REVIEWS

Species delimitation and relationships: The dance of the seven veils

Yamama Naciri[1] & H. Peter Linder[2]

1 *Plant Systematics and Biodiversity Laboratory, Conservatoire et Jardin botaniques & University of Geneva, Chemin de l'Impératrice 1, 1292 Chambésy, Geneva Switzerland*

2 *Institute of Systematic Botany, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland*

Author for correspondence: *Yamama Naciri, Yamama.Naciri@ville-ge.ch*

ORCID (http://orcid.org/): YN, 0000-0001-6784-8565; HPL, 0000-0002-1373-2708

DOI http://dx.doi.org/10.12705/641.24

Abstract An increasing number of studies using genetic data at the generic and species levels reveal complex patterns of relationships among populations and species. Incomplete lineage sorting and reticulation are, for instance, commonly observed in low-rank phylogenies. These two processes falsify the simplifying assumptions often used when reconstructing phylogenetic relationships or when assigning specimens to species using barcodes, i.e., the absence of ancestral polymorphism or the obligate dichotomous relationship among species. We recognize seven processes that act as veils to obscure species boundaries and relationships and lead to erroneous interpretations if not diagnosed and analysed properly. These processes include the transfer of genetic material from plastids or mitochondria to the nuclear genome (NuPt, NuMt), hybridization, the stochasticity of lineage sorting within and among species, genome organization, species demography, selection, genetic and geographic structure. Most of these processes have a direct impact on species effective population sizes, a central parameter to be considered. Failure to take the seven processes into account can affect the accuracy of barcode-based identification, in particular between closely related species. It can also bias the delimitation of species, result in inaccurate estimations of species relationships and lead to incorrect estimation of species ages. We suggest that these problems can be identified and, to some extent, mitigated by ensuring that the full spatial and genetic ranges of the species are sampled, and by the use of information from all genomes of the species. By analysing numerous loci, each as a separate entity in a coalescence framework, it is possible to take into account stochasticity as well as variation in genome effective population sizes. Such a critical approach would allow the use of genetic data to make realistic assessments of species delimitation, affinities and ages, and should promote a more biologically realistic view of species and speciation.

Keywords coalescence; effective population size; genetic structure; hybridization; incomplete lineage sorting; NuPt; phylogenetics; phylogeography; population genetics; speciation; species delimitation

■ INTRODUCTION

During the last decade, there has been much use of DNA sequence variation data at or near species level to explore the relationships among closely related species, to date species ages, or to delimit species (Bardy & al., 2011; Aguirre-Planter & al., 2012; Flores-Renteria & al., 2013; Zozomová-Lihová & al., 2014). This was often based on evolutionary assumptions that are appropriate at high taxonomic levels (families and above) but that may not be valid at the species level. These assumptions include, for instance, a dichotomous and hierarchical relationship among species, the absence of ancestral polymorphism, and a negligible influence of demography or mating systems on genetic variation within and among species. In the past, the effect of invalid assumptions at the species level was obscured by the use of few specimens per species (often a single one) and by the analysis of sequence data from only one genome or even only one locus. Recently, population genetics, applied at the species level to phylogeography (Jaramillo-Correa & al., 2009; Pandey & Rajora, 2012; Christe & al., 2014a) and barcoding (CBOL Plant Working Group, 2009; Ashfaq & al., 2013; Feng & al., 2013), led to concern about the consequences of these inappropriate assumptions at the species level or above.

Taxonomists' lives would be simple if a clean phylogenetic signal right down to species could be obtained, so that sequence data can be used to build a phylogeny to species level. However, there are several indications that a clean phylogenetic signal may only be rarely found at species level. Complex patterns were first assumed to be restricted to phylogenetically complicated groups that were referred to as taxonomically complex groups (TCG) by Federici & al. (2013). Groups that contain hybrid swarms, for instance *Veronica barrelieri* H.Schott ex Roem. & Schult. (Bardy & al., 2011), constitute typical TCGs. In TCGs the interaction amongst, *inter alia*, morphology, ploidy, phylogeographic pattern, demography and breeding system may be too complex to be adequately captured by dichotomous, hierarchical phylogenies based on few

Received: 24 Jul 2014 | returned for (first) revision: 10 Dec 2014 | (last) revision received: 12 Jan 2014 | accepted: 12 Jan 2014 || publication date(s): online fast track, 9 Feb 2015; in print and online issues, 2 Mar 2015 ||

genes and small sampling sizes at the species level. However, phylogenetic analyses revealed numerous cases for which a clear phylogenetic signal is not obtained at the species level, even for groups that were a priori less complex than TCGs (Morgan & al., 2009; Aguirre-Planter & al., 2012; Sessa & al., 2012; Wan & al., 2013).

Phylogeographic research also revealed many examples of processes that can blur phylogenetic signal. Ancestral polymorphism does exist among closely related species. For instance, in *Gentiana* sect. *Ciminalis* (Adans) Dumort. chloroplast haplotype sharing was found in four species, despite interspecific discrimination by the nuclear ribosomal ITS marker and an obvious morphological divergence (Christe & al., 2014a). Similar chloroplast sharing was seen among closely related species of *Solidago* subsect. *Humiles* (Ridb.) Semple, irrespective of their ploidy level (Peirson & al., 2013) and in *Salix* L. where 53 species display the same haplotype (Percy & al., 2014). This indicates that molecular and morphological rates of divergence might be uncoupled (Vanderpoorten & Shaw, 2010). Conversely, the coexistence of very distantly related haplotypes within single species may also occur due to hybridization (Hassel & al., 2013; Christe & al., 2014a), which might lead to incongruent signals between molecules and the accepted taxonomy, even in "well-curated" species. Finally, phylogeographic studies have highlighted the role of geography, ecology and demography in shaping the genetic diversity within species (Pandey & Rajora, 2012; Temunovic & al., 2012; Hilpold & al., 2014). Consequently, different tree reconstructions may be retrieved depending on which individual or which locus was sampled for phylogenetic analyses at the species level.

Similar issues have been highlighted by DNA barcoding studies. In plants, a consensus was reached for the use of a core of two chloroplast genes (*rbcL*, *matK*) plus the chloroplast spacer *trnH-psbA* (CBOL Plant Working Group, 2009), with ITS being additionally proposed by Hollingsworth & al. (2011). Even though it was observed that "plants are inherently harder to discriminate than animals using DNA barcode" (Fazekas & al., 2009), the global search for more efficient markers continued (Ford & al., 2009; Hollingsworth & al., 2009; Hollingsworth & al., 2011). Despite these concerted efforts, many publications indicate a suboptimal species assignment success with DNA barcodes (*Gossypium* L.: Ashfaq & al., 2013; *Thymus* L.: Federici & al., 2013; *Populus* L.: Feng & al., 2013; *Salvia* L.: Wang, M. & al., 2013; Arundinarieae: Zhang & al., 2012; *Pinus* L.: Hernandez-Leon & al., 2013—see the literature compilation in Naciri & al., 2012). It is evidently impossible to consistently capture species boundaries using a small number of standardized loci. Moreover, it seems that neither a minimum nor a maximum genetic distance among species can be defined to help in disentangling boundaries among closely related species, as shown in *Gentiana* L. using four barcode chloroplast loci (Christe & al., 2014a) or in *Salix* where 53 species share the same haplotype (Percy & al., 2014).

Beside the many controversies that DNA barcoding has provoked, it has had the positive effect of reviving the discussion on species concepts, species boundaries, and species discovery (Wheeler & Meier, 2000; Tautz & al., 2002, 2003; Seberg & al., 2003; Blaxter, 2004; Will & Rubinoff, 2004; Rubinoff, 2006). An increasing consensus now exists that species may be understood to be separately evolving metapopulations, as defined by De Queiroz (2007). This concept defines species as being composed of one or several populations, which share a common history (largely) and have a common future. Contrary to the biological species concept of Mayr (1942), this concept allows for some gene flow between species. It agrees with Templeton's definition (Templeton, 1989) in allowing for species to be distinct phenotypic and ecological entities, but emphasizes the common history of the constituent populations. It does not, contrary to the monophyletic species concept of Baum & Shaw (1995), demand species to be monophyletic. Although this concept is readily understood, it remains difficult to decide how species should be operationally delimited using sequence data.

In this paper, we evaluate the use and abuse of genetic data and the nature of phylogenetic signal at the species rank. We suggest appropriate assumptions for the use of molecular sequence data at the species level. We are particularly interested in how to discover and delimit species, and how to determine the relationships among very recently diverging species. We indicate the ways in which molecular data can mislead if used too simplistically. We are therefore investigating the border between population genetics and phylogeny, in the evolutionary time delimited by the first evidence of species divergence until lineage sorting is complete. Consequently, we are not dealing with the inference of the phylogeny among more distantly related species, or among genera and families.

■ THE SEVEN VEILS THAT OBSCURE SPECIES DELIMITATION AND RELATIONSHIPS

There would be no problem with using sequence data to delimit species if all gene trees matched the species tree. However, it has long been known that gene trees (the phylogenies of individual genes or partitions of the plant genomes) can be discordant with each other and therefore potentially with the species tree (Doyle, 1992; Maddison, 1997). Consequently, species could be delimited and grouped differently depending on which gene tree is utilized. This is referred to as gene tree heterogeneity (Cutter, 2013). We recognize seven processes ("veils") that can influence gene tree heterogeneity: (1) intergenomic transfers, (2) hybridization, (3) incomplete lineage sorting, (4) genome organisation, (5) demography, (6) selection, and (7) phylogeographic structure (Appendix 1).

