Plant Syst Evol (2009) 282:281–294
DOI 10.1007/s00606-008-0036-9

REVIEW

Application of non-coding DNA regions in intraspecific analyses

Thekla Pleines · Sabine S. Jakob · Frank R. Blattner

Received: 30 January 2008 / Accepted: 19 March 2008 / Published online: 1 August 2008
© Springer-Verlag 2008

Abstract In this review we discuss the use of non-coding DNA at the intraspecific level in plants. Both nuclear and organelle non-coding regions are widely used in interspecific phylogenetic approaches. However, they are also valuable in analyses on the intraspecific level. Besides taxonomy, that is, defining subspecies or varieties, large fields for the application of non-coding DNA are population genetic and phylogeographic studies. Population genetics tries to explain the genetic patterns within species mostly by the amount of extant gene flow among populations, while phylogeography explicitly tries to reconstruct historic events. Depending on the study different molecular markers can be used, varying between very fast evolving microsatellites or some more slowly changing regions like intergenic spacers and introns. Here, we focus mainly on the use of non-coding regions in phylogeographic analyses. Mostly used in this context are regions of the genomes of the chloroplasts and mitochondria. In phylogeography, the correct estimation of allele or haplotype relationships is particularly important. As tree-based methods are mostly insufficient to depict relationships within species, network approaches are better suitable to infer gene or locus genealogies. Problematic for phylogeographic studies are alleles shared among multiple species, which could result from either hybridization or incomplete lineage sorting. Especially the latter can severely influence the interpretation of the phylogeographic patterns. Therefore, it seems necessary for us to also include close relatives of the species under study in phylogeographic analyses. Not only the sample design but also the analysis methods are currently changing, as some new methods such as statistical phylogeography were emerging recently and widely used methods like nested clade analysis might not be reliable in every case. During the last few years, a multitude of studies were published, which mainly analyzed phylogeographic patterns in European and North American plants. Phylogeographic studies in other regions of the earth are still comparably rare, although questions like the influence of the ice age on the vegetation in the tropics or southern hemisphere are still open and phylogeography provides an excellent remedy to answer them.

Keywords Chloroplast DNA · Microsatellites (SSR) · Mitochondrial DNA · Population genetics · Phylogeography

T. Pleines · S. S. Jakob · F. R. Blattner (✉)
Leibniz-Institute of Plant Genetics and Crop Research (IPK), 06466 Gatersleben, Germany
e-mail: blattner@ipk-gatersleben.de

Introduction

DNA polymorphisms in non-coding regions are widely used for phylogenetic inferences of species relationships. In addition, some non-coding regions also exhibit enough variability for intraspecific studies, that is, to analyze phylogenetic relationships of subspecies, varieties, and domesticated forms or to analyze the structure of populations. A relatively old field of research using these polymorphisms is the classical population genetic approach (Wright 1951) that is still widely used in modern studies (Bachmann 2001). The main focus of this review, however, will be on the use of non-coding DNA in phylogeography. Phylogeography tries to infer population histories in space and time from the extant geographical distribution of genetic polymorphisms. During the last few years, phylogeography has been an important area of research (Avise 2000), and

since its foundation, the numbers of papers on this topic are steadily increasing, as can be seen in many recent issues of the major systematics and ecology journals.

As most non-coding parts of the plant genomes are free to vary without much restriction from selection (but see Halligan et al. 2003; Clark et al. 2006; Guo et al. 2007; Kelchner and Graham 2008) they can contain an ample amount of polymorphic sites (Bosch and Quandt 2008; Koch et al. 2008; Rein et al. 2008). However, this advantage also creates a major difficulty access these regions. Analyses today rely mainly on an initial PCR amplification step that requires prior knowledge of the flanking sequence regions where primers can be designed to bind. Therefore, the majority of non-coding regions used today in systematics and population biology are flanked by relatively conserved DNA regions (mostly genes), allowing the design of (nearly) universal PCR primers (White et al. 1990; Taberlet et al. 1991; Desmesure et al. 1995; Blattner 1999; Weising and Gardner 1999). Variable regions without conserved parts in a distance easily coverable by PCR are hard to access, as specific PCR primers have to be designed for each species or closely related species group under study. If no prior sequence information for a targeted region exists, this is tedious and often far beyond the technical capabilities of researchers working in organismic biology. However, current efforts to obtain sequences of the whole chloroplast genome for many plant groups will allow easier primer design for future studies, as chloroplast regions conserved in specific families or genera then can be simply found via nucleotide database searches. Moreover, non-coding DNA might contain small conserved regions involved in DNA transcription regulation (Clark et al. 2006), which can be used to bridge large parts of otherwise non-conserved DNA stretches via PCR or during sequencing.

Population genetics

Population genetics mainly studies contemporary gene flow and population structure. The basis for this field of research was laid by the work of Fisher (1930), Wright (1931, 1951), and Haldane (1932). Many factors influence the genetic structure of populations, as, for example, seed and pollen dispersal, breeding system, population size, genetic drift, selective pressure, and adaptations to habitats. The mating system is very important in this respect, since self-pollination leads to low gene flow between populations and, in the course of time, to a reduced number of polymorphisms in populations (Glémin et al. 2006), while outcrossing reduces genetic differentiation in and among populations. Furthermore, historical factors also, such as past genetic bottlenecks, contribute importantly to population structure, but are hard to access with these population genetic methods.

There are studies using a variety of genetic markers and dealing with a wide range of topics. After showing which markers are suitable for intraspecific studies, we will mention a few exemplary studies dealing with different topics associated with population or conservation genetics.

Genomes and markers

Generally, all DNA regions variable enough can be used for population genetic studies or intraspecific studies in general. Frequently, the organellar genomes are used, because chloroplasts and mitochondria are mostly uniparentally inherited in seed plants and thus have some great advantages over biparentally inherited nuclear markers. The main advantage is that there is typically only one allele per cell and organism, and consequently no recombination between two alleles occurs. Due to different dispersal distances, biparentally, maternally, and paternally inherited genomes also exhibit strong differences in genetic differentiation between populations. Especially, maternally inherited markers mostly show a much higher population subdivision (Petit et al. 2005).

The mitochondrial genome of plants is considerably larger than that of animals. Additionally, pronounced differences in size and organization of mitochondrial genomes exist among plant taxa. Intramolecular recombination, leading to complex genome rearrangements and, therefore, variable gene order even within single individuals, as well as duplications and deletions, are common (reviewed in Palmer 1992). Furthermore, base substitution rates in plant mitochondria are rather low (Wolfe et al. 1987), resulting in only minute differences within specific loci among individuals or even species.

Chloroplast genomes on the other hand exhibit a much more stable structure than those of mitochondria, and also higher substitution rates can be observed (Wolfe et al. 1987). However, chloroplast DNA (cpDNA) was for a long time considered too conserved for intraspecific studies (Banks and Birky 1985). This perspective changed at the beginning of the 1990s (Soltis et al. 1992), after a number of studies had found intraspecific or even intrapopulational chloroplast variation (Wagner et al. 1987; Milligan 1991). This intraspecific variation was shown to be high enough for population studies regarding gene flow (reviewed by McCauley 1995). An interesting approach is the contrast of paternally or biparentally inherited markers with maternally inherited markers. Using this combination, the ratio and the distances of pollen- vs. seed-based gene flow can be measured (Dong and Wagner 1994; Latta et al. 1998; Fénart et al. 2007).

In gymnosperms, the situation is somewhat different. Here, chloroplasts are inherited mainly paternally and are, therefore, dispersed through pollen and seed, while

mitochondria are mainly maternally inherited and thus dispersed via seeds only (Wagner 1992). Since pollen is normally distributed over far longer distances than seeds (Liepelt et al. 2002), mitochondrial markers exhibit a much stronger population differentiation than chloroplast markers and are important characters used for population genetic studies in gymnosperms (Johansen and Latta 2003), sometimes also used in connection with cpDNA markers (Chiang et al. 2006).

