroplast genome, depending on whether the species shows maternal, paternal, or biparental cpDNA transmission. (As reviewed by Harris & Ingram (1991), there are many documented exceptions to the rule that angiosperms show maternal chloroplast inheritance. In conifers, inheritance of cpDNA is typically paternal.) If the mode of inheritance is known, comparison of nuclear and chloroplast data can potentially reveal insights into the relative roles of pollen vs. seed movement as sources of gene flow (Ennos 1994; McCauley 1994; El Mousadik & Petit 1996). Strictly maternal inheritance of the chloroplast genome in many angiosperms limits chloroplast dispersal to seeds. When seed dispersal is limited, chloroplast differentiation will exclusively reflect this limited dispersal mechanism. Examples may be most evident in wind-pollinated trees with large seeds. Alternatively, seeds may be very well adapted to long-distance dispersal while pollen dispersal is quite limited.

Additional factors to consider are the difference in effective population size between the chloroplast and nuclear genomes and how such a difference is influenced by breeding system. As in the mitochondrion, the chloroplast genome is effectively haploid whereas nuclear genomes are diploid or polyploid. Thus, in a monoecious species, the effective population size of the nuclear genome is at least twice that of the chloroplast genome. Because the effects of genetic drift are greater as effective population size decreases, genetic drift will be a stronger force in shaping the population structure of the chloroplast genome than for nuclear genes. As a consequence, a cpDNA gene tree is more likely to reveal patterns of population differentiation and less likely to reflect ancestral polymorphisms than is a nuclear gene tree. If a species is dioecious, only half the population transmits a chloroplast genome, resulting in an effective population size of roughly one quarter that of the nuclear genome (assuming equal sex ratios in a population and uniparental cpDNA inheritance). Therefore, in a dioecious species, the differential effects of genetic drift on the chloroplast genome will be intensified.

We are not aware of any published intraspecific studies that have examined both cpDNA and nDNA phylogenies. However, a number of studies have compared cpDNA phylogenies with patterns of genetic differentiation inferred from allozyme allele frequency data (e.g. Soltis *et al.* 1997). Such studies typically find more evidence for population differentiation with chloroplast gene tree data than for nuclear genes. This pattern is consistent with

expectations, regardless of whether the lack of differentiation in the nuclear genes is the result of ancestral polymorphism maintained by the larger effective population size or higher rates of dispersal of nuclear genes. In order to understand the genetic structure of plant populations, comparative studies making use of multiple nDNA phylogenies, as well as phylogenies from cpDNA, are needed.

Conclusion

The application of phylogeography for studying plant evolution is still in its infancy. The few studies conducted to date demonstrate how this approach may be used to address unresolved issues concerning genetic exchange and differentiation within and among plant species. In the future, as technical advances facilitate the detection and measurement of DNA variation, we can expect more studies of diverse plant taxa to be initiated and ultimately to provide a more comprehensive picture of plant evolution.