(1) Intergenomic transfers (NuPt, NuMt). — Nuclear genomes are chimeric objects made of fragments of different origins and histories. This is partly due to intergenomic gene transfer from the plastid or the mitochondrion into the nucleus of the same organism (NuPt: nuclear copies of plastid DNA, NuMt: nuclear copies of mitochondrial DNA). Although transfers of numerous fragments from the chloroplast to the mitochondrion or to the nucleus have been demonstrated (Cummings & al., 2003; Richly & Leister, 2004; Noutsos & al.,

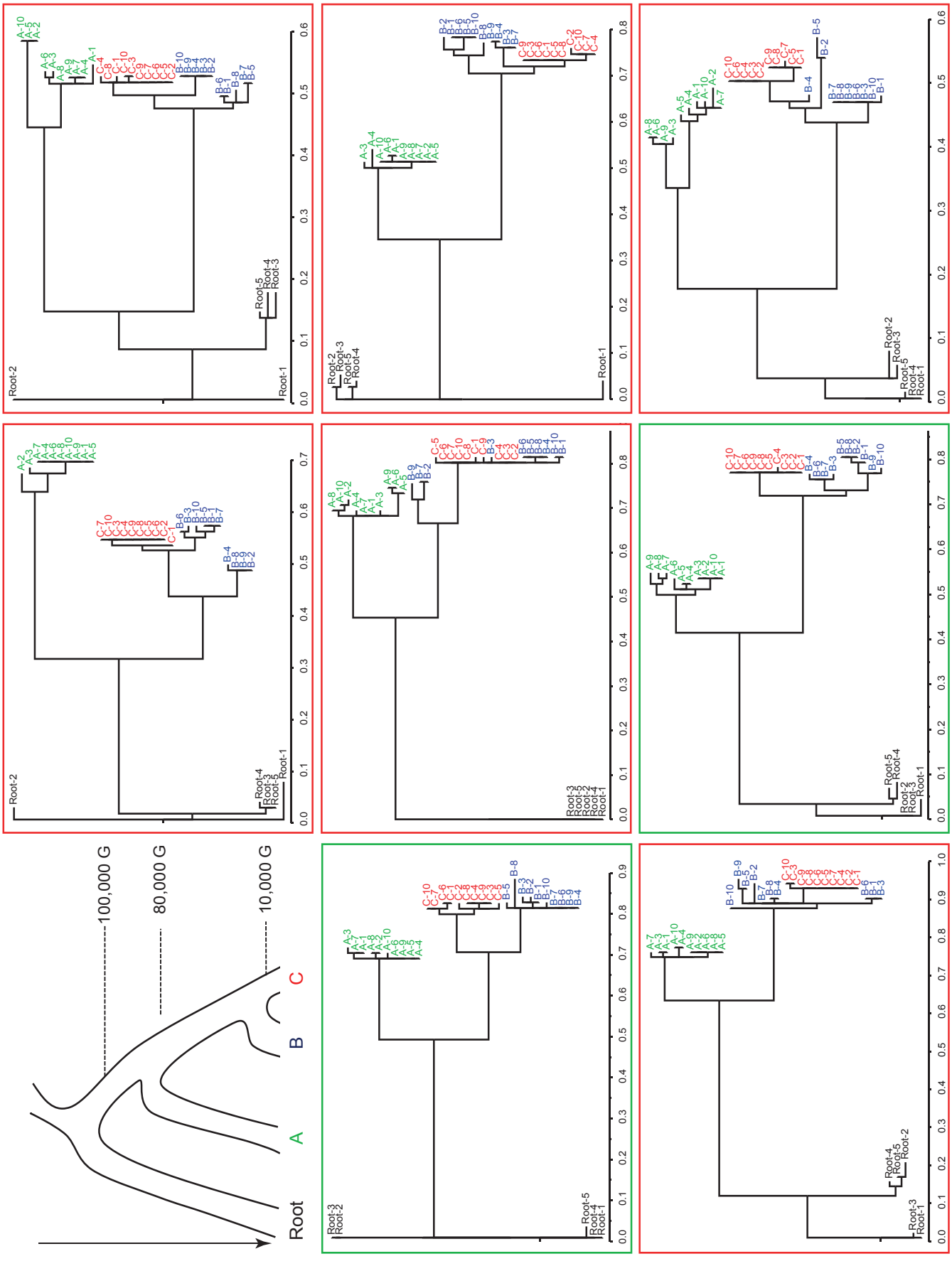
2005; Naciri & Manen, 2010), this process is largely ignored and its effect generally underestimated (Arthofer & al., 2010). Wang & Timmis (2013) showed that NuPts are most often recorded in regions of open chromatin and Wang & al. (2012) suggested that plastid DNA insertions might be favoured by environmental stress. Michalovova & al. (2013) recently demonstrated that the length of the sequence is inversely correlated to the time since the transfer, and suggested that NuPt are preferentially inserted as big pieces near the centromeres and later fragmented by transposable element insertions and redistributed across the genome. Arthofer & al. (2010) identified 34 to 1012 potential NuPt transfers in four model organisms (*Arabidopsis thaliana* (L.) Heynh., *Vitis vinifera* L., *Oryza sativa* L., *Populus trichocarpa* Torr. & A.Gray). A larger study recently confirmed these estimates on 17 plant species, with numbers of NuPts ranging from 38 in *A. thaliana* (L.) Heynh. to 1513 in *Solanum lycopersicum* Lam. (Yoshida & al., 2014). Both studies, moreover, reported high mean length for NuPts (from 0.5 to almost 900 kb). This might render their identification difficult using a long PCR strategy in the absence of other alerting signals such as ghost bands, double peaks or differences in GC contents for genes. In the absence of any other sequence to which a new amplification can be compared, as it can be the case when herbarium specimens of rare species are used, NuPt might not be recognized as such, and may be mistaken as a plastid sequence, therefore leading to false inferences in barcoding studies, as demonstrated by Naciri & Manen (2010). The diverging histories, since the time of transfer, of the targeted sequence and its NuMt or NuPt in terms of mutation rates, recombination and effective sizes may also lead to a biased overestimation of gene heterogeneity that may impact the efficiency of DNA barcoding studies as has been demonstrated on organisms other than plants (Song & al., 2008; Bertheau & al., 2011; Kim & al., 2013). Morevover, Song & al. (2008) showed how undetected NuMts amplifications can be interpreted as evidence of new cryptic species.

(2) Hybridization. — Another mechanism for making the nuclear genome chimeric and for generating gene and genome tree heterogeneity is the horizontal, inter-specific transfer of genetic material through hybridization. Hybridization is one of the most common mechanisms generating reticulated relationships (Doyle, 1992). Hybridization results in the transfer of genetic material between species, leading to the gene trees tracking several speciation histories (Petri & al., 2013). The most extreme case is when the plastid and nuclear genomes of the same species are sister to different species, a phenomenon referred to as plastid capture (Soltis & Kuzoff, 1995). Plastid capture among species results from two important characteristics of organelle genomes: uniparental inheritance in the vast majority of cases, generally maternally through seeds in angiosperms, and paternally through pollen in gymnosperms (Birky, 2008), and their circular DNA that is usually assumed not to recombine, although some intra-molecular recombination might sometimes occur (Kim & al., 2005). When two angiosperm species hybridize, the F1 hybrid obtains its plastid from the maternal, seed parent and half of the nuclear genome from the paternal, pollen parent. Backcrossing to the pollen parent species produces further admixed generations, and eventually the nucleus resembles the pollen parent, whereas the plastid is derived wholly from the original maternal species. This phenomenon could be enhanced when the pollen parent is invading the territory of a seed parent (Currat & al., 2008; Petit & Excoffier, 2009). Plants are known to hybridize frequently (Arnold, 1997) and hybridization may be enhanced by habitat disturbance (Choler & al., 2004), so consequently, plastid capture may not be a rare phenomenon. Recent studies have, moreover, shown that such plastid captures can also occur among species that are sexually incompatible, for example through grafting at root level (Stegemann & al., 2012; Greiner & Bock, 2013). Footprints of plastid capture can be identified when incongruences between chloroplast and nuclear phylogenies are found, and this is common across the angiosperms (e.g., *Satyrium* Sw.: Van der Niet & Linder, 2008; *Machaeranthera* Nees: Morgan & al., 2009; *Ilex* L.: Manen & al., 2010; *Nothofagus* Blume: Acosta & Premoli, 2010; Arundinarieae: Zhang & al., 2012; Brassicaceae: Rešetnik & al., 2013). A further indication of hybridization could be if chloroplast sequence variation is not congruent with the current taxonomy (e.g., *Verbena* L.: Yuan & Olmstead, 2008; *Carapa* Aubl.: Duminil & al., 2012; *Gentiana*: Christe & al., 2014a).

Possibly more common, but much more cryptic, is the evolution of heterogenous nuclear genomes, which contain DNA of two or more species. Hybridization between closely related species is common in some groups such as *Quercus* L. (Lagache & al., 2013), *Salix* (Percy & al., 2014), *Mimulus* Adans, *Silene* L. and *Populus* (Lexer & Widmer, 2008). For example, *Betula nana* L., *B. pubescens* Ehrh. and *B. pendula* Roth hybridize and introgress, despite ploidy differences (Wang, N. & al., 2013). Indeed, RAD-seq data from the British populations shows that genetic differentiation between populations of *B. nana* equals that found between *B. nana* and *B. pubescens*, suggesting that there is as much gene flow between as within species, potentially due to a hybrid zone that moved northwards through Britain after the last glacial maximum. However, the movement of DNA from one species into another through introgression can be quite complex and difficult to assess as the effect of contemporary gene flow can sometimes be confounded with the persistence of ancestral polymorphism. Introgressing alleles can indeed be selected for, or against, in the hybrid. The hybrids may also be very fit and can then persist in competition with the parents (Arnold & al., 2012).

Horizontal gene transfer occurs only rarely between distantly related taxa, for example between Asteraceae and *Gnetum* L. (Won & Renner, 2003), or between a parasitic species and its host as demonstrated for *Cistanche deserticola* Ma (Orobanchaceae) and *Haloxylon ammodendron* Bunge (Chenopodiaceae) by Li & al. (2013). However, such transfers usually do not impact species-level studies, since they are easier to recognize.