Chloroplast DNA regions that are often used in infraspecific studies encompass various intergenic spacers and introns, for which universal primers exist. A comparison of the variability at some non-coding cpDNA regions was reviewed by Shaw et al. (2005, 2007). Among the most frequently used cpDNA regions are the *trn*L intron and the *trn*L–*trn*F intergenic spacer, also often used in combination (Koch et al. 2006). Other variable chloroplast regions are the *atp*B–*rbc*L intergenic spacer (Hung et al. 2005; Chiang et al. 2006; Bänfer et al. 2006), and *trn*H–*psb*A (Xu et al. 2000).

Mitochondrial regions used in intraspecific studies of plants, especially gymnosperms, are, for example, the introns of the NADH dehydrogenase gene *nad*1 (Johansen and Latta 2003; Jaramillo-Correa et al. 2004; Chiang et al. 2006), *nad*7 intron 1 (Godbout et al. 2005), *nad*5 intron 4 (Liepelt et al. 2002) and the internal transcribed spacer (ITS) of mitochondrial ribosomal DNA (Huang et al. 2001).

Besides the organellar markers mentioned above, microsatellites (Tautz 1989; Tautz et al. 1986) or simple sequence repeats (SSR) are often used in population biology, and sometimes also for phylogeographic studies. Microsatellites are short tandem repeats of single bases (mononucleotide repeats; i.e., mostly runs of 8–20 T/A) or small DNA motives of up to five bases (in plants often GA/TC, GAA/TTC). Length variation at these loci originates mostly from slippage of the DNA strands during DNA replication, thus resulting in an increase or decrease of the number of repeat motifs. Uninterrupted long runs of SSR motifs (>10 repeats) normally show more length variation than shorter microsatellites or loci where a different base occur within the SSR motif, as the possibility of slippage increases with the number of consecutive repeated units. Microsatellites are much less common in plants than in animals (Lagercrantz et al. 1993). Nevertheless, they exist in the nuclear as well as in the organellar genomes. Microsatellites generally show high variability that can be useful in population genetic studies, when other sequences or fingerprint methods do not exhibit enough mutations (Tautz 1989; Powell et al. 1995). However, their high variability is also the major disadvantage of these markers, as microsatellite loci often show high levels of homoplasy, especially when distant populations or higher taxonomic levels are studied (Provan et al. 2001; Jakob et al. 2007).

Microsatellite primers designed for one species might as well amplify in closely related species (Guicking et al. 2006). However, for nuclear loci, no universal SSR primers exist, and even for the chloroplast genome universally variable loci are rare. Weising and Gardner (1999) published PCR primers for chloroplast SSR loci, which amplify in many plant families but do not always provide products with length variation. As repeat lengths might be quite different among different taxonomic groups, a locus selected for high repeat number in one species might be absent or much shorter and, therefore, less variable in most other taxa. This bias in selection procedure explains the often low variation at microsatellite loci when transferred to other taxa. Conversely, when using known sequences of taxa closely related to the target species to find SSR loci (e.g. using the wheat or rice chloroplast genome sequences to localize potentially variable loci in other grasses) it is often worth including loci with only small repetitive motifs in the screening process, as these might have evolved into longer and, therefore, more variable SSRs in the taxon under study.

Complete absence of variation at some chloroplast microsatellite loci was found, for example, by Provan et al. (1999) and Rendell and Ennos (2002). In such cases, it might be useful to re-sequence the locus to see if the absence of sequence length variation is due to shrinking or even complete loss of the microsatellite motif or if it is related to a population genetic bottleneck, erasing alleles of a potentially variable locus from the plant populations (Jakob et al. 2007). The latter should, however, influence variation at most loci in the genome, while loss of SSR motifs might be restricted to a single locus. Re-sequencing of different SSR alleles can also contribute to the understanding of the nature of length variation. As, not only the repeat number might vary, but also the motif itself, and insertions or deletions outside the microsatellite motif can change the length of the amplified fragment (Jakob et al. 2007), it seems advisable to generally check SSR loci, when comparisons among different species or even subspecies are conducted. A determination of the DNA sequence also seems necessary when two SSR loci occur in close proximity within an amplified locus. In this case, compensatory mutations, that is, the expansion of one repeat stretch is counterbalanced by shrinking of the second, can result in identical fragment length of different alleles. These complex SSRs should either be excluded from the analysis, or sequence identity has to be verified by re-sequencing, restriction digests, or single nucleotide sequencing (SNS) analysis (Guicking et al. 2008).

Frequently, fingerprint methods, such as amplified fragment length polymorphism (AFLP) (Vos et al. 1995) or random amplified polymorphic DNA (RAPD) (Welsh and McClelland 1990; Williams et al. 1990), were also used for

population studies (see Nybom 2004 for a comparison between different marker systems). However, AFLP, RAPD, and similar methods are anonymous markers, and even if the majority of polymorphic sites may be within non-coding regions, the exact location is unknown. Therefore, we will not go into detail about these markers.

Analysis methods

One of the main inferences made in population genetics is how genetic variation is distributed within and among natural populations of interbreeding organisms to study gene flow, genetic drift, mating systems, mutation rates, and natural selection (Templeton et al. 1995). This is often quantified by F statistics, using fixation indices like F_{ST} (Wright 1951) or related measures like G_{ST} (Nei 1973). From F_{ST} values, gene flow between populations can also be calculated (Beaumont 2005). As these are classical methods, they are included in many population genetic computer programs (Pearse and Crandall 2004). The main drawback of many traditional population genetic analysis methods is that the models used for the calculations of population genetic parameters are based on a variety of assumptions. Some of them, like constant population sizes and random mating in a Wright–Fisher population, are rarely met in natural populations (Whitlock and McCauley 1999; Hey and Machado 2003). Templeton et al. (1995) refer to a further major limitation of the use of F_{ST}, as the data used to estimate the F statistic often do not indicate which model of gene flow (e.g. "island model", "stepping stone model", "isolation by distance model") is appropriate for the populations being studied, particularly as the different models are not necessarily alternatives (see also Lynch and Crease 1990; Hudson et al. 1992). Besides, the geographical genetic variation measured by F_{ST} may not be caused by the current amount of gene flow but instead be shaped by events far back in time (Templeton et al. 1995).

An important analysis method is the hierarchical analysis of molecular variance (AMOVA) developed by Excoffier et al. (1992). AMOVA analyzes population subdivision through F-statistics by measuring the correlation between genetic variation drawn at different levels of a hierarchically subdivided population. These correlations can be influenced by several evolutionary forces, like mutation rates or migration (Excoffier and Heckel 2006). However, the procedure of AMOVA is based on several assumptions, like random mating and the absence of inbreeding, of which natural populations mostly depart. Moreover, AMOVA needs an a priori definition of the hierarchical structuring of populations, which might rely on wrong assumptions of the investigator. This problem might be solved by new analysis programs like STRUCTURE (Pritchard et al. 2000) or BAPS (Corander et al. 2003) which are based on Bayesian analysis algorithms and either deduce population structure directly from the data instead of using predefined settings or allow at least the testing of different assumptions.

Another frequently used statistical analysis is Mantel's test (Mantel 1967), which estimates the correlation between two distance matrices, for example, the genetic distance of the analyzed samples and the geographic distance of the collection sites.

For a review of population genetic methods and a compilation of software suitable for population genetic analyses see, for example, Pearse and Crandall (2004) and Excoffier and Heckel (2006). Additionally, Waples and Gaggiotti (2006) presented a review of methods to identify the number of gene pools or populations in a given sample.

Examples of studies

There are a number of questions that can be solved with the information about genetic structure in plants. It can be used to evaluate, for example, migration rates between populations, inbreeding coefficients, effective population sizes, spatial genetic structure, or hybridizations between taxa.

