THE ROYAL SOCIETY

Published online 16 December 2003

Impact of Ice Ages on the genetic structure of trees and shrubs

Martin Lascoux[1*], Anna E. Palmé[1], Rachid Cheddadi[2] and Robert G. Latta[3]

[1]*Program in Conservation Biology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 75326 Uppsala, Sweden*

[2]*Institut Mediterranéen d'Ecologie et de Paléoécologie, Faculté de St Jérôme, F-13397 Marseille, France*

[3]*Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, NS B3H 4J1, Canada*

Data on the genetic structure of tree and shrub populations on the continental scale have accumulated dramatically over the past decade. However, our ability to make inferences on the impact of the last ice age still depends crucially on the availability of informative palaeoecological data. This is well illustrated by the results from a recent project, during which new pollen fossil maps were established and the variation in chloroplast DNA was studied in 22 European species of trees and shrubs. Species exhibit very different levels of genetic variation between and within populations, and obviously went through very different histories after Ice Ages. However, when palaeoecological data are non-informative, inferences on past history are difficult to draw from entirely genetic data. On the other hand, as illustrated by a study in ponderosa pine, when we can infer the species' history with some certainty, coalescent simulations can be used and new hypotheses can be tested.

Keywords: coalescent; Ice Ages; phylogeography

1. INTRODUCTION

In 1959, Ernst Mayr derisively referred to the theoretical population genetics work of Wright, Fisher and Haldane as 'beanbag genetics' (Mayr 1959, 1963). In Mayr's eyes a coherent and mathematically rigorous theory of population genetics had been obtained at the cost of ecological and genetic realism and, worse, had led researchers to address questions of dubious biological relevance. As Felsenstein's spirited and amusing overview of the history of population genetics shows, there was undoubtedly more than a grain of truth in Mayr's assertion (Felsenstein 1999, see also Ewens 1993). In part owing to the lack of data, theoretical population genetics indeed started to acquire a life of its own in the 1960s, with minimal connections to ecological and experimental population genetics. As a consequence, models often relied on many questionable assumptions. However, as pointed out by Felsenstein, this situation changed drastically in the 1980s with the accumulation of molecular data, and population genetics shifted from a theory-driven field to one driven by data analysis.

This change in emphasis was followed by a no less important conceptual shift from the forward-looking view of Haldane, Fisher and Wright (using recurrence equations and analysing the approach to equilibrium) to the backward-looking view of the coalescent (Ewens 1990). A coalescent is the lineage of alleles in a sample traced backwards in time to their MRCA allele. Kingman (1982) first provided a mathematical description of the coalescent process for n genes sampled within a single population, the so-called n-coalescent. The n-coalescent was not inherently different from the forward-looking approach, as both modelled the same processes. But the coalescent emphasizes the role of mutation in creating variation and is more amenable to the analysis of non-equilibrium conditions and of the many sequences that would start to accumulate in the 1990s. In addition, the coalescent highlights a few striking features of gene genealogies that have profound implications for data interpretation (see Nordborg (2001) for a recent review).

First, the n-coalescent made obvious the intrinsic high variability of gene genealogies. The coalescence times are drawn from an exponential distribution, with a mean determined by the 'effective population size'. This implies that, for a given effective population size, one gene can have a much longer or much shorter genealogy than another, i.e. reveal a very different history (figure 1). Because a single, non-recombinating piece of DNA (such as mtDNA or cpDNA, which have been the markers of choice in intraspecific phylogeography), allows us to retrieve a single realization of the genealogical process, the value of this single genealogy for demographic inference can be, in general, debatable. Second, for a constant population size, half of the genealogy can, on average, be attributed to the last coalescent event. That is, for much of the time after the MRCA, only two branches are present in the tree (figure 2), so that mutations that occur along these branches will have the greatest impact on the distribution of the variation in the sample. Finally, a final feature worth noting here is that if a sample of n genes is taken from a population, the MRCA of these n genes coincides with that of the whole population with

* Author for correspondence (Martin.Lascoux@ebc.uu.se).