(3) Lineage sorting stochasticity. — Incongruent or unresolved relationships among closely related species can also result from incomplete lineage sorting (Posada & Crandall, 2001; Appendix 1). Incomplete lineage sorting occurs when


100,000 G
80,000 G
10,000 G
Root
A
B
C

the diverging species inherit alleles whose genealogy does not reflect the sequence of speciation events (Doyle, 1992; Maddison, 1997). This stochasticity is nicely described by the coalescent (Kingman, 1982, 2000), a theory developed within the framework of the Wright-Fisher model that gives the probability, at each generation back in time, that two given alleles were transmitted by the same parent (an event named coalescence; Appendix 1). In a neutral context, the rate at which gene lineages coalesce in the past only depends on the effective population size (N_e; Appendix 1), with larger populations having, on average, a deeper coalescence (requiring more time) than smaller N_e, with the variance to the most common ancestor (MRCA) being scaled by N_e^2. From the coalescence, it is therefore expected that genes with similar histories may give different realized genealogical trees (Fig. 1). Incomplete lineage sorting has been proposed to explain gene tree incongruence in lichens (Steinova & al., 2013), where a conflict between ITS and the nuclear gene beta-tubulin was reported. Similarly, Zhang & al. (2012) showed that lineage sorting is possibly the most important mechanism for generating disparate phylogenies in the bambusoid clade Arundinarieae.

Lineage sorting can be sometimes difficult to distinguish from hybridization, as both processes leave similar footprints in gene genealogies (Fig. 2). Low or absent sequence divergence also translates into the dominance of lineage sorting, as demonstrated for *Abies* Mill. in Mexico (Aguirre-Planter & al., 2012). Lineage sorting impacts not only species delimitation (and so species discovery) but also species phylogenies. This is particularly the case if coalescence is predicted to need much more time than has elapsed since the relevant speciation event. This is more likely the case in species with large N_e. Figure 1 shows cases of incomplete lineage sorting with a simple scenario of divergence at different time periods, and with equivalent effective sizes among species. For that scenario, reciprocal monophyly for the two most recent species was achieved in only 25% of the simulations. Accordingly, incomplete lineage sorting was shown to be particularly important in speciation events younger than $5N_e$ generations (Jakob & Blattner, 2006). Hudson & Coyne (2002) stated that mitochondrial or chloroplast markers should not be used to delineate species due to their average short coalescence time and that many nuclear genes should be used instead. With this setting, reciprocal monophyly among species is not expected to be attained for 95% of the genes before $9N_e$ to $12N_e$ generations have passed since speciation. Rosenberg (2003) suggested that $5.3N_e$ generations are needed for a given species to acquire monophyly at 99% of its loci given that all loci in its sister species are also monophyletic. For a species with N_e of 1 million individuals and a generation time of 10 years (typical for tree species), this means that full monophyly will only be reached 50 Ma after speciation. Mixtures of species with

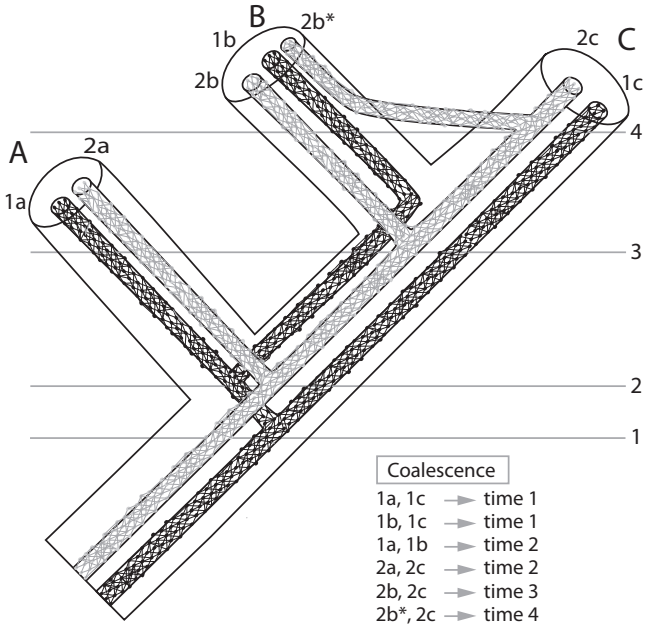


Fig. 2. Coalescence. The hollow tubes indicate the species phylogeny, and two genes (1 and 2) are included. Gene 1 has three orthologous loci: 1a, 1b and 1c in species A, B and C, respectively. The coalescence is dependent on the estimation of N_e at time 1, and estimates the age of divergence between A and C. However, this locus shows incomplete lineage sorting, so that the coalescence of 1b and 1c estimates the speciation between C and A plus B, and not between B and C. Furthermore, it suggests that A and B are more closely related to each other than B and C, and will give a speciation time between B and C of time 1 rather than time 3. Gene 2 represents a locus that has orthologous copies 2a (in species A), 2b in species B, and 2c in species C. Locus 2b* reflects a hybridization event between species C and B, where a copy of locus 2c is transferred from C to B (now labeled 2b*). If 2c is erroneously compared to 2b* (instead of 2b that also exists in the species) to estimate the speciation time between B and C, it will return a too young time (time 4 instead of time 3).

◀ **Fig. 1.** Impact of stochasticity on the coalescence process and on phylogenetic reconstructions. Three speciation events were simulated: the first one occurred 100,000 generations ago and separates the first species (A) from the root. The second event occurred 80,000 generations ago and led to a second species, which, 10,000 generations ago, split into two species (B and C). No migration among species was allowed after speciation. Ten individuals (haploid genomes) were sampled in each species and five for the root. All species have equivalent sizes (N = 10,000), with no changes with time. One hundred simulations were conducted under FastSimCoal v.2.1 (Excoffier & Foll, 2011) and sequence length was set to 1000 bp with a mutation rate of 2×10^{-7} per site per generation with no transversion/transition bias. Simulated sequences were aligned in BioEdit (Hall, 1999) and trees were built using the maximum likelihood program in BioEdit. Trees were then drawn in FigTree v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/). Eight trees were chosen at random to illustrate the stochasticity of lineage sorting. Over the one hundred trees, the time to the most recent common ancestor (TMRCA) ranged from 100,188 to 173,903 generations (mean = 111,188, standard deviation = 12,382), that is always older than the speciation event between the root and the ancestor of the three species A, B and C. Reciprocal monophyly of species B and C was observed in 25% of the trees (in green panels). This is expected for species that have diverged less than ~$5N_e$ ago (Rosenberg, 2003; Degnan & Rosenberg, 2009), as it is the case here.

low and high N_e can be also confusing, meaning that in some species, lineage sorting has been achieved by coalescence, while in closely related species incomplete lineage sorting might remain. This may be the case where widespread species are progenitors of a number of peripheral segregate species, forming so-called paraphyletic taxa (Brookfield, 2011) or in the case of rapid radiations (whether they are ancient or not; Fig. 3; Appendix 1) because they lead to short species tree branches (Degnan & Rosenberg, 2009).

(4) Genome organisation. — N_e is also influenced by genome organisation, or genome structure. This includes the number of chromosomes, the ploidy level, the position of the loci on the chromosomes relative to the centromere or to regions with many genes under selection, sex chromosomes, and other chromosomal regions with high linkage and low recombination rates. Genome organisation can be modified by whole genome duplication, translocations and chromosome fusions (Schneider & Grosschedl, 2007). Genome organisation influences the N_e of each locus. Loci near centromeres have a lower recombination rate than loci far from the centromeres (Mézard, 2006), and consequently have a shorter coalescence time. Loci located in regions with numerous genes under selection can, through hitchhiking effects, have lower N_e than loci in areas with few genes under selection (Cutter, 2013), and so have a shorter coalescence time. In contrast, loci near the tips of the chromosomes and distant from genes under selection could have a high N_e, and so a longer coalescence time.

Organelle genomes are haploid, and so have half the effective number of any chromosome in a diploid nuclear genome. In bisexual species, all individuals transmit the chloroplasts, therefore N_e for a given plastid locus will be half of that of a nuclear diploid locus. Plastids are generally transmitted by only one sex, so for a dioecious species N_e for a plastid locus will be scaled by one quarter. This scaling is obviously affected by the balance in the sex ratios. Sex chromosomes have a very high linkage rate, and also different effective sizes than autosomes (¼N_e or ¾N_e of autosomes depending on how many copies of the sex chromosomes exist (Palumbi & al., 2001). Consequently genes from the different genomes or even different chromosomes in the same genome have different coalescence times (Fig. 4), leading to gene tree heterogeneity (Corl & Ellegren, 2013). In the most bizarre case, a locus transferred from the plastid to the nuclear (NuPt, see above) in a hermaphrodite species will, after its transmission, have twice the coalescence time of its copy on the plastid, while accumulating a higher number of mutations. Palumbi & al. (2001) proposed to use differences in N_e, calling it the "three-times-rule",

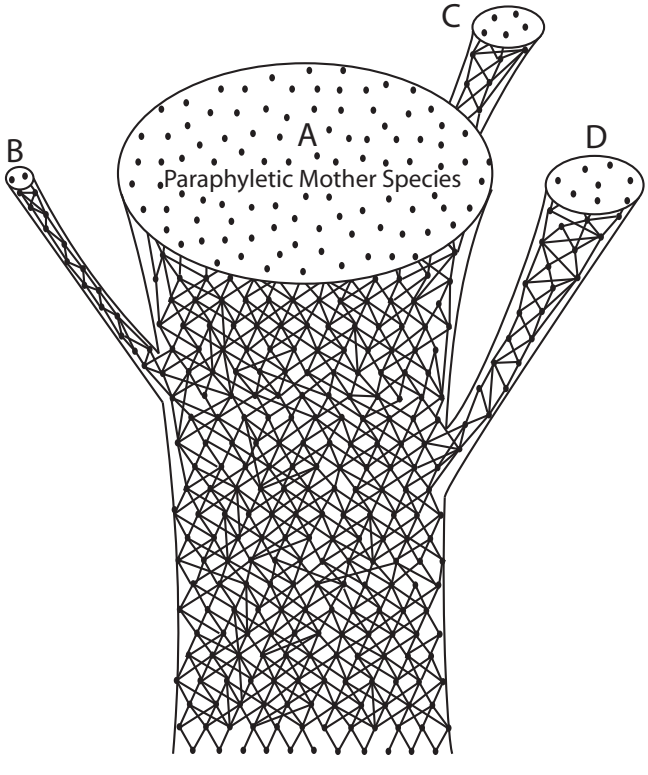


Fig. 3. Paraphyletic "mother species" A and peripheral derivatives, B, C and D. The diameter of the tubes is proportional to N_e. Note that the derivatives have a much smaller N_e than the mother species, consequently the coalescence times within species B, C and D will be much shorter than in A.