One fairly large field of research connected to population genetics is conservation biology. To preserve endangered species, information about their population genetic structure is important, because it can give hints about which populations are especially important to preserve (Maudet et al. 2002), or which measures should be taken to preserve a special population (Ellstrand and Elam 1993). The genetic structure of geographically very restricted species has been studied in many cases to evaluate possible genetic depletion. Genetic consequences of habitat degradation or fragmentation have also been studied (Kettle et al. 2007). For restoration of habitats and species reintroduction at a certain site, it is also often advisable to study the population genetics of the species, to choose the best source populations for seed transfer according to genetic diversity and possible local adaptations (Ramey et al. 2000; Smulders et al. 2000).

Other important fields of population genetics deal with the characteristics of invasive species and their provenance (Williams et al. 2005; Jahodová et al. 2007; Londo and Schaal 2007; Okada et al. 2007) or the study of recently originated allopolyploids (Abbott et al. 2007). Also, the population genetics of clonally growing species can be analyzed. For example, the relative importance of clonal versus sexual dispersal was studied by analyses of the population structure of sea grass (Alberto et al. 2005). Methods for assessing clonality were reviewed by Arnaud-Haond et al. (2007).

Population genetic data can also be combined with geographical and ecological data (for a review about

landscape genetics and related methods, see Manel et al. 2003) or be used to infer mating systems and factors such as inbreeding depression (Michalski and Durka 2007).

Phylogeography

Phylogeography as a distinct discipline arose in the late 1980s and combines microevolutionary (population genetics) and macroevolutionary (phylogenetics, systematics) concepts with the distribution of genetic variation in space and time (Avise et al. 1987; Avise 2000). In principle, the same genetic markers can be used in population genetics and phylogeography. The goals and analysis methods differ, however, even if population genetic methods are sometimes also used in phylogeographic analyses. The major difference between both fields is that population genetics interprets differences in allele distribution under the assumption of recent gene flow, while phylogeographic analyses explicitly seek to find out the historical processes that shaped the extant distribution of genetic variation. Phylogeography can, therefore, complement the analysis of fossil remains, such as pollen, or give insights into the history of species for which fossil remains are scarce or indistinguishable from other taxa.

Phylogeography and its basic assumptions

To reconstruct a species' history with a phylogeographic approach, the genetic variation within this species is organized into a genealogy and overlaid by the geographical distribution of the alleles of the marker region under study (Avise 1989). Genealogy here refers to the progenitor–derivative relationships among these alleles, mostly depicted via genealogical networks. The analysis then interprets patterns of congruence or incongruence between the extant geographic distribution of alleles and their genealogical relationships on the background of different recent and historical processes influencing the structuring of genetic diversity within and among populations, that is, geographic barriers, dispersal events, population size changes, and gene flow.

The basic assumptions of phylogeography are mostly derived from coalescent theory, the formal mathematical and statistical treatment of gene genealogies within and among related species (Felsenstein 1971; Griffiths 1980; Tavaré 1984; Hudson 1990, 1998). Coalescent theory describes the merging of allele lineages in common progenitor alleles when going back in time. It allows the recognition of the polarity (old vs. young haplotypes) from the topology of an unrooted tree or network together with frequencies of haplotypes. Castelloe and Templeton (1994) showed that tip haplotypes or clades (which are connected to the remaining network by only one connecting branch) are almost always younger than interior ones (which possess more than one connecting branch to the remaining cladogram). Thus, contrasting tips versus interiors strongly tends to contrast younger versus older alleles. Predictions from coalescent theory show further that older alleles should prevail in populations, and be characterized by a higher number of descending lineages and a geographically wider distribution than younger alleles (Neigel et al. 1991; Neigel and Avise 1993; Castelloe and Templeton 1994; Posada and Crandall 2001). Although often true, the validity of these assumptions, which were tested in simulation models (Castelloe and Templeton 1994), depends strongly on sampling design, as well as life history traits and differences in the history of the taxa under study (Jakob and Blattner 2006; Jakob et al. 2007). As alleles become extinct at random over time, unless the population permanently grows, the oldest alleles are not necessarily the most frequent ones within a taxon (Avise 2000). Also the number of descending alleles has to be considered critically, since a certain allele may not have any descendants (or not any more) within the species under study, but descendants might occur in close relatives, when incomplete lineage sorting, that is, the persistence of ancestral polymorphisms through speciation events, is present (Jakob and Blattner 2006). This can invert the time axis of a network and result in completely different conclusions about historic processes (Jakob and Blattner 2006; Liston et al. 2007).

The differences in genetic diversity among populations in different areas are used for the reconstruction of a species' history. The rationale behind this is the observation that during an expansion of the distribution area of a species, genetic diversity declines with distance from the starting point or center of distribution. This matches the hypothesis that the number of different haplotypes per area should be reduced during population expansion through repeated genetic bottlenecks occurring on the leading edge due to low population sizes and repeated founder events (Hewitt 2000) and the lower penetration rates of newly arising or arriving alleles in areas already occupied by conspecific individuals. A greater evenness in geographic haplotype frequency distribution can be expected if expansion started from multiple sources (Song et al. 2006) or if the population expansion is quite old, resulting in the slow admixture of many haplotypes within specific areas. As demography is critically important for allele numbers and their distribution, different historical events can be reconstructed from the allele patterns. For example, the foundation of a new population after long-distance dispersal should start with a single allele (or a very low number of alleles), which might be quite common in the source population. Thus, this allele will be shared between both populations. All alleles evolving after the

dispersal event will, however, be exclusive to the respective population and are private alleles or haplotypes of these groups, as long as no introgression occurs. Besides, the allelic richness should be much higher in the source population in comparison to the population that originated with the initial genetic bottleneck of a founder event. This situation of haplotype distribution can similarly be found in speciation events beginning with a strong bottleneck for the new species. The picture looks different if population differentiation or speciation happened through vicariance, that is, if large populations became separated. In this case we could expect nearly equal haplotype diversity in the sister taxa, and for a long time the occurrence of shared alleles (Hudson and Coyne 2002; Jakob and Blattner 2006; Syring et al. 2007). Also, hybridization might be recognized by haplotype distribution. If tip haplotypes are found in species where none of their progenitor alleles occur and/or if haplotypes are area-specific instead of taxon-specific, this is a strong indication for hybridization (Bänfer et al. 2006). For internal haplotypes, however, it is often impossible to discern incomplete lineage sorting, that is, shared alleles, from hybridization events. Another mechanism contributing to geographical haplotype differences is genetic drift, which might result in fixation of different alleles in fragmented populations. Such populations might show a low gene diversity, whereas diversity among populations is high (Jakob et al. 2007).

Genomes and markers useful for phylogeographic studies

Here, we will focus mostly on non-coding parts of the organelle genomes, as they are most frequently used in phylogeographic studies in plants. However, some recent studies also used the variation at coding or non-coding nuclear regions to infer intraspecific differentiation (Olsen and Schaal 1999; Morrell et al. 2003; Caicedo and Schaal 2004; Schmuths et al. 2004; Bartish et al. 2006; Joly et al. 2006; Koch et al. 2006; Schmid et al. 2006; Gurushidze et al. 2007). Nuclear loci might create problems in data interpretation as recombination occurs and, therefore, mosaic sequences might be included in a data set. These can distort gene or loci genealogies if they remain undetected. The recognition (and probably exclusion) of recombinant alleles is, therefore, crucial for data analysis (Schaal and Olsen 2000).