One contribution of 14 to a Discussion Meeting Issue 'The evolutionary legacy of the Ice Ages'.

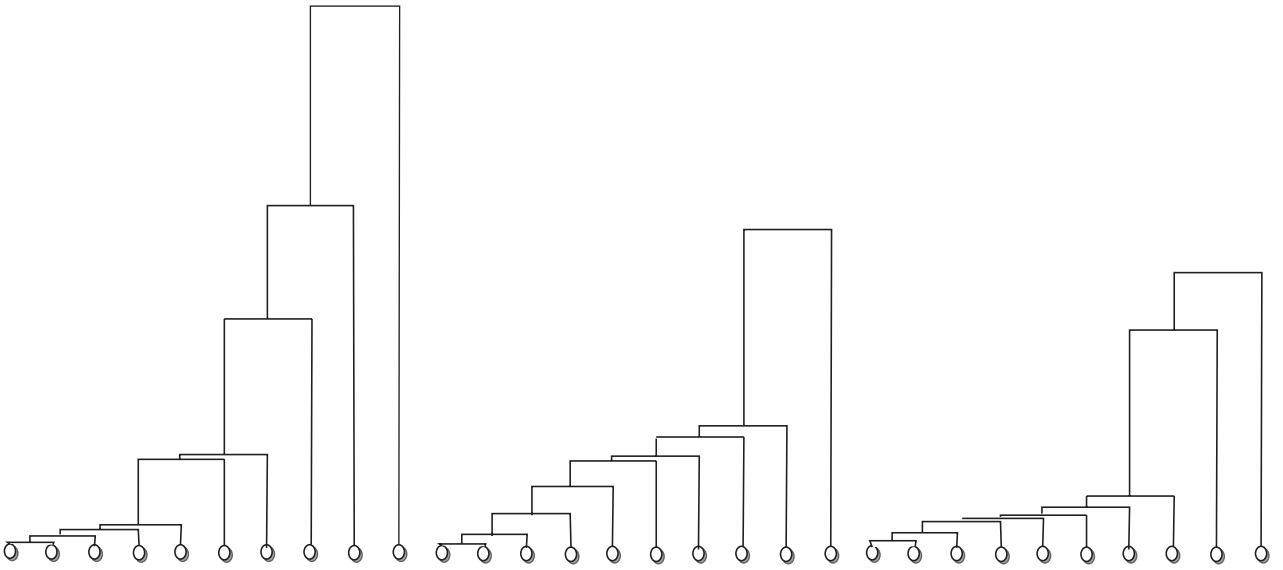

Figure 1. Three realizations of the standard coalescent process for a sample of 10 genes when the population size is constant. Note the variability in tree length.

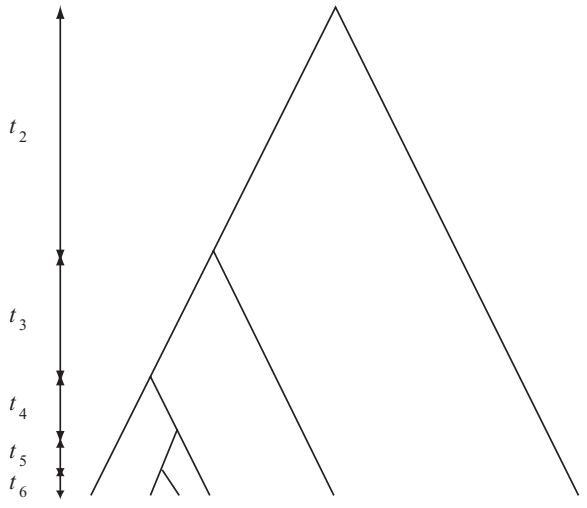

Figure 2. A hypothetical coalescent genealogy of a sample of size 6 when the population size is constant. The lengths of the coalescent intervals, t_6 through t_2, are drawn in proportion to their expected values. Note that t_2 is expected to be half of the total height when the sample size $n \rightarrow \infty$.

probability $(n - 1)/(n + 1)$. For example, if the six lineages depicted in figure 2 form a population, a random choice of two alleles has a one-third probability of tracing to the MRCA of the entire population. Consequently, at least for constant population size, it should be more valuable to increase the number of genes, than the number of individuals, when estimating population parameters. The former will average out the variability among genes whereas the latter will add little resolution because it will basically amount to adding small external branches to the tree (figure 2). We note, however, that this result should not be used uncritically. For example, when making inferences on past demographics, a sampling covering all the main areas of the natural range of a species may be necessary to sample rare variants, whose proportion is used in some test statistics as an indication of past population growth or selection (Ptak & Przeworski 2002).