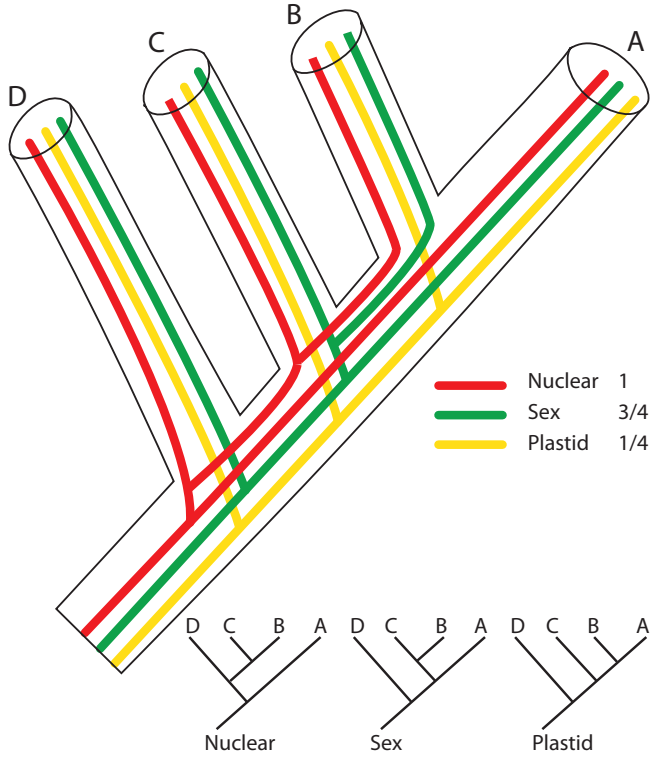


Fig. 4. Lineage sorting, showing the different fates of single nuclear, plastid, and sex chromosome genes for a dioecious species. The plastid gene shows the shortest coalescence time (smallest N_e), and tracks the divergence between A and B. The sex chromosome gene is the next slowest, and tracks divergence between A and C. Due to lineage sorting, the relationships between A, B and C are incorrectly represented. This incorrect inference of relationships due to lineage sorting is even more extreme for the nuclear gene.

and suggested that complete lineage sorting using autosomal chromosomes can be expected when the branch leading to the mitochondrial DNA sequences of a given species is three times longer than the average mtDNA diversity within this species. This rule was rejected by Hudson & Turelli (2003) who argued that stochasticity of the coalescence process can lead to even higher (or lower) differences in realized times to MRCA among genomes.

Polyploidy can impact N_e in different ways. On one hand, duplicating the genome doubles the effective population size N_e, resulting in longer coalescence time and less drift. On the other, a species with several ploidy levels will have, for each ploidy, a different N_e than apparent from a census size of the whole species. If, for instance, a widespread diploid species has one tetraploid population, the census size of tetraploids is not that of the whole species, but only that of the single tetraploid population.

(5) Demography. — The most obvious process that determines the effective population size N_e (and consequently lineage sorting and coalescence time) is the population size and its variation through time. All demographic changes, such as bottlenecks, range expansions, demographic expansions, variation in the reproductive success or unbalanced sex ratios for hermaphrodite species have an influence on N_e and therefore on the time to coalescence. Such events and processes might be responsible for the discrepancy often recorded for N_e versus N_b, the census size of reproductive units within species (Waples & al., 2013).

It is often assumed that at speciation both daughter species have the same N_e. However, this will rarely be the case. Where speciation follows a long-distance dispersal event, or consists of the divergence of a peripheral population, one species may have a much smaller population size than the other, consequently also a much smaller N_e (Cutter, 2013). This is all the more important as peripatric speciation might be a common form of speciation (Vanderpoorten & Shaw, 2010). The smaller segregate will have shorter coalescence times, and higher drift, than the larger sister species. This will lead to different estimates of the time since speciation for the two sisters, and could also result in biased estimates of the common ancestral conditions.

(6) Selection. — Changes in selection intensity is expected to affect N_e, and so the coalescence depth. However, to date contrasting results have been obtained concerning the detectability of selection on genealogies. Some authors showed that selection has no effect on genealogies and on the time to coalescence unless the strength of selection is strong to very strong ($N_e s \gg 1$), depending on the type of selection (purifying or balancing), N_e being the population size and s the selection pressure (Barton & Etheridge, 2004). This is because selection must overwrite the inherent stochasticity of the coalescence process before it can be detected. Other authors demonstrated that selection can have a detectable effect on the way mutations are distributed on the genealogies (Williamson & Orive, 2002) therefore affecting branch length rather than topology in a phylogenetic framework. Walczak & al. (2012) furthermore showed using a modified coalescence approach (the fitness-class coalescent) that purifying selection produces patterns of variation that mimic a population having experienced varying population size changes in the past. Therefore, selection and changes in demography can both leave similar signals on the genealogy of sequences, as already pinpointed for summary statistics that test for sequence neutrality such as Tajima's D or Fu's indices (Tajima, 1983; Fu, 1997; Appendix 1). An extreme case has been described by Percy & al. (2014) who suggested that the extensive chloroplast sharing observed among 53 species is the footprint of a severe selective sweep. In that case, selection erased the pre-existing polymorphism if it ever existed, and covered up species relationships and delimitation

(7) Phylogeographic structure. — In many species a strong spatial genetic structure is observed, and the analysis of such structure forms the foundation of population genetics and phylogeography. Several phenomena contribute to this patterning, which also influences the effective population size N_e in various ways (Waples, 2010). Indeed, a structure within a species can lead to higher N_e for that species than would be expected from the species census size (Waples, 2010; Waples & al., 2013). This is, for instance, the case when subdivision within a species of a given census size allows for the retention of rare alleles that can become fixed in different subpopulations due to genetic drift but that would have been lost if the species existed as a single panmictic population of the same census size. However, without additional knowledge such as the location from which individuals were sampled, the effect of population subdivision can be difficult to disentangle from balancing selection and this is also the case for population expansion and selective sweep (Brookfield, 2011).

The process of range expansion can strongly influence genetic composition within species. When expansions occur rapidly and are mediated by successive founder events leading to the establishment of a series of new small populations, selection is relaxed or completely absent at the leading edge populations, which can result in sometimes otherwise rare alleles becoming very widespread and common. This phenomenon is known as surfing (Klopfstein & al., 2006; Excoffier & Ray, 2008) and might have been quite common in the Northern Hemisphere after the last glacial maximum when land became available after the glaciers retreated. These expanding fronts function as a repetitive series of founder effects, and consequently are often genetically impoverished, compared with the much more genetically diverse central areas of the distribution ranges. Furthermore, genetic drift in these expanding populations can also result in genetic differentiation between the expansion zone and the core region. Expansions can consequently result in zones which are relatively homogenous, mimicking selective sweeps, and which differ genetically from each other. At suture zones between previously geographically isolated lineages, genetic exchange can happen, and in refugia, population size reduction may occur. Jaramillo-Correa & al. (2009) reported such patterns for numerous species of North American tree species. Similarly Christe & al. (2014a) suggested that spatial expansions may have occurred for two siliceous *Gentiana* species that showed reduced levels of genetic

variation all over the Alps using four chloroplast markers. This pattern of decreasing genetic diversity with more recent populations is also well illustrated in *Picea abies* (L.) H.Karst. (Tollefsrud & al., 2008). Therefore, delineating sister species that experienced contrasting colonization scenarios may be difficult, due to differing levels of intraspecific diversities with contrasting N_e and structuring.

Consequences of the seven veils. — These seven processes can influence the estimation of phylogenetic relationships and the efficacy of barcodes, but in different ways. We identified three main consequences. The first relates to the impact of comparing several loci from different organisms in assessing their relationships, and indicates whether the different loci specify the same sister relationships. The second refers to the consequences of comparing several loci on the branch length distribution, which is the same as estimating the ages of the divergence events. The third corresponds to the impact of using a given locus from different individuals, populations or regions and therefore tests the reliability of a barcoding approach. The seven processes outlined above can severely blur the signal of divergence among populations and species, and may take place before, during or after speciation. They can profoundly affect our ability to reconstruct the evolutionary history of species. Here we present a thought experiment to illustrate the interaction of these processes.

A species originating as a peripheral off-shoot of a spatially widespread species may experience a bottleneck reducing genetic diversity at speciation and form a small population with short coalescence times, in which drift could play a role. If the new segregate species is successful, the range expands, allowing surfing effects to increase the frequency of some possibly non-adaptive alleles and potentially allowing polyploids to survive. Polyploidy causes changes in the genome organisation, modifying the N_e. This allows the species to go through a period of relaxed selection, as the larger N_e protects suboptimal alleles against removal. If the expanding new species meets other species from the same group, then hybridization is possible (possibly also with polyploidy), increasing gene tree heterogeneity, and adding some captured chloroplasts to the genetic diversity. Accidental amplification of NuPts could also happen and add to the problems in understanding what happened in this species. It is then not surprising that the history of peripheral segregate species relative to their "ancestral" species (which have experienced none of the above adventures) may be difficult to track.