References

- Anderson E (1949) *Introgressive Hybridization*. Wiley, New York.
- Arnold ML, Robinson JJ, Buckner CM, Bennett BD (1992) Pollen dispersal and interspecific gene flow in Louisiana irises. *Heredity*, **68**, 399–404.
- Avise JC (1994) *Molecular Markers, Natural History And Evolution*. Chapman and Hall, NY.
- Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.
- Baldwin BG, Kyhos DW, Dvorak J (1990) Chloroplast DNA evolution and adaptive radiation in the Hawaiian silversword alliance (Asteraceae–Madiinae). *Annals of the Missouri Botanical Garden*, **77**, 96–109.
- Berry PE, Calvo RN (1994) An overview of the reproductive biology of Espeletia (Asteraceae) in the Venezuelan Andes. In *Tropical Alpine Environments* (eds Rundel PW, Smith AP, Meinzer FC), pp. 229–249. Cambridge University Press, Great Britain.
- Bradley D, Ratcliffe O, Vincent C, Carpenter R, Coen E (1997) Inflorescence commitment and architecture in Arabidopsis. *Science*, **275**, 80–82.
- Bradshaw HD, Wilbert SM, Otto KG, Schemske DW (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature*, **376**, 762–765.
- Byrne M, Moran GF (1994) Population divergence in the chloroplast genome of *Eucalyptus nitens*. *Heredity*, **73**, 18–28.
- Carr GD (1987) Beggar's ticks and tarweeds: masters of adaptive radiation. *Trends in Ecology and Evolution*, **2**, 192–195.
- Carr GD, Kyhos DW (1981) Adaptive radiation in the Hawaiian silversword alliance (Compositae–Madiinae). I. Cytogenetics of spontaneous hybrids. *Evolution*, **35**, 543–556.
- Carr GD, Kyhos DW (1986) Adaptive radiation of the Hawaiian silversword alliance (Compositae–Madiinae). II. Cytogenetics of artificial and natural hybrids. *Evolution*, **40**, 959–976.
- Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution*, **3**, 102–113.
- Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution*, **7**, 111–122.
- Cuatrecasas J (1986) Speciation and radiation of the Espeletiinae in the Andes. In: *High Altitude Tropical Biogeography* (eds Vuilleumier F, Monasterio M), pp. 267–303. Oxford University Press, New York.
- Dawson IK, Waugh R, Simons AJ, Powell W (1997) Simple sequence repeats provide a direct estimate of pollen-mediated gene dispersal in the tropical tree *Gliricidia sepium*. *Molecular Ecology*, **6**, 179–183.
- Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515–2520.
- Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, **4**, 129–131.
- Doebly JF (1989) Molecular evidence for a missing wild relative of maize and the introgression of its chloroplast genome into *Zea perennis*. *Evolution*, **43**, 1555–1558.
- Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Molecular Ecology*, **5**, 615–627.
- Dumolin-Lapegue S, Pemonge M-H, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, **6**, 393–397.
- Ehrlich PR, Raven PH (1969) Differentiation of populations. *Science*, **165**, 1228–1232.
- El Mousadik A, Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology*, **5**, 547–555.
- Ellstrand NC, Devlin B, Marshall DL (1989) Gene flow by pollen into small populations: data from experimental and natural stands of wild radish. *Proceedings of the National Academy of Sciences of the USA*, **86**, 9044–9047.
- Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.
- Ewens WJ (1990) Population genetics theory – the past and the future. In: *Mathematical and Statistical Developments of Evolutionary Theory* (ed. Lessard S), pp. 177–227. Kluwer Press, New York.
- Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology*, **2**, 337–344.
- Gaut BS, Muse SV, Clegg MT (1993) Relative rates of nucleotide substitution in the chloroplast genome. *Molecular Phylogenetics and Evolution*, **2**, 89–96.
- Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: non-coding vs. rbcL sequences. *Molecular Biology and Evolution*, **11**, 769–777.
- Golding GB (1987) The detection of deleterious selection using ancestors inferred from a phylogenetic history. *Genetics Research*, **49**, 71–82.
- Gottlieb LD (1984) Genetics and morphological evolution in plants. *The American Naturalist*, **123**, 681–709.
- Grant V (1981) *Plant Speciation*. Columbia University Press, New York.
- Hamrick JL (1987) Gene flow and distribution of genetic variation