In a certain sense, the argument started by Mayr in 1963 has survived to this day. Practitioners of phylogeography, the use of estimated gene genealogies to study the geographical history and structure of populations and species, tend to view their own field of enquiry as developing apart from, if not against, 'classical' population genetics (Petit *et al.* 2001; Knowles & Maddison 2002). Phylogeography would be the realm of pragmatic, ecologically realistic, historical re-constructions whereas population genetics would be the domain of abstract and fundamental disputes on the relative importance of mutation, recombination, selection, drift and migration in evolution, largely ignoring the contingencies introduced by historical events.

In truth, the most significant difference between many, if not most, practitioners of intraspecific phylogeography and population geneticists, particularly students of coalescent theory, lies in their appreciation of the amount of historical information that can be retrieved from extant polymorphism, especially when data come from a single locus, or more precisely, from a single gene genealogy or gene tree. Whereas the former group tends to be overly optimistic (see e.g. Hewitt 2000, p. 908: 'Population bottlenecks are inferred by mismatch comparisons, aspects of demography and phylogeography can be discerned by nested clade analysis, and spanning haplotype networks can reveal both demographics and geographic history. These are now being applied effectively to human populations.'), the latter group tends to be far more cautious (Nordborg 2001, see below; Hudson & Turelli 2003).

It is true that *in the expectation*, a gene tree does capture the main historical events (bottlenecks, population splits, etc. ...) that the population experienced. However, when the variability around the expectation is considered, drawing conclusions on *population* history from *single gene* genealogies can be problematic because, as the coalescent clearly shows, a single gene genealogy is a single realization of an extremely variable process. Hence, a single gene genealogy is equivalent to a sample of size of one

drawn from a large and diverse population no matter how many individuals we sample, and there is still only a *single* underlying genealogy to estimate. It could of course be that this single genealogy contains a lot of information about the interesting aspect of the evolutionary process, but if it does not, then our sample will be as good as one would normally expect from a sample of size one (Nordborg 2001, p. 182)!

In addition, the aim of population geneticists is generally to estimate *population parameters* (such as the time to the MRCA, the time at which a population started growing or the scaled mutation parameter, θ) using formal statistical models, rather than simply drawing a tree between the sampled genes. In that case, the tree relating randomly sampled genes is interesting because it informs us about the demographic history (Stephens 2001). This distinction between estimating the gene genealogy and making inference about the population is crucial but is not always clear. Nonetheless, it should be obvious that in phylogeography we are seldom interested in the former, always in the latter. The whole purpose of phylogeography is, after all, the reconstruction of the population history, not the study of the gene genealogy of a specific segment of DNA! The only case when the history of a particular locus would be interesting is when selection is involved. But we usually start by assuming that the gene under study is neutral (even when there is mounting evidence that this is unlikely to be the case, e.g. *Drosophila* mtDNA (Hudson & Turelli 2003)).

Until recently, most coalescent-based computer programs could only accommodate fairly simple demographic models, and, under these circumstances, phylogeographers (like morphologists before them (Felsenstein 1999)) could be forgiven for concluding that population genetics was of little use to them. This is, however, quickly changing; new coalescent-based computer programs are far more versatile than their predecessors and two recent articles well illustrate how a bridge between the two fields could be built (Knowles & Maddison 2002; Wakeley 2004). Wakeley (2004), for instance, assessed under which demographic circumstances the drawing of demographic inferences from single gene genealogies would be reasonable. He concluded that 'at the intraspecific level, roughly speaking, the circumstances favourable to using inferred gene trees are those in which random genetic drift is relatively unimportant compared to non-equilibrium factors like the splitting of populations'. But, how can we, *a priori*, tell these two situations apart? More generally, as coalescent-based estimates are obtained by conditioning on a given population history, which historical situations are plausible? That is where palaeoecological data enter the picture.