■ IMPACTS OF INVALID PHYLOGENETIC ASSUMPTIONS

Ignoring the assumptions (dichotomous branching, no persistent polymorphism, no lateral gene transfer, and no variation in effective population size) in phylogenetic constructions among closely related taxa could lead to wrong interpretations, particularly affecting barcode-based identifications, species delimitations, species dating and the assessment of the relationships among closely related species.

Specimen identification using sequence data. — Using barcodes for identification requires that species have to be uniquely defined by a limited set of sequence data. Such a unique definition is only possible if the species are older than the coalescent date. This is the case in the relict tree genus *Zelkova* Spach (Christe & al., 2014b), where the three Mediterranean and Eurasian species can be unambiguously distinguished from each other using two chloroplast loci. The low rate of definite identifications between closely related species and the common absence of a barcode gap in plants (Ashfaq & al., 2013) indicates that this is often not the case. This suggests that a full identification of species using barcodes might not be expected for most species. However, within regional floras (which include only a small proportion of closely related species) identification rates can be much higher, for example the 75%–100% success rate for African rainforest trees (Parmentier & al., 2013) or for the regionally restricted flora of the Kruger National Park (Lahaye & al., 2008).

Species delimitation. — It is evident from the phylogenetic species definition used here that there is no simple protocol by which species can be delimited or "discovered". We prefer "delimited" since species are always falsifiable hypotheses, which are diagnosed against other species. Therefore, species are not absolutes waiting to be "discovered".

Heterogeneous gene trees suggest that species could be compared to languages. Languages assimilate words from other languages (= lateral transfer of genes). As in introgressed genes, these words stay in the language mostly for a limited time. Languages evolve through time. Languages develop regional dialects (= phylogeographical pattern), some of which can diverge to form new languages. Languages can be seen as boxes with holes through which bits from other languages can enter, similar to foreign genes entering species. This model retains the evolving metapopulation concept of De Queiroz (2007), but not the genetically isolated species concept of Mayr (1942).

We advocate a flexible definition of species: species are entities which can incorporate DNA from other species, which can undergo dynamic genetic restructuring, change in genetic variability, donate DNA to other species, even assimilate plastids from other species, but still stay the same recognizable morphological and ecological entity.

Phylogenetic relationships in species clusters. — The more closely related a group of species is, the more difficult it might be to infer the species tree from the collection of gene trees, as shown on Fig. 1. Any single gene tree might give an apparently robust solution, but there may be conflict among the individual gene trees. Problems due to incomplete lineage sorting are likely to occur when species diverged less than $5N_e$ generations ago (Rosenberg, 2003; Degnan & Rosenberg, 2009; Rosenberg & Degnan, 2010) which can represent a long span of time for species with high census sizes. For example, in the European red beech *Fagus sylvatica* L. this could be several millions of years, whereas it could be less than 1000 years for an annual plant species restricted to a single small population. This situation would be particularly difficult if there is variation in the mean coalescence time among a set of closely related

species, such as may be found in paraphyletic species resulting from peripheral speciation. This can be exacerbated by hybridization among sympatric species obscuring the phylogenetic pattern. This situation was nicely illustrated in *Armeria* Willd. (Plumbaginaceae) by Nieto Feliner & al. (2004) or in *Pinus* by Hernandez-Leon & al. (2013).

Species dating. — Estimating the age of a species is reliant on estimating the coalescence age, but this will only be valid if the speciation event and coalescence time are largely similar (but see Fig. 1). This applies to old events (where the time difference between coalescence and speciation is negligible compared to the time since speciation), but not to recent events. For recent events, the coalescent age will also depend on the locus and genome used, as these can have different N_e. Species should not be dated by the divergence from each other, since this assumes that they are reciprocally monophyletic, an assumption which is probably false in most cases. In the absence of hybridization, the youngest coalescent among species might be the closest approach to the correct speciation age of those species. However, if hybridization occurred subsequent to speciation, the ages indicated by the different loci would need to be treated with much more circumspection since they could be the result of horizontal transfer.

■ POSSIBLE SOLUTIONS

Several approaches exist that can reveal problems in the genetic data, and avoid misleading signals.

Recommendation 1. — Each species should be represented by several individuals that cover the differing ploidy levels (when they exist), as well as the geographic and ecological ranges of the species (Maddison & Knowles, 2006; Knowles & Kubatko, 2010; Corl & Ellegren, 2013). Both the core regions, as well as the margins of the species should be sampled. It is important to sample as many populations as possible, as this affects the coalescent, especially when geographical structure is significant or is suspected to be so (Cutter, 2013). The benefits of a comprehensive sampling for disentangling the history of closely related *Pinus* species in a phylogenetic framework were nicely illustrated by Flores-Renteria & al. (2013). With next-generation sequencing techniques (NGS) the cost of individual sequences has become so cheap that the main difficulty remains sampling numerous individuals per species.

Recommendation 2. — For each sample, data should be taken from as many genomes and loci as possible. Within each genome, and especially within the nuclear genome, different loci, experiencing different rates of selection and recombination (Maddison & Knowles, 2006; Heled & Drummond, 2010), should be sampled. This is now possible as NGS techniques give new opportunities to analyse many specimens at many nuclear loci at affordable costs (Zimmer & Wen, 2013). McCormack & al. (2013) recently published a review of the main techniques that can be successfully applied to phylogenetics and phylogeography. In addition to gene information, morphological, breeding system, and genomic architectural information should also be used, for several reasons: (1) these parameters give information on each other (e.g., ploidy level on the expected variation among low-copy nuclear genes); (2) for species delimitation divergence, only one set of characters is insufficient; (3) gene tree heterogeneity, as defined by Cutter (2013), can only be resolved by the separate analysis of many genes or loci.

Recommendation 3. — All loci should be treated as separate data, since concatenation is more likely to result in an incorrect tree species (Degnan & Rosenberg, 2009). The null assumption should be that the gene trees are heterogeneous due to the inherent stochasticity of the coalescence process. Keeping them separate allows avoiding reconstructing false, albeit robust, trees from concatenated data as demonstrated by Kubatko & Degnan (2007). It furthermore allows a much more detailed population genetic analysis and helps to explore the biological processes that drive divergence such as selection or gene flow. Moreover, studying incongruences between nuclear, chloroplast and possibly mitochondrial phylogenies gives insights into hybridization and chloroplast and/or mitochondrial capture events, and helps identify NuPts or NuMts, which might blur the genetic signal. It also allows researchers to select loci appropriate for particular problems, such as those with high intraspecific gene-flow (thus high connectivity within species) to delimit species (Petit & Excoffier, 2009; Naciri & al., 2012).

Recommendation 4. — A coalescence framework should be used whenever studying closely related or sister species (Knowles & Kubatko, 2010). This method overcomes several of the problems associated with hierarchical tree building and incorporates demographic and effective population size parameters and their effects on tree topologies and heterogeneity. Several software packages are now available that use the multispecies coalescent framework (MSC; Appendix 1) formulized by Yang & Rannala (2010). These include *BEAST (Heled & Drummond, 2010) and DISSECT (Jones & al., 2014). *BEAST was successfully used in *Silene* to recognise a new species within *S.* sect. *Cryptoneurae* Aydin & Oxelman (Aydin & al., 2014). It was also used within subtribe Leucanthemopsidinae (Asteraceae) to infer the interspecific phylogenetic relationships (Tomasello & al., 2015). DISSECT was used to delimit species in the *Silene aegyptiaca* complex (Aydin, 2014). Another possibility is to use networks, as advised by many authors (Posada & Crandall, 2001; Corl & Ellegren, 2013), to keep trace of reticulation events. Network topologies are shaped by demography and can be also interpreted in a coalescence framework. Moreover, new methods such as filtered supernetworks can be used to distinguish hybridization from lineage sorting (Holland & al., 2008).

■ CONCLUSION

Species are "natural" entities, but there is some arbitrariness in their ranking. This is best accommodated by approaches integrating all available data. In this context, molecular data are very important for interpreting species limits and relationships.

Due to the large number of characters, and the high level of resolution, it is often possible to distinguish populations or groups of populations (phylogeographic patterns), a level of resolution not achieved by other (e.g., morphology) datasets. However, these very rich data can also thoroughly mislead. It is important to analyse these data using the appropriate methods and assumptions, and this can be achieved only by merging population genetics and phylogenetic approaches. Fortunately, coalescence methods are suitable for both and are becoming increasingly available as a series of newly developed software that estimate population parameters such as the effective population size or the population growth rate (Kuhner, 2009).

■ ACKNOWLEDGEMENTS

We wish to thank the University of Zurich and the Conservatoire et Jardin botaniques de la Ville de Genève for funding this research. We also thank Mathias Currat, Richard Bateman and an additional reviewer for their valuable comments on the manuscript. Many thanks also to Melanie Ranft for drawing the figures.

■ LITERATURE CITED

Acosta, M.C. & Premoli, A.C. 2010. Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Molec. Phylogen. Evol.* 54: 235–242. http://dx.doi.org/10.1016/j.ympev.2009.08.008

Aguirre-Planter, E., Jaramillo-Correa, J.P., Gomez-Acevedo, S., Khasa, D.P., Bousquet, J. & Eguiarte, L.E. 2012. Phylogeny, diversification rates and species boundaries of Mesoamerican firs (*Abies*, Pinaceae) in a genus-wide context. *Molec. Phylogen. Evol.* 62: 263–274. http://dx.doi.org/10.1016/j.ympev.2011.09.021

Arnold, M.L. 1997. *Natural hybridization and evolution*. New York: Oxford University Press.