Uniparentally inherited organelle markers have specific qualities for phylogeographic studies, as effective population size should be reduced in these markers compared to nuclear markers, since they can be considered as effectively haploid (Birky et al. 1989; Petit et al. 2005). Smaller effective populations sizes should result in faster turnover rates for newly evolving genotypes, resulting in a clearer picture of past migration history than nuclear markers (Rendell and Ennos 2002; Hudson and Coyne 2002; Kadereit et al. 2005). Moreover, problems associated with recombination are mostly absent in these markers. Initially, mainly animal species were studied phylogeographically using mitochondrial markers (Avise 2000). These studies provided, for example, interesting insights into the origin and evolutionary history of modern human populations (Richards et al. 1998). Contrary to the situation in animals, the use of mitochondrial markers in plants, especially in angiosperms, is more restricted (Tomaru et al. 1998). Today cpDNA markers are the most commonly used markers in phylogeographic studies of angiosperms, although mitochondrial markers are widespread in studies of gymnosperm taxa (see above).

As no chloroplast region variable in all angiosperms or let alone all land plants exists, often several loci have to be tested to find regions with a suitable variability in the group under study (Borsch and Quandt 2008). These tests involve sequencing of different chloroplast intergenic spacers or introns for a certain number of individuals. Candidate regions are accessible via PCR using universal primers (Taberlet et al. 1991; Desmesure et al. 1995; Weising and Gardner 1999; variability of some regions reviewed in Shaw et al. 2005, 2007) or by screening published chloroplast genome sequences of closely related species, if available (Shaw et al. 2007; Jakob et al. 2007; Sacks and Louie 2008). As described earlier for microsatellite loci, it might be worth to look for core motifs with slightly repetitive structures, as these might have evolved into highly variable loci in the taxa under study.

Laboratory methods

Early studies of chloroplast variation often applied restriction fragment length polymorphism (RFLP) with fragment detection via Southern blotting (Wagner et al. 1987). PCR technique and development of universal chloroplast primers amplifying introns and intergenic spacers by Tablerlet et al. (1991), Demesure et al. (1995), and others, led to a growing number of studies finding more and more variation in the chloroplast genome. PCR–RFLP, that is, the PCR amplification of defined DNA regions and their digestion with restriction fragments, was thus often used (Tremblay and Schoen 1999; Stehlik et al. 2002). As the number of available restriction enzymes increased and PCR provided nearly unlimited amounts of DNA of the target loci, these studies could arrive at high numbers of detected polymorphisms and, therefore, the phylogeographic resolution could be nearly as good as with DNA sequencing (Stehlik et al. 2002). This method, although currently still cheaper than sequencing, involves a lot of hands-on time and is, therefore, successively replaced by direct sequencing of PCR

products of the respective target regions. This leads to another increase in detected polymorphisms and genetic resolution. However, also with sequencers using classical Sanger sequencing technology (i.e., four gel lanes to sequence one DNA strand) like, for example, radioactively labeled sequencing reactions or detection on ALF and machines, single nucleotide sequencing (Guicking et al. 2008) allows high throughput without losing much information within the analyzed region.

Microsatellite loci are sometimes also used in phylogeographic studies. However, due to their high variability, microsatellites are prone to homoplasy (Doyle et al. 1998; Ingvarsson et al. 2003), which particularly complicates the estimation of gene genealogies. According to some authors, they should rather be omitted if a reasonable number of other mutations are present because of the high risk of scoring non-homologous characters (Provan et al. 2001; Jakob and Blattner 2006). This depends, however, on the genetic distances among the studied populations or taxa, as SSR homoplasy should increase with the age of divergence and accordingly, the genetic distance. Bänfer et al. (2006) and Jakob et al. (2007) used a novel two-step approach to combine length variation at chloroplast SSR loci with sequence variation of intergenic spacer regions of the chloroplast genome, by first building a backbone genealogy on the basis of the sequence-based chloroplast haplotypes and adding the SSR variation for each so-defined haplotype creating subhaplotypes. Assuming lower mutation rates in the non-repetitive parts of the genome, this approach allows the use of chloroplast SSRs without introducing much homoplasy into the data sets.

Anonymous markers like AFLPs or RAPDs are less frequently used in phylogeographic studies than sequence-based marker regions because of their unknown genomic background. However, they are useful in cases were other markers reveal only very low genetic variation and have been used to analyze, for example, population histories in Arctic and alpine plant species (Gabrielsen et al. 1997; Tollefsrud et al. 1998; Friesen and Blattner 2000; Zhang et al. 2001; Stehlik 2002; Kropf et al. 2006).

Estimation of haplotype relationships

Beside the geographical distribution of the haplotypes, their relatedness is critically important for phylogeographic studies. Therefore, phylogenetic trees are often used to display these relationships. However, intraspecific gene evolution cannot be accurately depicted by bifurcating trees, but has to be represented by gene genealogies, that is, multifurcating networks (Clement et al. 2000; Posada and Crandall 2001), as ancestral alleles mostly coexist with multiple younger descendants. This phenomenon frequently results in zero-length branches within bifurcating trees, indicating that tips and internal nodes of a tree are occupied by extant taxa, alleles or individuals. In theory this should not be the case in phylogenies on higher taxonomic levels due to the longer time since population separation and, thus, the loss of ancestral alleles. However, a quick look through currently published phylogentic analyses of closely related species complexes immediately shows that zero-length branches are common even in interspecific studies. Also in these cases, network approaches often provide more information than phylogenetic trees and should, therefore, be used instead of or in addition to tree-based analyses (Gurushidze et al. 2007). Posada and Crandall (2001) discussed these problems in their detailed review and referred to distinct network approaches and currently available software packages to calculate gene genealogies.

The coexistence of progenitor and derivative alleles in a tree also causes problems for programs used to estimate absolute or relative ages of taxa and nodes in a tree like, for example, R8S (Sanderson 2002). The relative age of haplotypes can, however, be assessed using predictions of coalescent theory and their genealogical relationships (see above). Each haplotype network provides a time axis, from the oldest haplotypes at central and internal positions towards the tips of the network, where the youngest haplotypes are placed. Thus, rooting of a network is not always essential, as, contrary to phylogenetic trees, the reading direction is already inherent in a genealogy. As haplotypes will ultimately get lost after some time within a certain species due to lineage sorting, they appear in a genealogical network as missing intermediates, that is, their prior existence can be inferred from the character state differences of neighboring alleles. If the analyzed individuals cover the variation within a species quite well, the existence and distribution of these missing intermediate alleles in a genealogical network allows the inference of population history, as during population expansion, lineage sorting and, therefore, the loss of alleles is low, while with shrinking population size or an extreme genetic bottleneck lineage sorting and, therefore, allele extinction steeply increases (Jakob and Blattner 2006). Comparable to branch length distributions in phylogenetic trees (Barraclough and Vogler 2000; Barraclough and Nee 2001), missing intermediates reflect parameters of the population history in gene genealogies and networks. Deviations from the predictions of coalescent theory can, therefore, be expected if taxa are analyzed, where some ancient alleles are still present in some individuals, or if taxa or populations are compared that experienced rather different evolutionary influences during their histories (Jakob and Blattner 2006).

As haplotype relationships in a genealogical network are calculated by a statistical parsimony approach that represents each mutation as a step in the network, it is necessary to have single mutations represented as single characters in

the data matrix. In DNA sequence alignments of non-coding DNA, often insertions or deletions (indels) are the main informative characters. As these can partly be quite long but anyway evolved via a single mutation event, they have, therefore, to be represented as single alignment positions. This means that longer indels occurring in some individuals have to be shortened to a single character state in the alignment, which might cause problems when several informative positions occur in that stretch of DNA where present. In this case, it is sometimes necessary to code this part so that all information can be represented and adjust the necessary mutational steps in the network afterwards manually to represent the correct haplotype relationships. As already mentioned, it could be necessary to exclude parts of the data matrix from the analysis if homology of the alignment positions (or more general, characters) could not be estimated safely (Morrison 2008). This holds particularly true for microsatellite loci but also some other repetitive parts might cause problems. Homoplasy in genealogical networks results in closed loops, that is, relationships are represented by several different connections between the haplotypes. Although these loops can sometimes be resolved based on coalescence assumptions, this problem is not different from homoplasy in phylogenetic analysis (Kelchner and Graham 2008) and can best be solved by using marker regions with adequate variability, exclusion of non-alignable sequence parts or involving a kind of weighting scheme, as described before for the two-step procedure of network construction when microsatellites are analyzed together with non-repetitive DNA parts (Bänfer et al. 2006).