In plants it is sometimes possible to know, from a network of fossil records spanning the past 15 kyr, whether a single gene genealogy is likely to be informative or not, and what were the main historical events that the population experienced since the LGM. For instance, the fossil pollen records provide a fairly good picture of the history of thermophilous species such as oaks since the LGM (Brewer *et al.* 2002). Oaks were confined to three main southern European refugia during the LGM: the Iberian peninsula; the Balkans; and Italy. They left these as climate became warmer and they reached Scandinavia 6 ka (Brewer *et al.* 2002; Taberlet & Cheddadi 2002). Once they had left the glacial refugia, which were probably isolated, and went through bottlenecks, the population sizes grew rapidly and it is safe to assume that random genetic drift was relatively unimportant compared with non-equilibrium factors like the splitting of populations. Alternatively, other species, such as birches, pines, spruces or willows are cold tolerant and were apparently able to survive the LGM at much higher latitudes. In birches or willows, the fossil pollen record provides very little information on possible LGM refugia and on the fate of the populations after the LGM, and all reliable information might have to come from serendipitous finds of macrofossils (Willis *et al.* 2000; Stewart & Lister 2001; Birks 2003). As we shall see, in the absence of a fossil record, it will be very difficult to make any inference from genetic data alone and it will be difficult to justify the use of a specific demographic model in coalescent-based inferences. Hence, at the bottom of many phylogeographical inferences, there is implicitly, if not explicitly, a close interplay between the information provided by the fossil record and the information provided by genetics.

We shall now try to illustrate this implicit or explicit interplay between the fossil record and the genetic data. First we shall give an overview of what has been inferred on the impact of Ice Ages on forest trees and shrubs from genetic data and the fossil record. For this we will draw heavily from a recent project, involving seven European research groups, during which new fossil pollen maps were established (EU Project CYTOFOR, unpublished data) and the variation in cpDNA was studied in 22 species of trees and shrubs (Petit *et al.* 2003). The main conclusion from this comparison is that species exhibit very different levels of genetic variation between and within populations and probably went through very different postglacial histories. We will then show, using the same dataset, that when the pollen fossil record is non-informative, conclusions on putative refugia and colonization routes are difficult to draw solely from genetic data. Finally, we will present an example from North America and show that when palaeoecological information allows a fairly good sketch of the species' recent history and when the species' history fulfils the requirements spelled out by Wakeley (2004), new hypotheses can be tested when analysing the genetic data.

2. DIFFERENT SPECIES, DIFFERENT HISTORIES

Since the pioneering work of Huntley & Birks (1983), pollen fossil maps have been continuously upgraded for many European species. Keeping in mind the limitations due to extensive pollen dispersal in some species (Birks 2003), the resolution is now sufficient to infer the pace and direction of postglacial re-colonization of some species to a satisfying degree of accuracy. This is the case, for instance, for *Quercus* (Brewer *et al.* 2002). Unfortunately, for species such as birch or willow, the fossil pollen record is much less informative, and can even be rather misleading in the absence of macrofossils (Birks 2003). As stated in § 1, such discrepancies between these temperate species are in part related to the location of their glacial refugia that represent the starting points of the postglacial migrational process. Both fossil pollen (figure 3) and