Arnold, M.L., Ballerini, E.S. & Brothers, A.N. 2012. Hybrid fitness, adaptation and evolutionary diversification: Lessons learned from Louisiana Irises. *Heredity* 108: 159–166. http://dx.doi.org/10.1038/hdy.2011.65

Arthofer, W., Schüler, S., Steiner, F. & Schlick-Steiner, B.C. 2010. Chloroplast DNA-based studies in molecular ecology may be compromised by nuclear-encoded plastid sequence. *Molec. Ecol.* 19: 3853–3856. http://dx.doi.org/10.1111/j.1365-294X.2010.04787.x

Ashfaq, M., Asif, M., Anjum, Z.I. & Zafar, Y. 2013. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). *Molec. Ecol. Resources* 13: 573–582. http://dx.doi.org/10.1111/1755-0998.12089

Aydin, Z. 2014. *Species delimitation and phylogenetic relationships: A study of* Silene *sections* Atocion *and* Cryptoneurae. Ph.D. thesis, University of Gothenburg, Sweden.

Aydin, Z., Marcussen, T., Ertekin, A.S. & Oxelman, B. 2014. Marginal likelihood estimate comparisons to obtain optimal species delimitations in *Silene* sect. *Cryptoneurae* (Caryophyllaceae). *PLOS ONE* 9: e106990. http://dx.doi.org/10.1371/journal.pone.0106990

Bardy, K.E., Schönswetter, P., Schneeweiss, G.M., Fischer, M.A. & Albach, D.C. 2011. Extensive gene flow blurs species boundaries among *Veronica barrelieri*, *V. orchidea* and *V. spicata* (Plantaginaceae) in southeastern Europe. *Taxon* 60: 108–121.

Barton, N.H. & Etheridge, A.M. 2004. The effect of selection on genealogies. *Genetics* 166: 1115–1131. http://dx.doi.org/10.1534/genetics.166.2.1115

Baum, D.A. & Shaw, K.L. 1995. Genealogical perspectives on the species problem. Pp. 289–303 in: Hoch, P.C. & Stephenson, A.G. (eds.), *Experimental and molecular approaches to plant biosystematics*. St. Louis: Missouri Botanical Garden.

Bertheau, C., Schuler, H., Krumböck, S., Arthofer, W. & Stauffer, C. 2011. Hit or miss in phylogeographic analyses: The case of the cryptic NUMTs. *Molec. Ecol. Resources* 11: 1056–1059. http://dx.doi.org/10.1111/j.1755-0998.2011.03050.x

Birky, J.C.W. 2008. Uniparental inheritance of organelle genes. *Curr. Biol.* 18: R692–R695.

Blaxter, M.L. 2004. The promise of a DNA taxonomy. *Philos. Trans., Ser. B* 359: 669–679. http://dx.doi.org/10.1098/rstb.2003.1447

Brookfield, J. 2011. Coalescence: The sharing of ancestry of alleles. *eLS*. http://dx.doi.org/10.1002/9780470015902.a0001775.pub2

CBOL Plant Working Group 2009. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106: 12794–12797. http://dx.doi.org/10.1073/pnas.0905845106

Choler, P., Erschbamer, B., Tribsch, A., Gielly, L. & Taberlet, P. 2004. Genetic introgression as a potential to widen a species' niche: Insights from alpine *Carex curvula*. *Proc. Natl. Acad. Sci. U.S.A.* 101: 171–176. http://dx.doi.org/10.1073/pnas.2237235100

Christe, C., Caetano, S., Aeschimann, D., Kropf, M., Diadema, K. & Naciri, Y. 2014a. The intraspecific genetic variability of siliceous and calcareous *Gentiana* species is shaped by contrasting demographic and re-colonization processes. *Molec. Phylogen. Evol.* 70: 323–336. http://dx.doi.org/10.1016/j.ympev.2013.09.022

Christe, C., Kozlowski, G., Frey, D., Bétrisey, S., Maharramova, E., Garfì, G., Pirintsos, S. & Naciri, Y. 2014b. Footprints of past intensive diversification and structuring for the genus *Zelkova* (Ulmaceae) in south-western Eurasia. *J. Biogeogr.* 41: 1081–1093. http://dx.doi.org/10.1111/jbi.12276

Corl, A. & Ellegren, H. 2013. Sampling strategies for species trees: The effects on phylogenetic inference of the number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. *Molec. Phylogen. Evol.* 67: 358–366. http://dx.doi.org/10.1016/j.ympev.2013.02.002

Cummings, M.P., Nugent, J.M., Olmstead, R.G. & Palmer, J.D. 2003. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms. *Curr. Genetics* 43: 131–138.

Currat, M.M., Ruedi, M.M., Excoffier, L.L. & Petit, R.J.R. 2008. The hidden side of invasions: Massive introgression by local genes. *Evolution* 62: 1908–1920.

Cutter, A.D. 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molec. Phylogen. Evol.* 69: 1172–1185. http://dx.doi.org/10.1016/j.ympev.2013.06.006

De Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56: 879–886. http://dx.doi.org/10.1080/10635150701701083

Degnan, J.H. & Rosenberg, N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescence. *Trends Ecol. Evol.* 24: 332–340. http://dx.doi.org/10.1016/j.tree.2009.01.009

Doyle, J.J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Syst. Bot.* 17: 144–163. http://dx.doi.org/10.2307/2419070

Duminil, J., Kenfack, D., Viscos, V., Grumiau, L. & Hardy, O.J. 2012. Testing species delimitation in sympatric species complexes: The case of an African tropical tree, *Carapa* spp. (Meliaceae). *Molec. Phylogen. Evol.* 62: 275–285. http://dx.doi.org/10.1016/j.ympev.2011.09.020

Excoffier, L. & Foll, M. 2011. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27: 1332–1334. http://dx.doi.org/10.1093/bioinformatics/btr124

Excoffier, L. & Ray, N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* 23: 347–351. http://dx.doi.org/10.1016/j.tree.2008.04.004

Fazekas, A.J., Kesankurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C., Newmaster, S.G., Hajibabei, M. & Husband, B.C. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molec. Ecol. Resources* 9: 130–139. http://dx.doi.org/10.1111/j.1755-0998.2009.02652.x

Federici, S., Galimberti, A., Bartolucci, F., Bruni, I., De Mattia, F., Cortis, P. & Labra, M. 2013. DNA barcoding to analyse taxonomically complex groups in plants: the case of *Thymus* (Lamiaceae). *Bot. J. Linn. Soc.* 171: 687–699. http://dx.doi.org/10.1111/boj.12034

Feng, J., Jiang, D., Shang, H., Dong, M., Wang, G., He, X., Zhao, C. & Mao, K. 2013. Barcoding poplars (*Populus* L.) from western China. *PLOS ONE* 8: e71710. http://dx.doi.org/10.1371/journal.pone.0071710

Flores-Renteria, L., Wegier, A., Ortega Del Vecchyo, D., Ortiz-Medrano, A., Pinero, D., Whipple, A.V., Molina-Freaner, F. & Dominguez, C.A. 2013. Genetic, morphological, geographical and ecological approaches reveal phylogenetic relationships in complex groups, an example of recently diverged pinyon pine species (Subsection *Cembroides*). *Molec. Phylogen. Evol.* 69: 940–949. http://dx.doi.org/10.1016/j.ympev.2013.06.010

Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Stahl, J.V., Kelly, L.J., Wikstrom, N., Hollingsworth, P.M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W. & Wilkinson, M.J. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. *Bot. J. Linn. Soc.* 159: 1–11. http://dx.doi.org/10.1111/j.1095-8339.2008.00938.x

Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.

Greiner, S. & Bock, R. 2013. Tuning a menage a trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* 35: 354–365. http://dx.doi.org/10.1002/bies.201200137

Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41: 95–98.

Hassel, K., Segreto, R. & Ekrem, T. 2013. Restricted variation in plant barcoding markers limits identification in closely related bryophyte species. *Molec. Ecol. Resources* 13: 1047–1057. http://dx.doi.org/10.1111/1755-0998.12074

Heled, J. & Drummond, A.J. 2010. Bayesian inference of species trees from multilocus data. *Molec. Biol. Evol.* 27: 570–580. http://dx.doi.org/10.1093/molbev/msp274

Hernandez-Leon, S., Gernandt, D.S., Perez de la Rosa, J.A. & Jardon-Barbolla, L. 2013. Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae* inferrred from plastid DNA. *PLoS ONE* 8: e70501. http://dx.doi.org/10.1371/journal.pone.0070501

Hilpold, A., Vilatersana, R., Susanna, A., Meseguer, A.S., Boršic´, I., Constantinidis, T., Filigheddu, R., Romaschenko, K., Suárez-Santiago, V.N., Tugay, O., Uysal, T., Pfeil, B.E. & Garcia-Jacas, N. 2014. Phylogeny of the *Centaurea* group (*Centaurea*, Compositae): Geography is a better predictor than morphology. *Molec. Phylogen. Evol.* 77: 195–215. http://dx.doi.org/10.1016/j.ympev.2014.04.022

Holland, B.R., Benthin, S., Lockhart, P.J., Moulton, V. & Huber, K.T. 2008. Using supernetworks to distinguish hybridization from lineage-sorting. *B. M. C. Evol. Biol.* 8: 202. http://dx.doi.org/10.1186/1471-2148-8-202

Hollingsworth, M.L., Andra Clark, A., Forrest, L.L., Richardson, J., Pennington, R.T., Long, D.G., Cowan, R., Chase, M.W., Gaudeul, M. & Hollingsworth, P.M. 2009. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molec. Ecol. Resources* 9: 439–457. http://dx.doi.org/10.1111/j.1755-0998.2008.02439.x

Hollingsworth, P.M., Graham, S.W. & Little, D.P. 2011. Choosing and using a plant DNA barcode. *PLOS ONE* 6: e19254. http://dx.doi.org/10.1371/journal.pone.0019254

Hudson, R.R. & Coyne, J.A. 2002. Mathematical consequences of the genealogical species concept. *Evolution* 56: 1557–1565. http://dx.doi.org/10.1111/j.0014-3820.2002.tb01467.x

Hudson, R.R. & Turelli, M. 2003. Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* 57: 182–190.