Phylogeographic analysis methods and their problems

As mentioned before, traditional population genetic approaches also, that deal with the spatial frequencies of alleles and are mostly based on equilibrium expectations derived from the theoretical model of population structure under neutrality theory (reviewed in Felsenstein 1982; Slatkin 1985; Slatkin and Barton 1989; Neigel 1997), are partly used in phylogeographic analyses. Particularly, the hierarchical analysis of molecular variance (AMOVA) by Excoffier et al. (1992) and Mantel's test (Mantel 1967) must be named here (see above). To describe the genetic diversity of populations and their differentiation, several diversity measures can also be evaluated, for example, Nei's gene diversity H (Nei 1987), the haplotype richness R (El Mousadik and Petit 1996), or the number and distribution of population-specific haplotypes, so called private alleles (Stehlik 2002).

Templeton et al. (NCA; Templeton et al. 1995; Templeton 1998 and references therein) developed a method specifically for phylogeographic purposes, called nested clade analysis that tries to include the principles and basic coalescence assumptions mentioned before. This method is supposed to be able to distinguish between recurrent gene flow and a variety of historical processes, like past fragmentation, long distance colonization events, and range expansions. There is also a computer program available (GeoDis; Posada et al. 2000), which, together with an inference key, implements NCA. Due to its simple usage and ready-to-use results, the method was, and is, still the most popular in phylogeographic studies. Recently, different approaches have been made to automate the whole procedure of NCA (Zhang et al. 2006; Panchal 2007).

Nested clade analysis is, however, not undisputed. First, it was proposed that the choice of a randomization strategy for NCA greatly affects the outcome of the analyses (Petit and Grivet 2002). Secondly, Knowles and Maddison (2002) stressed that, although statistical significance is computed for association between the geographical distribution of haplotypes and their nested clade, no confidence limits can be assessed for the interpretations drawn from the inference key. In an analysis of simulated data sets with NCA, they found that NCA, in most cases, did not identify the processes that were used to simulate the data. Templeton (2004) counter argued that NCA performs reasonably well concerning hypotheses of range expansions and past fragmentations. However, Panchal and Beaumont (2007) conducted another study on a larger number of real and simulated datasets by using the software provided by Panchal (2007). They found that NCA returned a high number of false positives (but see Templeton 2008). Another potential problem concerns possible local population extinctions. Hung et al. (2005) showed that this can lead to erroneous identification of long distance dispersal in NCA. For these reasons, Petit (2008) recommended to refrain from using NCA until the method is further evaluated. A probable solution might be the use of statistical phylogeography (Knowles 2004; Knowles and Carstens 2007) where a wide range of demographical and biogeographical processes can be accommodated, provided biologically realistic models are available. However, as currently no easy usable software package or at least a protocol for these methods exists, we assume that it will take some time until statistical phylogeography will become a widespread analysis method.

As mentioned above, it is sometimes found that haplotypes are not restricted to the species under study. One possible explanation for such shared haplotypes is hybridization, which was inferred, for example, in European species of *Quercus* (Petit et al. 2002), *Betula* (Palme et al. 2004), and *Arabis* (Dobeš et al. 2004). However, evidence is accumulating that incomplete or differential lineage sorting also contributes to the distribution of haplotypes among different species (Mason-Gamer et al.

1999; Wendel and Doyle 1999; Comes and Abbott 2001; Linder and Rieseberg 2004; Syring et al. 2007). This means that the population involved in speciation was of considerable size, or a new species was formed more than once, so that more than one haplotype of the progenitor species is also present in the new species. This pattern is common in the New World species of the genus *Hordeum* (Jakob and Blattner 2006), where virtually all interior haplotypes are shared among up to six species. We assume that the same phenomena will be found in many young or rapidly radiating species groups, when sampling is expanded and more close relatives are also included in phylogeographic studies. This is supported by findings that chloroplast alleles might survive for quite a long time. Jakob and Blattner (2006) estimated survival times of chloroplast alleles within the genus *Hordeum* to up to 4 million years (Ma). This value is an order of magnitude higher than in, for example, Mediterranean *Senecio* (Asteraceae) where minimum survival times of shared chloroplast haplotypes of 0.44 Ma were found (Comes and Abbott 2001). Therefore, it is not surprising that some of the polymorphisms can be much older than the species in which they are found (Hudson and Coyne 2002). Generally, haplotypes shared between different species lead to considerable problems for phylogeographic studies, as it is normally difficult or even impossible to discern between hybridization and incomplete lineage sorting (Bänfer et al. 2006). The history of each haplotype (e.g. geographical restriction due to a population bottleneck, expansion or even extinction) depends strongly on the fate of the species in which the haplotype occurs. For example, geographical patterns arisen during the history of an ancestral taxon may still be visible and may conceal historic processes, which underlay the distribution of genetic diversity of derived extant taxa. Thus, also for intraspecific analyses, it might be advisable to include a certain amount of closely related taxa and to check carefully the geographic distribution of shared haplotypes on different spatial scales, which may allow distinguishing between both processes (Jakob and Blattner 2006).

Examples of phylogeographic studies

The rationale of most phylogeographic studies is, for example, to identify putative Pleistocene refugia, re-colonized areas, migration and dispersal corridors and, thus, to explain the observed population structure and geographic distribution of genetic diversity as well as to enlighten the evolutionary history of a certain species. However, phylogeographic methods have also been applied to study speciation and hybridization events (Joly et al. 2006; Jakob and Blattner 2006) or the history of polyploid formation (van Dijk and Bakx-Schotman 1997).

Phylogeographic studies have been conducted in many European taxa, studying the number and location of potential ice age refugia and re-colonization routes (Johansen and Latta 2003; Rendell and Ennos 2002; Bartish et al. 2006), sometimes in conjunction with fossil pollen data (Petit et al. 2002). In North America, a number of studies have been conducted on the same topic (Soltis et al. 1997 and references therein; Johansen and Latta 2003; Griffin and Barrett 2004; Godbout et al. 2005). There are also studies on plant species with special distribution patterns, like arctic plants (Tremblay and Schoen 1999; Abbott et al. 2000; Abbott and Brochmann 2003; Alsos et al. 2005) or alpine plants (Holderegger et al. 2002; Comes and Kadereit 2002; Stehlik 2003; Schönswetter et al. 2005; Dixon et al. 2007), and the post-ice age re-colonization of the continents by trees in Europe and North America (Dumolin-Lapegue et al. 1997; Petit et al. 2002; Grivet et al. 2006). The number of comparative phylogeographical studies increased remarkably during the last years. More recently, these studies started to look for congruence among phylogeographical patterns of a number of different species sharing the same distribution range (Soltis et al. 1997, 2006; Taberlet et al. 1998; Calsbeek et al. 2003; Kadereit et al. 2005), or to analyze the phylogeographic pattern in widely distributed species, where the genetic consequences of the same event (e.g. the Pleistocene) could have been quite different for each of the subpopulations (Ehrlich et al. 2007). For reviews on the genetic consequences of the ice ages for plants from different regions of the world see Hewitt (2000, 2004).