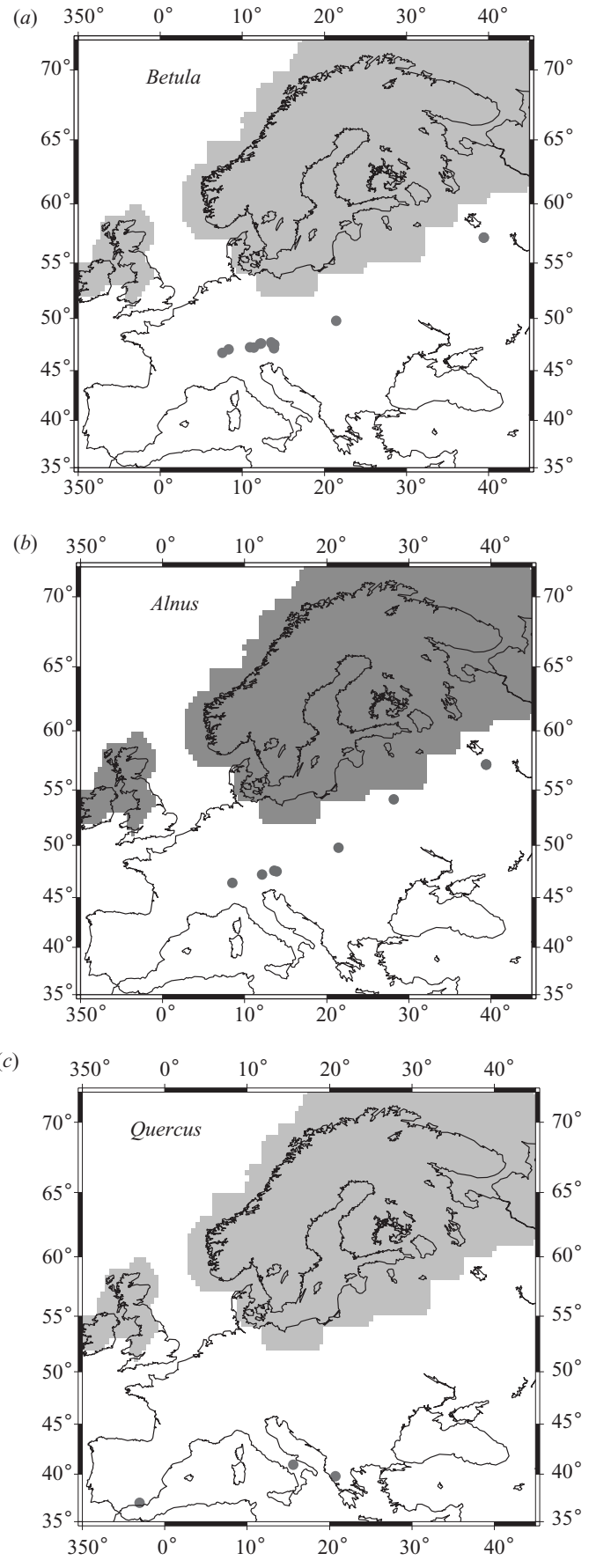

Figure 3. Presence of fossil pollen at the LGM in (*a*) *Betula* spp., (*b*) *Alnus glutinosa*, (*c*) *Quercus* spp. Grey areas are glaciated areas during the LGM (data from European Pollen Database).

macrofossils (Willis *et al.* 2000) indicate, for instance, that birch and alder (figure 3*a*,*b*) survived at higher latitudes than oaks (figure 3*c*), in areas where a supply of moisture was available from the summer melting of ice. Oak trees, on the other hand, were not able to survive in such areas during the last glacial maximum because they are not as cold tolerant as birch or alder. Thus, cold-tolerant species of genera such as *Betula*, *Alnus*, *Populus* and *Salix* apparently had more scattered and northerly glacial refugia, close to the ice cap in some cases, than other more warmth-demanding deciduous temperate species. The difference between, for instance, birches and oaks in post-glacial re-colonization pattern was also magnified by the different dynamics of these two species. Birches started recolonizing northern latitudes quite early and therefore encountered few hinderances to their rapid migration northwards. They also have a much shorter generation time than oaks.