Jakob, S.S. & Blattner, F.R. 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. http://dx.doi.org/10.1093/molbev/msl018

Jaramillo-Correa, J.P., Beaulieu, J., Khasa, D.P. & Bousquet, J. 2009. Inferring the past from the present phylogeographic structure of North American forest trees: Seeing the forest for the genes. *Canad. J. Forest Res.* 39: 286–307. http://dx.doi.org/10.1139/X08-181

Jones, G., Aydin, Z. & Oxelman, B. 2014. DISSECT: An assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*. http://dx.doi.org/10.1093/bioinformatics/btu770

Kim, K.-J., Choi, K.-S. & Jansen, R.K. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molec. Biol. Evol.* 22: 1783–1792. http://dx.doi.org/10.1093/molbev/msi174

Kim, S.-J., Lee, K.Y. & Ju, S.-J. 2013. Nuclear mitochondrial pseudogenes in *Austinograea alayseae* hydrothermal vent crabs (Crustacea: Bythograeidae): Effects on DNA barcoding. *Molec. Ecol. Resources* 13: 781–787. http://dx.doi.org/10.1111/1755-0998.12119

Kingman, J.F.C. 1982. The coalescent. *Stochastic Processes Applic.* 13: 235–248. http://dx.doi.org/10.1016/0304-4149(82)90011-4

Kingman, J.F.C. 2000. Origins of the coalescent: 1974–1982. *Genetics* 156: 1461–1463.

Klopfstein, S., Currat, M. & Excoffier, L. 2006. The fate of mutations surfing on the wave of a range expansion. *Molec. Biol. Evol.* 23: 482–490. http://dx.doi.org/10.1093/molbev/msj057

Knowles, L.L. & Kubatko, L.S. 2010. *Estimating species trees: Practical and theoretical aspects*. Hoboken: Wiley.

Kubatko, L.S. & Degnan, J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56: 17–24. http://dx.doi.org/10.1080/10635150601146041

Kuhner, M.K. 2009. Coalescent genealogy samplers: Windows into population history. *Trends Ecol. Evol.* 24: 86–93. http://dx.doi.org/10.1016/j.tree.2008.09.007

Lagache, L., Leger, J.-B., Daudin, J.-J., Petit, R.J. & Vacher, C. 2013. Putting the biological species concept to the test: Using mating networks to delimit species. *PLOS ONE* 8: e68267. http://dx.doi.org/10.1371/journal.pone.0068267

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G. & Savolainen, V. 2008. DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 105: 2923–2928. http://dx.doi.org/10.1073/pnas.0709936105

Lexer, C. & Widmer, A. 2008. The genic view of plant speciation: Recent progress and emerging questions. *Philos. Trans., Ser. B* 363: 3023–3036. http://dx.doi.org/10.1098/rstb.2008.0078

Li, X., Zhang, T.-C., Qiao, Q., Ren, Z., Zhao, J., Yonezawa, T., Hasegawa, M., Crabbe, M.J.C., Li, J. & Zhong, Y. 2013. Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PLOS ONE* 8: e58747. http://dx.doi.org/10.1371/journal.pone.0058747

Maddison, W.P. 1997. Gene trees in species trees. *Syst. Biol.* 46: 523–536. http://dx.doi.org/10.1093/sysbio/46.3.523

Maddison, W.P. & Knowles, L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* 55: 21–30. http://dx.doi.org/10.1080/10635150500354928

Manen, J.-F., Barriera, G., Loizeau, P.-A. & Naciri, Y. 2010. The history of extant *Ilex* species (Aquifoliaceae): Evidence of hybridization within a Miocene radiation. *Molec. Phylogen. Evol.* 57: 961–977. http://dx.doi.org/10.1016/j.ympev.2010.09.006

Mayr, E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C. & Brumfield, R.T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molec. Phylogen. Evol.* 66: 526–538. http://dx.doi.org/10.1016/j.ympev.2011.12.007

Mézard, C. 2006. Meiotic recombination hotspots in plants. *Trans. Biochem. Soc.* 34: 531–534. http://dx.doi.org/10.1042/BST0340531

Michalovova, M., Vyskot, B. & Kejnovsky, E. 2013. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: Size, relative age and chromosomal localization. *Heredity* 111: 314–320. http://dx.doi.org/10.1038/hdy.2013.51

Morgan, D.R., Korn, R.L. & Mugleston, S.L. 2009. Insights into reticulate evolution in Machaerantherinae (Asteraceae: Astereae): 5S ribosomal RNA spacer variation, estimating support for incongruence, and constructing reticulate phylogenies. *Amer. J. Bot.* 96: 920–932. http://dx.doi.org/10.3732/ajb.0800308

Naciri, Y. & Manen, J.-F. 2010. Potential DNA transfer from the chloroplast to the nucleus in *Eryngium alpinum* L. (Apiaceae). *Molec. Ecol. Resources* 10: 728–731. http://dx.doi.org/10.1111/j.1755-0998.2009.02816.x

Naciri, Y., Caetano, S. & Salamin, N. 2012. Plant DNA barcodes and the influence of gene flow. *Molec. Ecol. Resources* 12: 575–580. http://dx.doi.org/10.1111/j.1755-0998.2012.03130.x

Nieto Feliner, G., Larena, B.G. & Aguilar, J.F. 2004. Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). *Ann. Bot. (Oxford)* 93: 189–200. http://dx.doi.org/10.1093/aob/mch027

Noutsos, C., Richly, E. & Leister, D. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res.* 15: 616–628. http://dx.doi.org/10.1101/gr.3788705

Palumbi, S.R., Cipriano, F. & Hare, M.P. 2001. Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. *Evolution* 55: 859–868. http://dx.doi.org/10.1554/0014-3820(2001)055[0859:PNGCFM]2.0.CO;2

Pandey, M. & Rajora, O.P. 2012. Genetic diversity and differentiation of core vs. peripheral populations of eastern white cedar, *Thuja occidentalis* (Cupressaceae) *Amer. J. Bot.* 99: 690–699. http://dx.doi.org/10.3732/ajb.1100116

Parmentier, I., Duminil, J., Kuzmina, M., Philippe, M., Thomas, D.W., Kenfack, D., Chuyong, G.B., Cruaud, C. & Hardy, O.J. 2013. How effective are DNA barcodes in the identification of African rainforest trees? *PLOS ONE* 8: e54921. http://dx.doi.org/10.1371/journal.pone.0054921

Peirson, J.A., Dick, C.W. & Reznicek, A.A. 2013. Phylogeography and polyploid evolution of North American goldenrods (*Solidago* subsect. *Humiles*, Asteraceae). *J. Biogeogr.* 40: 1887–1898. http://dx.doi.org/10.1111/jbi.12136

Percy, D.M., Argus, G.W., Cronk, Q.C., Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Husband, B.C., Newmaster, S.G., Barrett, S.C.H. & Graham, S.W. 2014. Understanding the spectacular failure of DNA barcoding in willows (*Salix*): Does this result from a trans-specific selective sweep? *Molec. Ecol.* 19: 4737–4756. http://dx.doi.org/10.1111/mec.12837

Petit, R.J. & Excoffier, L. 2009. Gene flow and species delimitation. *Trends Ecol. Evol.* 24: 386–393. http://dx.doi.org/10.1016/j.tree.2009.02.011

Petri, A., Pfeil, B.E. & Oxelman, B. 2013. Introgressive hybridization between anciently diverged lineages of *Silene* (Caryophyllaceae). *PLOS ONE* 8: e67729. http://dx.doi.org/10.1371/journal.pone.0067729

Posada, D. & Crandall, K.A. 2001. Intraspecific gene genealogies: Trees grafting into networks. *Trends Ecol. Evol.* 16: 37–45. http://dx.doi.org/10.1016/S0169-5347(00)02026-7

Rešetnik, I., Satovic, Z., Schneeweiss, G.M. & Liber, Z. 2013. Phylogenetic relationships in Brassicaceae tribe Alysseae inferred from nuclear ribosomal and chloroplast DNA sequence data. *Molec. Phylogen. Evol.* 69: 772–786. http://dx.doi.org/10.1016/j.ympev.2013.06.026

Richly, E. & Leister, D. 2004. NUPT in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Molec. Biol. Evol.* 21: 1972–1980. http://dx.doi.org/10.1093/molbev/msh210

Rosenberg, N.A. 2003. The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescence model. *Evolution* 57: 1465–1477. http://dx.doi.org/10.1111/j.0014-3820.2003.tb00355.x

Rosenberg, N.A. & Degnan, J.H. 2010. Coalescent histories for discordant gene trees and species trees. *Theor. Populat. Biol.* 77: 145–151. http://dx.doi.org/10.1016/j.tpb.2009.12.004

Rubinoff, D. 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biol.* 20: 1026–1033. http://dx.doi.org/10.1111/j.1523-1739.2006.00372.x

Schneider, R. & Grosschedl, R. 2007. Dynamics and interplay of nuclear architecture, genome organization and gene expression. *Genes & Developm.* 21: 3027–3043. http://dx.doi.org/10.1101/gad.1604607

Seberg, O., Humphries, C.J., Knapp, S., Stevenson, D.W., Petersen, G., Scharff, N. & Andersen, N.M. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18: 63–65. http://dx.doi.org/10.1016/S0169-5347(02)00059-9

Sessa, E.B., Zimmer, E.A. & Givnish, T.J. 2012. Reticulate evolution on a global scale: A nuclear phylogeny for New World *Dryopteris* (Dryopteridaceae). *Molec. Phylogen. Evol.* 64: 563–581. http://dx.doi.org/10.1016/j.ympev.2012.05.009

Soltis, D.E. & Kuzoff, R.K. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49: 727–742. http://dx.doi.org/10.2307/2410326

Song, H., Buhay, J.E., Whiting, M.F. & Crandall, K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci. U.S.A.* 105: 13486–13491. http://dx.doi.org/10.1073/pnas.0803076105

Stegemann, S.S., Keuthe, M.M., Greiner, S.S. & Bock, R.R. 2012. Horizontal transfer of chloroplast genomes between plant species. *Proc. Natl. Acad. Sci. U.S.A.* 109: 2434–2438. http://dx.doi.org/10.1073/pnas.1114076109

Steinova, J., Stenroos, S., Grube, M. & Skaloud, P. 2013. Genetic diversity and species delimitation of the zeorin-containing red-fruited *Cladonia* species (lichenized Ascomycota) assessed with ITS rDNA and beta-tubulin data. *Lichenologist* 45: 665–684. http://dx.doi.org/10.1017/S0024282913000297

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. 2002. DNA points the way ahead of taxonomy. *Nature* 418: 479. http://dx.doi.org/10.1038/418479a

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.* 18: 70–74. http://dx.doi.org/10.1016/S0169-5347(02)00041-1

Templeton, A.R. 1989. The meaning of species and speciation: A genetic perspective. Pp. 3–27 in: Otte, D. & Endler, J.A. (eds.), *Speciation and its consequences*. Sunderland, Massachusetts: Sinauer.