All in all, after more than a decade of phylogeographic studies on European or North American taxa, the knowledge of the response of plants and animals to Pleistocene climate changes, of the localization of ice age refugia, and patterns of postglacial re-colonization of formerly glaciated areas, increased considerably. However, not much is known about the impact of Pleistocene climate oscillations on plant species in the mountain ranges of Asia and Africa. Generally, phylogeographic studies in Africa, Asia, and the southern hemisphere are still scarce compared to Europe or North America. Cannon and Manos (2003), for example, described the current geographical patterns of genetic diversity and inferred the historical population dynamics of the stone oaks (*Lithocarpus*) in Southeast Asia. Hung et al. (2005) used a phylogeographic study to enlighten the change of genetic diversity within *Lithocarpus konishii* after an earthquake. Aizawa et al. (2007) analyzed the phylogeography of *Picea jezoensis* by a combination of mitochondrial and chloroplast markers. Takayama et al. (2006) scrutinized the phylogeography and genetic structure of *Hibiscus tiliaceus*, a pantropical plant with an interesting modus of seed dispersal. Nettel and Dodd (2007) used a phylogeographic study to reveal dispersal

patterns of mangrove species along tropical coasts. Some phylogeographic studies on species from the southern hemisphere have also been conducted. For example, Shepherd et al. (2007) studied the volcanic and glacial impacts on the distribution of the forest fern *Asplenium hookerianum* and Gardner et al. (2004) studied the late Quaternary phylogeography of *Metrosideros* (Myrtaceae) in New Zealand. Most of the existing phylogeographic studies from South America concern tree species and/or species from the Andes (Pastorino and Gallo 2002; Marchelli and Gallo 2004, 2006; Muellner et al. 2005), while rainforest species were rarely included (Olsen and Schaal 1999), and studies of the taxa of the huge pampa and steppe areas extending between the southern Andes and the Atlantic coast are completely absent up to now. Generally, studies of tropical plant groups, no matter if herbs, trees or epiphytes, are comparably rare, owing mostly to the problem of getting the good population representation necessary for phylogeographic studies. Thus, in groups where it is hard to obtain even samples of single individuals, as mostly included in phylogenetic studies, phylogeographic data will, for a long time, not become available unless major efforts are put into collecting good population samples. We assume, therefore, that future phylogeographic analyses in remote or hardly accessible regions will mostly be restricted to economically important species (Olsen and Schaal 1999). However, we think that more studies conducted on species outside Europe and North America are necessary to elucidate species and vegetation history, particularly in the background of the extant global climate change.

References

Abbott RJ, Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hultén. Molec Ecol 12:299–313

Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, Balfour J (2000) Molecular analysis of plant migration and refugia in the Arctic. Science 289:1343–1346

Abbott RJ, Ireland HE, Rogers HJ (2007) Population decline despite high genetic diversity in the new allopolyploid species *Senecio cambrensis* (Asteraceae). Molec Ecol 16:1023–1033

Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, Shi F, Kaji M (2007) Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. Molec Ecol 16:3393–3405

Alberto F, Gouveia L, Arnaud-Haond S, Pérez-Lloréns JL, Duarte CM, Serrão EA (2005) Within-population spatial genetic structure, neighbourhood size and clonal subrange in the seagrass *Cymodocea nodosa*. Molec Ecol 14:2669–2681

Alsos IG, Engelskjøn T, Gielly L, Taberlet P, Brochmann C (2005) Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. Molec Ecol 14:2739–2753

Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA (2007) Standardizing methods to address clonality in population studies. Molec Ecol 16:5115–5139

Avise JC (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution 43:1192–1208

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Rev Ecol Syst 18:489–522

Bachmann K (2001) Evolution and the genetic analysis of populations: 1950–2000. Taxon 50:7–45

Bänfer G, Moog U, Fiala B, Mohamed M, Weising K, Blattner FR (2006) A chloroplast genealogy of myrmecophytic *Macaranga* species (Euphorbiaceae) in Southeast Asia reveals hybridization, vicariance and long-distance dispersals. Molec Ecol 15:4409–4424

Banks JA, Birky CW (1985) Chloroplast DNA diversity is low in a wild plant, *Lupinus texensis*. Proc Naturalist Acad Sci USA 82:6950–6954

Barraclough TG, Nee S (2001) Phylogenetics and speciation. Trends Ecol Evol 16:391–399

Barraclough TG, Vogler AP (2000) Detecting the geographical pattern of speciation from species-level phylogenies. Amer Naturalist 155:419–434

Bartish IV, Kadereit JW, Comes HP (2006) Late Quaternary history of *Hippophaë rhamnoides* L. (Elaeagnaceae) inferred from chalcone synthase intron (*Chsi*) sequences and chloroplast DNA variation. Molec Ecol 15:4065–4083

Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? Trends Ecol Evol 20:435–440

Birky CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121:613–627

Blattner FR (1999) Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. Biotechniques 29:1180–1186

Bosch T, Quandt D (2008) Mutational dynamics and phylogenetic utility of chloroplast non-coding DNA. Pl Syst Evol Suppl (this issue)

Caicedo AL, Schaal BA (2004) Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene. Molec Ecol 13:1871–1882

Calsbeek R, Thompson JN, Richardson JE (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. Molec Ecol 12:1021–1029

Cannon CH, Manos PS (2003) Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*). J Biogeogr 30:211–226

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. Molec Phylogenet Evol 3:102–113

Chiang Y-C, Hung K-H, Schaal BA, Ge X-J, Hsu T-W, Chiang T-Y (2006) Contrasting phylogeographical patterns between mainland and island taxa of the *Pinus luchuensis* complex. Molec Ecol 15:765–779

Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. Nature Genetics 38:594–597

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Molec Ecol 9:1657–1660

Comes HP, Abbott RJ (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). Evolution 55:1943–1962

Comes HP, Kadereit JW (2002) Spatial and temporal patterns in the evolution of the flora of the European alpine system. Taxon 52:451–462

Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163:367–374

Demesure B, Sodzi N, Petit R (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Molec Ecol 4:129–131

Dixon CJ, Schönswetter P, Schneeweiss GM (2007) Traces of ancient range shifts in a mountain plant group (*Androsace halleri* complex, Primulaceae). Molec Ecol 16:3890–3901

Dobeš CH, Mitchell-Olds T, Koch MA (2004) Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A.* × *divaricarpa*, and *A. holboellii* (Brassicaceae). Molec Ecol 13:349–370

Dong J, Wagner DB (1994) Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. Genetics 136:1187–1194

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). Molec Biol Evol 15:215–218

Dumolin-Lapegue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475–1487

Ehrlich D, Gaudeul M, Assefa A, Koch MA, Mummenhoff K, Nemomissa S, Intrabiodiv Consortium, Brochmann C (2007) Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. Molec Ecol 16:2542–2559

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annual Rev Ecol Syst 24:217–242

El Mousadik A, Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. Molec Ecol 5:547–555

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. Nature Rev Genet 7:745–758

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Felsenstein J (1971) The rate of loss of multiple alleles in finite haploid populations. Theor Pop Biol 2:391–403

Felsenstein J (1982) How can we infer geography and history from gene frequencies? J Theor Biol 96:9–20

Fénart S, Austerlitz F, Cuguen J, Arnaud J-F (2007) Long distance pollen-mediated gene flow at a landscape level: the weed beet as a case study. Molec Ecol 16:3801–3813

Fisher RA (1930) The genetical theory of natural selection. Clarendon Press, Oxford

Friesen N, Blattner FR (2000) RAPD analysis reveals geographic differentiations within *Allium schoenoprasum* L. (Alliaceae). Pl Biol 2:297–305

Gabrielsen TM, Bachmann K, Jakobsen KS, Brochmann C (1997) Glacial survival does not matter: RAPD phylogeography of nordic *Saxifraga oppositifolia*. Molec Ecol 6:831–842

Gardner RC, De Lange PJ, Keeling DJ, Bowala T, Brown HA, Wright SD (2004) A late Quaternary phylogeography for *Metrosideros* (Myrtaceae) in New Zealand inferred from chloroplast DNA haplotypes. Biol J Linn Soc 83:399–412

Glémin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc Roy Soc Lond Series B 273:3011–3019

Godbout J, Jaramillo JP, Beaulieu J, Bousquet J (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. Molec Ecol 14:3497–3512