In parallel to the establishment of a new generation of pollen fossil maps, genetic data from several woody angiosperms were gathered across Europe, using standardized sampling and molecular screening methods (Petit *et al.* 2003). Briefly, about 10 individuals per species were sampled according to a standard procedure in 25 European forests selected on the basis of their high species richness and limited human influence. Polymorphism in the cp genome was assessed in all 22 species by using PCR–RFLP to screen for indels and point mutations in regions amplified by universal primers and automated DNA sequencers to identify variation in the number of repeat units in cp microsatellite regions. For each species, polymorphism was measured by genetic diversity and number of haplotypes; the degree of subdivision among population was assessed with Nei's coefficient of gene differentiation, G_{ST} (Nei 1987, p. 190; table 1; Petit *et al.* 2003). A new generation of fossil pollen maps was also constructed. Some general conclusions emerge when the 22 species are analysed jointly. First, population differentiation is higher than average among Corsica, Italy and the Balkans (figure 4). Second, both mean number of haplotypes and gene diversity were higher than average in central France, southern Germany and Slovakia, whereas the southern- and northern-most populations generally had low or average diversity, with the exception of southwestern Sweden (figure 5).

However, this general conclusion conceals important differences between species. This is indicated by the correlations between the pattern of divergence of individual species and the pattern of divergence for all species combined, excluding that particular species (r_D) and the same correlation for allelic richness (r_H; table 1): in both cases, the correlation is non-significant in about half of the cases (11/22 and 8/21, respectively). Because the 22 species involved in this study were chosen to represent variants in an array of life-history traits (e.g. seed dispersal and pollination modes, cold tolerance, see table 1), the resulting diversity in genetic patterns is perhaps not surprising; as the fossil record had already suggested, 'plant communities are impermanent assemblages resulting from the individualistic behaviour of taxa in response to environmental changes' (Bennett 1997).

Table 1. List and parameter values for the 22 species investigated in the CYTOFOR project.
(The seed dispersal mode was classified as follows: Ac, animal cached; Ai, animal ingested; Wc, cottony, wind dispersed; Ww, winged, wind dispersed; Wa, wind dispersed and animal ingested; and Ed, explosive dehiscence. r_D measures the correlation between the pattern of divergence of a given species and the pattern of divergence for all species combined, excluding that particular species, whereas r_H corresponds to the same correlation for allelic richness. nc, not computed (see Petit *et al.* (2003) for details).)

species	family	seed dispersal mode	number of populations	total number of haplotypes	G_{ST}	r_D	r_H
Acer campestre	Aceraceae	Ww	16	14	0.71	0.65*	0.07
Acer pseudoplatanus	Aceraceae	Ww	19	22	0.66	0.62*	0.60*
Alnus glutinosa	Betulaceae	Ww	25	12	0.81	0.62*	0.26
Betula pendula	Betulaceae	Ww	23	9	0.42	–0.02	0.48*
Calluna vulgaris	Ericaceae	Wa	17	12	0.59	0.22	0.39
Carpinus betulus	Betulaceae	Ww	18	4	1.00	0.81*	nc
Corylus avellana	Betulaceae	Ac	24	5	0.89	0.173*	0.66*
Crataegus monogyna	Rosaceae	Ai	21	4	0.24	0.14	0.76*
Cytisus scoparius	Fabaceae	Ed	18	24	0.57	0.54*	0.61*
Fagus sylvatica	Fagaceae	Ac	23	6	0.74	0.70*	–0.07
Fraxinus sp.	Oleaceae	Ww	24	7	0.86	0.08	0.41*
Hedera sp.	Araliaceae	Ai	22	11	0.57	0.21	0.58*
Ilex aquifolium	Aquifoliaceae	Ai	16	8	0.60	0.18	0.47*
Populus tremula	Salicaceae	Wc	23	30	0.11	0.47*	0.54*
Prunus avium	Rosaceae	Ai	23	16	0.29	0.78*	0.42*
Prunus spinosa	Rosaceae	Ai	25	50	0.32	0.44*	0.57*
Quercus sp.	Fagaceae	Ac	25	10	0.84	0.40*	0.31
Rubus sp.	Rosaceae	Ai	23	15	0.31	–0.04	0.34
Salix caprea	Salicaceae	Wc	25	29	0.09	–0.16	0.01
Sorbus torminalis	Rosaceae	Ai	17	26	0.33	–0.11	0.69*
Tilia cordata	Tiliaceae	Ww	16	16	0.57	0.45	0.66*
Ulmus sp.	Ulmaceae	Ww	25	41	0.47	0.34	0.10
mean			21.3	16.9	0.54	0.36	0.42

* $p < 0.05$.