- in plant populations. In: *Differentiation Patterns in Higher Plants* (ed. Urbanska KM), pp. 53–67. Academic Press, London.
- Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates Inc., Sunderland, MA.
- Harris SA, Ingram R (1991) Chloroplast DNA and biosystematics: The effects of intraspecific diversity and plasmid transmission. *Taxon*, **40**, 393–412.
- Hein J (1993) A heuristic method to reconstruct the history of sequences subject to recombination. *Journal of Molecular Evolution*, **36**, 396–405.
- Hennig W (1966) *Phylogenetic Systematics*. University of Illinois Press, Urbana.
- Hillis DM, Maple BK, Larson A, Davis SK, Zimmer EA (1996) Nucleic acids IV: sequencing and cloning. In: *Molecular Systematics*, 2nd edn (eds Hillis DM, Moritz C, Maple BK), pp. 321–381. Sinauer Associates, Sutherland, MA.
- Hilu KW (1983) The role of single gene mutations in the evolution of flowering plants. In: *Evolutionary Biology* (eds Hecht MK, Wallace B, Prance GT), pp. 97–128. Plenum Press, New York.
- Hong YP, Hipkins VD, Strauss SH (1993) Chloroplast DNA diversity among trees, populations and species in the California closed-cone pines (*Pinus radiata*, *Pinus muricata* and *Pinus attenuata*). *Genetics*, **135**, 1187–1196.
- Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Surveys in Evolutionary Biology* (eds Futuyma D, Antonovics J), pp. 1–44. Oxford University Press, New York.
- Lavin M, Matthews S, Hughes C (1991) Chloroplast DNA variation in *Gliricidia sepium* (Leguminosae): Intraspecific phylogeny and tokogeny. *American Journal of Botany*, **78**, 1576–1585.
- Lessa EP (1992) Rapid surveying of DNA sequence variation in natural populations. *Molecular Biology and Evolution*, **9**, 323–330.
- Lessa EP, Applebaum G (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**, 119–129.
- Levin DA, Kerster HW (1974) Gene flow in seed plants. In: *Evolutionary Biology* (ed. Dobzhansky T), pp. 139–220. Plenum, New York.
- Levy F, Antonovics J, Boyton JE, Gilham NW (1996) A population genetic analysis of chloroplast DNA in *Phacelia*. *Heredity*, **76**, 143–155.
- Matos J (1992) *Evolution within the Pinus montezumae complex of Mexico: population subdivision, hybridization and taxonomy*. PhD thesis, Washington University, St. Louis, Missouri.
- Mayer MS, Soltis PS (1994) The evolution of serpentine endemics: a chloroplast DNA phylogeny of the *Streptanthus glandulosus* complex (Cruciferae). *Systematic Botany*, **19**, 557–574.
- McCaughey DE (1994) Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the USA*, **91**, 8127–8131.
- Monasterio M, Sarmiento L (1991) Adaptive radiation of *Espeletia* in the cold Andean tropics. *Trends in Ecology and Evolution*, **6**, 387–391.
- Neigel JE, Ball RM, Avise JC (1991) Estimation of single generation migration distances from geographic variation of animal mitochondrial DNA. *Evolution*, **45**, 423–432.
- Palmer JD (1992) Mitochondrial DNA in plant systematics: applications and limitations. In: *Molecular Systematics of Plants* (eds Soltis PS, Soltis DE, Doyle JJ), pp. 36–49. Chapman and Hall, New York.
- Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: *Molecular Systematics*, 2nd edn (eds Hillis DM, Moritz C, Maple BK), pp. 205–247. Sinauer Associates, Sutherland, MA.
- Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA from humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.
- Rieseberg LH, Beckstrom-Sternberg S, Doan K (1990) *Helianthus annuus* ssp. *texasus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences of the USA*, **87**, 593–597.
- Rieseberg LH, Fossen CV, Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*, **375**, 313–316.
- Rieseberg LH, Wendell JF (1993) Introgression and its consequences in plants. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 70–109. Oxford University Press, Oxford.
- Schaal BA (1980) Measurement of gene flow in *Lupinus texensis*. *Nature*, **284**, 450–453.
- Slade RW, Moritz C, Heideman A (1994) Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Molecular Biology and Evolution*, **11**, 341–356.
- Slade SW, Moritz C, Heideman A, Hale PT (1993) Rapid assessment of single-copy nuclear DNA variation in diverse species. *Molecular Ecology*, **2**, 359–373.
- Slatkin M (1989) Population structure and the evolutionary process. *Genome*, **31**, 196–202.
- Slatkin M (1991) Inbreeding coefficients and coalescent times. *Genetical Research*, **58**, 167–175.
- Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, **47**, 264–279.
- Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–613.
- Slatkin M, Maddison WP (1990) Detecting isolation by distance using phylogenies of genes. *Genetics*, **126**, 249–260.
- Smith SL, Sytsma KJ (1990) Evolution of *Populus nigra* (sect. *Aigeiros*): introgressive hybridization and the chloroplast contribution of *Populus alba* (sect. *Populus*). *American Journal of Botany*, **77**, 1176–1187.
- Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**, 353–373.
- Soltis DE, Mayer MS, Soltis PS, Edgerton M (1991a) Chloroplast DNA variation in *Tellima grandiflora* (Saxifragaceae). *Plant Systematics and Evolution*, **181**, 203–216.
- Soltis DE, Soltis PS, Collier TG, Edgerton ML (1991b) Chloroplast DNA variation within and among genera of the Heuchera group (Saxifragaceae): evidence for chloroplast transfer and parafly. *American Journal of Botany*, **78**, 1091–1112.
- Soltis DE, Soltis PS, Kuzoff RK, Tucker TL (1992a) Geographic structuring of chloroplast DNA genotypes in *Tiarella trifoliata* (Saxifragaceae). *Plant Systematics and Evolution*, **181**, 203–216.