Griffin SR, Barrett SCH (2004) Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: inferences from phylogeography. Amer J Bot 91:465–473

Griffiths RC (1980) Lines of descent in the diffusion approximation of neutral Wright–Fischer models. Theor Pop Biol 17:40–50

Grivet D, Deguilloux M-F, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. Molec Ecol 15:4085–4093

Guicking D, Rana TS, Blattner FR, Weising K (2006) Microsatellite markers for the paleotropical pioneer tree genus *Macaranga* (Euphorbiaceae) and their cross-species transferability. Molec Ecol Notes 6:245–248

Guicking D, Kröger-Kilian T, Weising K, Blattner FR (2008) Single nucleotide sequence analysis: a cost- and time-effective protocol for the analysis of microsatellite- and indel-rich chloroplast DNA regions. Molec Ecol Res 8:62–65

Guo XY, Wang Y, Keightley PD, Fan LJ (2007) Patterns of selective constraints in noncoding DNA of rice. BMC Evol Biol 7:208

Gurushidze M, Mashayekhi S, Blattner FR, Friesen N, Fritsch RM (2007) Phylogenetic relationships of wild and cultivated species of *Allium* section *Cepa* inferred by nuclear rDNA ITS sequence analysis. Pl Syst Evol 269:259–269

Haldane JBS (1932) The causes of evolution. Longmans, Green and Co., London

Halligan DL, Eyre-Walker A, Andolfatto P, Keightley PD (2004) Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. Genome Res 14:273–279

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Phil Trans Roy Soc Lond B 359:183–195

Hey J, Machado CA (2003) The study of structured populations—new hope for a difficult and divided science. Nature Rev Genet 4:535–543

Holderegger R, Stehlik I, Abbott RJ (2002) Molecular analysis of the Pleistocene history of *Saxifraga oppositifolia* in the Alps. Molec Ecol 11:1409–1418

Huang S, Chiang YC, Schaal BA, Chou CH, Chiang TY (2001) Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. Molec Ecol 10:2669–2681

Hudson RR (1990) Gene genealogies and the coalescent process. Oxford Surv Evol Biol 7:1–44

Hudson RR (1998) Island models and the coalescent process. Molec Evol 7:413–418

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589

Hung K-H, Hsu T-W, Schaal BA, Chiang T-Y (2005) Loss of genetic diversity and erroneous phylogeographical inferences in *Lithocarpus konishii* (Fagaceae) of Taiwan caused by the Chi-Chi earthquake: implications for conservation. Ann Missouri Bot Gard 92:52–65

Ingvarsson PK, Ribstein S, Taylor DR (2003) Molecular evolution of insertions and deletion in the chloroplast genome of *Silene*. Molec Biol Evol 20:1737–1740

Jahodová Š, Trybush S, Pyšek P, Wade M, Karp A (2007) Invasive species of *Heracleum* in Europe: an insight into genetic relationships and invasion history. Diversity Distrib 13:99–114

Jakob SS, Blattner FR (2006) A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. Molec Biol Evol 23:1602–1612

Jakob SS, Ihlow A, Blattner FR (2007) Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae)—niche

differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. Molec Ecol 16:1713–1727
Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004) Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. Molec Ecol 13:2735–2747
Johansen AD, Latta RG (2003) Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. Molec Ecol 12:293–298
Joly S, Starr JR, Lewis WH, Bruneau A (2006) Polyploid and hybrid evolution in roses east of the Rocky Mountains. Amer J Bot 93:412–425
Kadereit JW, Arafeh R, Somogyi G, Westberg E (2005) Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. Taxon 54:861–876
Kelchner SA, Graham SW (2008) Room for improvement: phylogenetic models and model selection for noncoding DNA. Pl Syst Evol Suppl (this issue)
Kettle CJ, Hollingsworth PM, Jaffré T, Moran B, Ennos RA (2007) Identifying the early genetic consequences of habitat degradation in a highly threatened tropical conifer, *Araucaria nemorosa* Laubenfels. Molec Ecol 16:3581–3591
Knowles LL (2004) The burgeoning field of statistical phylogeography. J Evol Biol 17:1–10
Knowles LL, Carstens BC (2007) Estimating a geographically explicit model of population divergence. Evolution 61:477–493
Knowles LL, Maddison WP (2002) Statistical phylogeography. Molec Ecol 11:2623–2635
Koch MA, Kiefer C, Ehrich D, Vogel J, Brochmann C, Mummenhoff K (2006) Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). Molec Ecol 15:825–839
Koch MA, Dobes C, Kiefer C, Schmickl R (2008) Evolution of plastidic *trn*F pseudogenes in cruciferous plants. Pl Syst Evol Suppl (this issue)
Kropf M, Comes HP, Kadereit JW (2006) Long-distance dispersal vs vicariance: the origin and genetic diversity of alpine plants in the Spanish Sierra Nevada. New Phytol 172:169–184
Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115
Latta RG, Linhart YB, Fleck D, Elliot M (1998) Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. Evolution 52:61–67
Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial gene flow among refugia. Proc Natl Acad Sci USA 99:14590–14594
Linder CR, Rieseberg LH (2004) Reconstructing patterns of reticulate evolution in plants. Amer J Bot 91:1700–1708
Liston A, Parker-Defeniks M, Syring JV, Willyard A, Cronn R (2007) Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: a case study in *Pinus lambertiana*. Molec Ecol 16:3926–3937
Londo JP, Schaal BA (2007) Origins and population genetics of weedy red rice in the USA. Molec Ecol 16:4523–4535
Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Molec Biol Evol 7:377–394
Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189–197
Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220
Marchelli P, Gallo LA (2004) The combined role of glaciation and hybridization in shaping the distribution of genetic variation in a Patagonian southern beech. J Biogeography 31:451–460
Marchelli P, Gallo LA (2006) Multiple ice-age refugia in a southern beech of South America as evidenced by chloroplast DNA markers. Conservation Genet 7:591–603
Mason-Gamer RJ, Holsinger KE, Jansen RK (1999) Chloroplast DNA variation in *Coreopsis nuecensoides* and *C. nuecensis* (Asteraceae), a presumed progenitor-derivative species pair. Pl Syst Evol 218:5–12
Maudet C, Miller C, Bassano B, Breitenmoser-Wüsten C, Gauthier D, Obexer-Ruff G, Michallet J, Taberlet P, Luikart G (2002) Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in Alpine ibex [*Capra ibex* (*ibex*)]. Molec Ecol 11:421–436
McCauley DE (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends Ecol Evol 10:198–202
Michalski SG, Durka W (2007) High selfing and high inbreeding depression in peripheral populations of *Juncus atratus*. Molec Ecol 16:4715–4727
Milligan BG (1991) Chloroplast DNA diversity within and among populations of *Trifolium pratense*. Curr Genet 19:411–416
Morrell PL, Lundy KE, Clegg MT (2003) Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. Proc Natl Acad Sci USA 100:10812–10817
Morrison D (2008) Homology assessment in non-coding genomic regions. Pl Syst Evol Suppl (this issue)
Muellner AN, Tremetsberger K, Stuessy T, Baeza CM (2005) Pleistocene refugia and recolonization routes in the southern Andes: insights from *Hypochaeris palustris* (Asteraceae, Lactuceae). Molec Ecol 14:203–212
Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323
Nei M (1987) Molecular evolutionary genetics. Columbia Univ. Press, New York
Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. Annual Rev Ecol Syst 28:105–128
Neigel JE, Avise JC (1993) Application of a random-walk model to geographic distribution of animal mitochondrial DNA. Genetics 135:1209–1220
Neigel JE, Ball RM, Avise JC (1991) Estimation of single generation migration distances from geographic variation in animal mitochondrial DNA. Evolution 45:423–432
Nettel A, Dodd RS (2007) Drifting propaguls and receding swamps: genetic footprints of mangrove recolonization and dispersal along tropical coasts. Evolution 61:958–971
Nybom H (2004) Comparison of different nuclear markers for estimating intraspecific genetic diversity in plants. Molec Ecol 13:1143–1155
Okada M, Ahmad R, Jaseniuk M (2007) Microsatellite variation points to local landscape plantings as sources of invasive pampas grass (*Cortaderia selloana*) in California. Molec Ecol 16:4956–4971
Olsen KM, Schaal BA (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proc Natl Acad Sci USA 96:5586–5591
Palme AE, Su Q, Palsson S, Lascoux M (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. Molec Ecol 13:167–178
Palmer JD (1992) Mitochondrial DNA in plant systematics: applications and limitations. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 26–49
Panchal M (2007) The automation of nested clade phylogeographic analysis. Bioinformatics 4:509–510