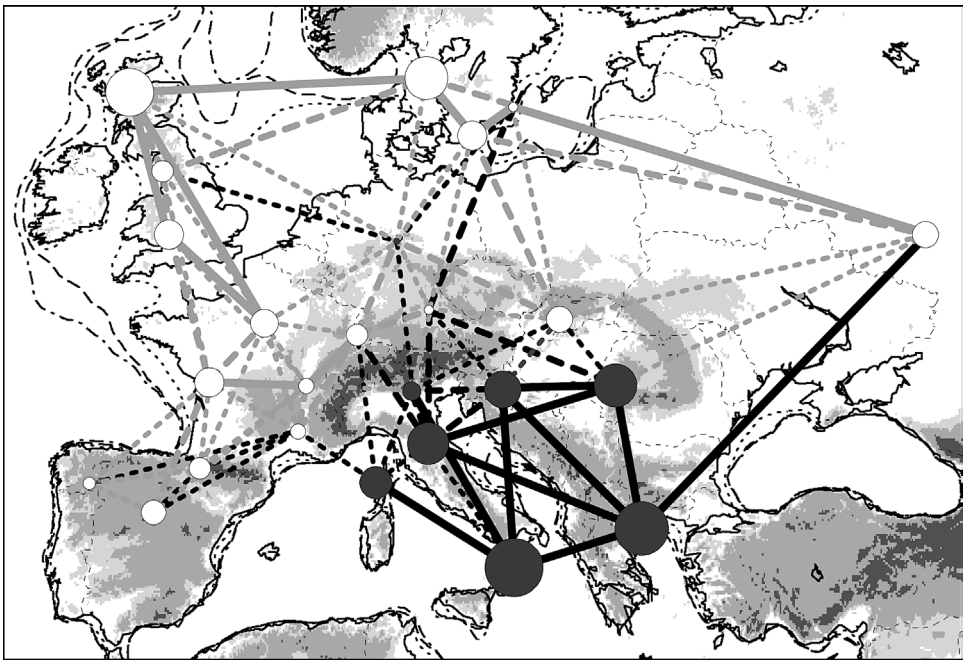

Figure 4. Multispecies genetic divergence of each of the 25 European forests studied. Higher than average values are in black, lower than average are in white, with diameter proportional to the difference from the mean value. For all forests, the level of divergence with each of the five nearest forests was represented by connecting lines, with continuous black lines indicating comparatively high divergence, dotted lines intermediate divergence (black, higher than the mean; grey, lower than the mean) and continuous grey lines low divergence. The altitude is indicated by grey shadings (250–500, 500–1000, > 1000 m) and past sea-levels at 21 kyr BP, 15 kyr BP and 12 kyr BP are indicated by black dotted lines (from Petit *et al.* 2003).