- Soltis DE, Soltis PS, Milligan BG (1992b) Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. In: *Molecular Systematics of Plants* (eds Soltis DE, Soltis PS, Doyle JJ), pp. 117–150. Chapman and Hall, New York.
- Soltis DE, Soltis PS, Ranker TA, Ness BD (1989) Chloroplast DNA variation in a wild plant, *Tolmiea menziesii*. *Genetics*, **121**, 819–826.
- Stebbins GL (1950) *Variation and Evolution in Plants*. Columbia University Press, New York.
- Strand AE, Leebens-Mack J, Milligan BG (1997) Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology*, **6**, 113–118.
- Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.
- Tavaré S (1984) Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical Population Biology*, **26**, 119–164.
- Templeton AR (1993) The 'Eve' hypothesis: a genetic critique and reanalysis. *American Anthropologist*, **95**, 51–72.
- Templeton AR (1994) The role of molecular genetics in speciation studies. In: *Molecular Ecology and Evolution: Approaches And Applications* (eds Schierwater B, Streit B, Wagner GP, DeSalle R), pp. 455–477. Birkhauser Verlag Basel, Switzerland.
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.
- Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA sequence data. III. Cladogram estimation. *Genetics*, **140**, 767–782.
- Terauchi R, Terachi T, Tsunewaki K (1991) Intraspecific variation of chloroplast DNA in *Dioscorea bulbifera* L. *Theoretical and Applied Genetics*, **81**, 461–470.
- Van Dijk P, Bakx-Schotman T (1997) Chloroplast DNA phylogeography and cytotypic geography in autopolyploid *Plantago media*. *Molecular Ecology*, **6**, 345–352.
- Wagner DB, Sun Z-X, Govindaraju DR, Dancik BP (1991) Spatial patterns of chloroplast DNA and cone morphology variation within populations of a *Pinus banksiana* – *Pinus contorta* sympatric region. *The American Naturalist*, **138**, 156–170.
- Whittemore AT, Schaal BA (1991) Interspecific gene flow in oaks. *Proceedings of the National Academy of Sciences of the USA*, **88**, 2540–2544.
- Witter MS, Carr GD (1988) Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution*, **42**, 1278–1287.
- Wolfe AD, Elisens WJ (1994) Nuclear ribosomal DNA restriction-site variation in *Penstemon* section *Peltanthera* (Scrophulariaceae): an evaluation of diploid hybrid speciation and evidence for introgression. *American Journal of Botany*, **81**, 1627–1635.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

This paper represents a collaborative effort of Barbara Schaal and several of her graduate students. Barbara Schaal is interested in the evolutionary biology of plants. Doug Hayworth is examining hierarchical levels of concerted evolution using nuclear rDNA intergenic spacers in species of *Arabidopsis*. Ken Olsen is investigating the phylogeography of the wild relatives of cassava (*Manihot esculenta*) in Brazil. Jason Rauscher is researching adaptive radiation and the role of hybridization in the evolution of the *Espeletia* complex of the northern Andes. Allen Smith studies the population genetics and molecular evolution of self-incompatibility in wild populations of *Lycopersicon chilense*.