Temunovic, M., Franjic, J., Satovic, Z., Grgurev, M., Frascaria-Lacoste, N. & Fernandez-Manjarre, J.F. 2012. Environmental

heterogeneity explains the genetic structure of continental and Mediterranean populations of *Fraxinus angustifolia* Vahl. *PLOS ONE* 7: e42764. http://dx.doi.org/10.1371/journal.pone.0042764

Tollefsrud, M.M., Kissling, R., Gugerli, F., Johnsen, Ø., Skrøppa, T., Cheddadi, R., Van der Knaap, W.O., Latałowa, M., Terhürne-Berson, R., Litt, T., Geburek, T., Brochmann, C. & Sperisen, C. 2008. Genetic consequences of glacial survival and postglacial colonization in Norway spruce: Combined analysis of mitochondrial DNA and fossil pollen. *Molec. Ecol.* 17: 4134–4150. http://dx.doi.org/10.1111/j.1365-294X.2008.03893.x

Tomasello, S., Álvarez, I., Vargas, P. & Oberprieler, C. 2015. Is the extremely rare Iberian endemic plant species *Castrilanthemum* debeauxii (Compositae, Anthemideae) a 'living fossil'? Evidence from a multi-locus species tree reconstruction. *Molec. Phylogen. Evol.* 82: 118–130. http://dx.doi.org/10.1016/j.ympev.2014.09.007

Van der Niet, T. & Linder, H.P. 2008. Dealing with incongruence in the quest for the species tree: A case study from the orchid genus *Satyrium*. *Molec. Phylogen. Evol.* 47: 154–174. http://dx.doi.org/10.1016/j.ympev.2007.12.008

Vanderpoorten, A. & Shaw, A.J. 2010. The application of molecular data to the phylogenetic delimitation of species in bryophytes: A note of caution. *Phytotaxa* 9: 229–237. http://dx.doi.org/10.11646/phytotaxa.9.1.12

Walczak, A.M., Nicolaisen, L.E., Plotkin, J.B. & Desai, M.M. 2012. The structure of genealogies in the presence of purifying selection: A fitness-class coalescent. *Genetics* 190: 753–779. http://dx.doi.org/10.1534/genetics.111.134544

Wan, Y., Schwaninger, H.R., Baldo, A.M., Labate, J.A., Zhong, G.-Y. & Simon, C.J. 2013. A phylogenetic analysis of the grape genus (*Vitis* L.) reveals broad reticulation and concurrent diversification during Neogene and Quaternary climate change. *B. M. C. Evol. Biol.* 13: 141. http://dx.doi.org/10.1186/1471-2148-13-141

Wang, D. & Timmis, J.N. 2013. Cytoplasmic organelle DNA preferentially inserts into open chromatin. *Genome Biol. Evol.* 5: 1060–1064. http://dx.doi.org/10.1093/gbe/evt070

Wang, D., Lloyd, A.H. & Timmis, J.N. 2012. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc. Natl. Acad. Sci. U.S.A.* 109: 2444–2448. http://dx.doi.org/10.1073/pnas.1117890109

Wang, M., Zhao, H.X., Wang, L., Wang, T., Yang, R.W., Wang, X.L., Zhou, Y.H., Ding, C.B. & Zhang, L. 2013. Potential use of DNA barcoding for the identification of *Salvia* based on cpDNA and nrDNA sequences. *Gene* 528: 206–215. http://dx.doi.org/10.1016/j.gene.2013.07.009

Wang, N., Thomson, M., Bodles, W.J.A., Crawford, R.M.M., Hunt, H.V., Featherstone, A.W., Pellicer, J. & Buggs, R.J.A. 2013. Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Molec. Ecol.* 22: 3098–3111. http://dx.doi.org/10.1111/mec.12131

Waples, R.S. 2010. Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. *Molec. Ecol. Resources* 10: 785–796. http://dx.doi.org/10.1111/j.1755-0998.2010.02876.x

Waples, R.S., Luikart, G., Faulkner, J.R. & Tallmon, D.A. 2013. Simple life-history traits explain key effective population size ratios across diverse taxa. *Proc. Roy. Soc. London, Ser. B, Biol. Sci.* 280: 20131339. http://dx.doi.org/10.1098/rspb.2013.1339

Wheeler, Q.D. & Meier, R. 2000. *Species concepts and phylogenetics theory: A debate*. New York: Columbia University Press.

Will, K.W. & Rubinoff, D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20: 47–55. http://dx.doi.org/10.1111/j.1096-0031.2003.00008.x

Williamson, S. & Orive, M.E. 2002. The genealogy of a sequence subject to purifying selection at multiple sites. *Molec. Biol. Evol.* 19: 1376–1384. http://dx.doi.org/10.1093/oxfordjournals.molbev.a004199

Won, H. & Renner, S.S. 2003. Horizontal gene transfer from flowering plants to *Gnetum*. *Proc. Natl. Acad. Sci. U.S.A.* 100: 10824–10829. http://dx.doi.org/10.1073/pnas.1833775100

Yang, Z. & Rannala, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 107: 9264–9269. http://dx.doi.org/10.1073/pnas.0913022107

Yoshida, T., Furihata, H.Y. & Kawabe, A. 2014. Patterns of genomic integration of nuclear chloroplast DNA fragments in plant species. *DNA Res.* 21: 127–140. http://dx.doi.org/10.1093/dnares/dst045

Yuan, Y.-W. & Olmstead, R.G. 2008. A species-level phylogenetic study of the *Verbena* complex (Verbenaceae) indicates two independent intergeneric chloroplast transfers. *Molec. Phylogen. Evol.* 48: 23–33. http://dx.doi.org/10.1016/j.ympev.2008.04.004

Zhang, Y.-X., Zeng, C.-X. & Li, D.-Z. 2012. Complex evolution in *Arundinarieae* (Poaceae: Bambusoideae): Incongruence between plastid and nuclear GBSSI gene phylogenies. *Molec. Phylogen. Evol.* 63: 777–797. http://dx.doi.org/10.1016/j.ympev.2012.02.023

Zimmer, E.A. & Wen, J. 2013. Reprint of: Using nuclear gene data for plant phylogenetics: Progress and prospects. *Molec. Phylogen. Evol.* 66: 539–550. http://dx.doi.org/10.1016/j.ympev.2013.01.005

Zozomová-Lihová, J., Marhold, K. & Španiel, S. 2014. Taxonomy and evolutionary history of *Alyssum montanum* (Brassicaceae) and related taxa in southwestern Europe and Morocco: Diversification driven by polyploidy, geographic and ecological isolation. *Taxon* 63: 562–591. http://dx.doi.org/10.12705/633.18

Appendix 1. Glossary.

Coalescence: the merging of two lineages in a single individual at a particular generation back in time.

Fu's Fs: a summary statistic, based on the infinite site model, which detects the effect of natural selection on DNA samples within populations. It calculates the probability of observing a similar allele number as the one observed one given the recorded number of pairwise differences in the population. This statistic assumes population equilibrium and is also sensitive to population demographic changes.

Incomplete lineage sorting: the discordance between gene tree and species tree. It is due to the stochastic segregation of alleles at a polymorphic locus at the time of speciation. By chance only, the genealogy of alleles retained by each species may not equal the species tree.

Monophyletic species: a species for which the analysed genes show complete lineage sorting.

Multispecies coalescent model: a theoretical framework developed by Yang & Rannala (2010) who extended the coalescent theory, initially developed within a population (Kingman, 1982), to multiple populations. The model assumes a strict divergence after speciation (no migration, no hybridization, no gene flow, no horizontal gene transfer). It also assumes that populations are ideal, i.e., following the Wright-Fisher model, with constant sizes, no selection and no overlapping generations.

***N*:** the census size of a population or species. This size can be different from N_b the number of breeding individuals in a given generation or from N_e, the effective population size.

***N*$_e$:** The effective population size corresponds to the size of a theoretical population under the Wright-Fisher model that would have the same genetic diversity as the one recorded in the population. N_e is usually smaller than the census size (N), due to variation in the reproductive success, to unbalanced sex ratios for dioecious species

Appendix 1. Continued.

or to overlapping generations. It can, however, be larger than N, when structuring allows for the retention of alleles that would have been otherwise lost.

Paraphyletic species: a species that has given rise to another phylogenetically nested species (see Fig. 3).

Phylogeography: the study of the historical processes that may be responsible for the contemporary geographic distributions of individuals.

Tajima's D: a summary statistic, based on the infinite site model, which detects the effect of natural selection on DNA samples within populations. It compares two estimates of the parameter θ. This statistic is also sensitive to population expansions, bottlenecks or heterogeneity of mutation rates.

Theta (θ): a parameter that measures the capacity of a population to maintain genetic diversity. In diploids, $\theta = 4N_e\mu$, whereas in haploids it equals $2N_e\mu$, where μ is the mutation rate at the haploid level. In a sample of DNA sequences, θ can be approximated using the number of segregating sites, when assuming an infinite site model and equilibrium.

Wright-Fisher model: a model of random genetic drift. It describes an ideal population of finite size with no overlapping generations, in which all individuals mate randomly to produce the next generation.