Panchal M, Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. Evolution 61:1466–1480
Pastorino MJ, Gallo LA (2002) Quaternary evolutionary history of *Austrocedrus chilensis*, a cypress native to the Andean Patagonian forest. J Biogeography 29:1167–1178
Pearse DE, Crandall KA (2004) Beyond F_{ST}: analysis of population genetic data for conservation. Conservation Genet 5:585–602
Petit RJ (2008) The coup de grâce for the nested clade phylogeographic analysis? Molec Ecol 17:516–518
Petit RJ, Grivet D (2002) Optimal randomizaton strategies when testing the existence of a phylogeographic structure. Genetics 161:469–471
Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Svejgaard Jensen J, König AO, Lowe AJ, Flemming Madsen S, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu J-L, Kremer A (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecol Manag 156:49–74
Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Molec Ecol 14:689–701
Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol 16:37–45
Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Molec Ecol 9:487–488
Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–77763
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999) A low mutation rate for chloroplast microsatellites. Genetics 153:943–947
Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147
Ramey RR, Luikart G, Singer FJ (2000) Genetic bottlenecks resulting from restoration efforts: the case of bighorn sheep in Badlands National Park. Rest Ecol 8:85–90
Rein T, Warmund U, Groth-Malonke M, Frahm JP, Knoop V (2008) Fifty mosses and three trees: comparing non-coding loci in the mitochondrial DNA in bryophyte phylogeny. Pl Syst Evol Suppl (this issue)
Rendell S, Ennos RA (2002) Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. Molec Ecol 11:69–78
Richards MB, Macauly VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260
Sacks BN, Louie S (2008) Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. Molec Ecol Res 8:35–49
Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molec Biol Evol 19:101–109
Schaal BA, Olsen KM (2000) Gene genealogies and population variation in plants. Proc Natl Acad Sci USA 97:7024–7029
Schmid K, Torjek O, Meyer R, Schmuths H, Hoffmann MH, Altmann T (2006) Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. Theor Appl Genet 112:1104–1114
Schmuths H, Hoffmann MH, Bachmann K (2004) Geographic distribution and recombination of genomic fragments on the short arm of chromosome 2 of *Arabidopsis thaliana*. Pl Biol 6:128–139
Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. Molec Ecol 14:3547–3555
Shaw J, Lickey AB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Amer J Bot 92:142–166
Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Amer J Bot 94:275–288
Shepherd LD, Perrie LR, Brownsey J (2007) Fire and ice: volcanic and glacial impacts on the phylogeography of the New Zealand forest fern *Asplenium hookerianum*. Molec Ecol 16:4536–4549
Slatkin M (1985) Gene flow in natural populations. Annual Rev Ecol Syst 16:393–430
Slatkin M, Barin NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349–1368
Smulders MJM, van der Schoot J, Geerts RHEM, Antonisse-de Jong AG, Korevaar H, van der Werf A, Vosman B (2000) Genetic diversity and the reintroduction of meadow species. Pl Biol 2:447–454
Soltis DE, Soltis PS, Milligan BG (1992) Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 117–150
Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Pl Syst Evol 206:353–373
Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. Molec Ecol 15:4261–4293
Song B-H, Clauss MJ, Pepper A, Mitchell-Olds T (2006) Geographic patterns of microsatellite variation in *Boechera stricta*, a close relative of *Arabidopsis*. Molec Ecol 15:357–369
Stehlik I (2002) Glacial history of the alpine herb *Rumex nivalis* (Polygonaceae): a comparison of common phylogeographic methods with nested clade analysis. Amer J Bot 89:2007–2016
Stehlik I (2003) Resistance or emigration? Response of alpine plants to the ice ages. Taxon 52:499–510
Stehlik I, Blattner FR, Holderegger R, Bachmann K (2002) Nunatak survival of the high Alpine plant *Eritrichium nanum* (L.) Gaudin in the central Alps during the ice ages. Molec Ecol 11:2027–2036
Syring J, Farrell K, Businsky R, Cronn R, Liston A (2007) Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Syst Biol 56:163–181
Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Pl Molec Biol 17:1105–1109
Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molec Ecol 7:453–464
Takayama K, Kajita T, Murata J, Tateishi Y (2006) Phylogeography and genetic structure of *Hibiscus tiliaceus*—speciation of a pantropical plant with sea-drifted seeds. Molec Ecol 15:2871–2881
Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652–656

Tavaré S (1984) Line-of-descent and genealogical processes, and their applications in population genetic models. Theor Pop Biol 26:119–164

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Molec Ecol 7:381–397

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. Molec Ecol 13:789–809

Templeton AR (2008) Nested clade analysis: an extensively validated method for strong phylogeographic inference. Molec Ecol 17:1877–1880

Templeton AR, Routmann E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrum*. Genetics 140:767–782

Tollefsrud MM, Bachmann K, Jakobsen KS, Brochmann C (1998) Glacial survival does not matter II: RAPD phylogeography of nordic *Saxifraga cespitosa*. Molec Ecol 7:1217–1232

Tomaru N, Takahashi M, Tsumura Y, Takahashi M, Ohba K (1998) Intraspecific variation and phylogeographic patterns of *Fagus crenata* (Fagaceae) mitochondrial DNA. Amer J Bot 85:629–636

Tremblay NO, Schoen DJ (1999) Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. Molec Ecol 8:1187–1198

van Dijk P, Bakx-Schotman T (1997) Chloroplast DNA phylogeography and cytotype geography in autopolyploid *Plantago media*. Molec Ecol 6:345–352

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Freuters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP, a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wagner DB (1992) Nuclear, chloroplast, and mitochondrial DNA polymorphisms as biochemical markers in population genetic analyses of forest trees. New For 6:373–390

Wagner DB, Furnier GR, Saghai-Maroof MA, Willians SM, Dancik BP, Allard RW (1987) Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc Natl Acad Sci USA 84:2097–2100

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molec Ecol 15:1419–1439

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Wendel JF, Doyle JJ (1999) Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis DE, Soltis PS, Doyle JJ (eds) Molecular systematics of plants II: DNA sequencing. Kluwer, Dordrecht, pp 265–296

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 315–322

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4 Nm + 1)$. Heredity 82:117–125

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Williams DA, Overholt WA, Cuda JP, Hughes CR (2005) Chloroplast and microsatellite DNA diversities reveal the introduction history of Brazilian peppertree (*Schinus terebinthifolius*) in Florida. Molec Ecol 14:3643–3656

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Wright S (1931) Evolution in Mendelian population. Genetics 16:97–159

Wright S (1951) The genetical structure of populations. Eugenics 15:323–354

Xu DH, Abe J, Sakai M, Kanazawa A, Shimamoto Y (2000) Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. Theoret Appl Genet 101:742–732

Zhang L-B, Comes HP, Kadereit JW (2001) Phylogeny and Quaternary history of the European montane/alpine endemic *Soldanella* (Primulaceae) based on ITS and AFLP variation. Amer J Bot 88:2331–2345

Zhang A-B, Tan S, Sota T (2006) AUTOINFER 1.0: a computer program to infer biogeographical events automatically. Molec Ecol Notes 6:597–599