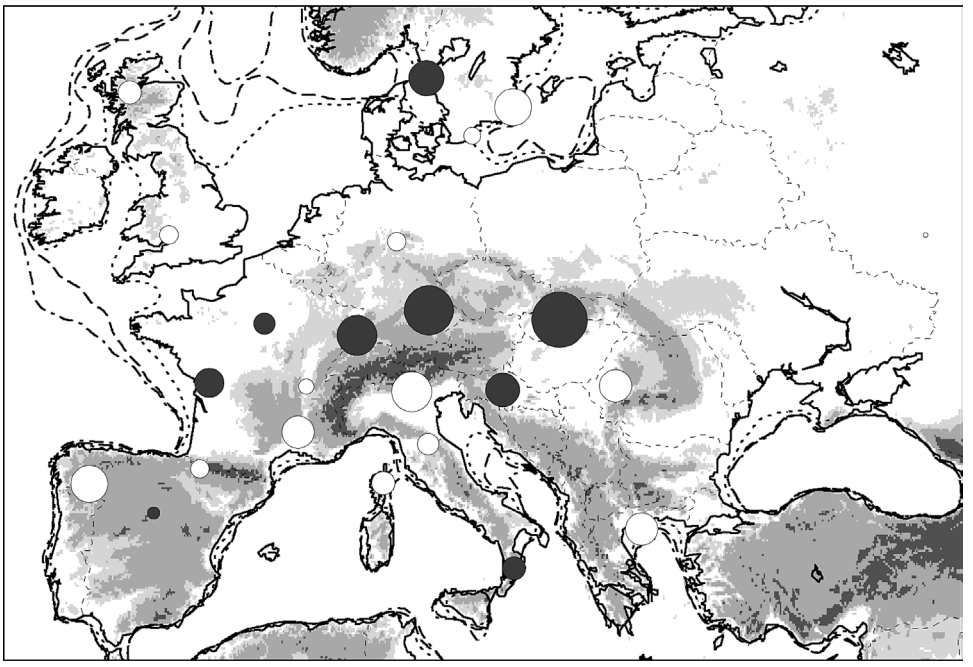

Figure 5. Mean number of haplotypes per forest, averaged across species. Legend as for figure 4. Diversity is highest at relatively high latitudes, north of the three European peninsulas (from Petit *et al.* 2003).

3. DEPARTING SPECIES: *BETULA*, *SALIX* AND OTHERS

What was possibly more surprising was the fairly high number of species that did not fit the average patterns of divergence and diversity. Furthermore, this lack of fit seems to have had very different causes. For instance, neither sallow (*Salix caprea*) nor silver birch (*Betula pendula*) fit the average pattern, but *S. caprea* is characterized by very many haplotypes and a complete absence of geographical structure (Palmé *et al.* 2003*a*), whereas the number of haplotypes in *B. pendula* is limited but rare haplotypes allow the delineation of different areas (Palmé *et al.* 2003*b*). The lack of structure in *S. caprea* may reflect both extensive gene flow through seeds and introgression (as also in *Rubus* sp.), whereas the resulting pattern in *B. pendula* (figure 6) may simply be a consequence of the presence of more northerly and more scattered refugia than were available to other species, such as oaks, which were confined to Spain, Italy and the Balkans during the LGM. Interestingly, the genetic structure of heather (*Calluna vulgaris*) was similar to that observed in birch and the explanation outlined here could also apply to it (Rendell & Ennos 2002).

Unfortunately, in most of these cases, a descriptive analysis of the haplotypes' distribution and limited inferences on post-glacial history seem to be the only available options because the fossil record provides very limited information on putative refugia and colonization routes for most species. Of course, this does not imply that nothing is learned from these studies. Undoubtedly, interesting conclusions can still be drawn from the distribution of cp genetic variation of *Betula* as follows:

(i) rare haplotypes do allow the delineation of different areas in *Betula*; and
(ii) populations from the Iberian peninsula apparently did not participate in postglacial recolonization (A. E. Palmé, unpublished data).

But this is still a long way from a satisfying reconstruction of the species history since the LGM.

4. THE FOSSIL RECORD PROVIDES A DEMOGRAPHIC SITUATION

In North America there has been no integrated attempt to compare phylogeographical with palaeoecological results across several species to nearly the same extent as seen in Europe. This is partly because of the difficulty of access to some parts, but some regional comparisons have been made in several areas (Avise 1992; Soltis *et al.* 1997). In montane forest trees there is a clear east–west biogeographical separation in Western USA, seen across several species such as Douglas fir, lodgepole pine, whitebark pine and ponderosa pine (Critchfield 1984*a*; Richardson *et al.* 2002).

In ponderosa pine (*Pinus ponderosa* Laws), the combination of biogeographical, pollen and macrofossil (packrat middens) evidence shows a relatively simple history through the Wisconsin (last) glaciation, with the species being separated into two populations followed by recent secondary contact. Two varieties are recognized—*P. p.* var. *ponderosa* in the Sierra Nevada and inland Northwest, and *P. p.* var. *scopulorum*, present in the interior Rockies and the mountain islands of the desert southwest (figure 7). These two varieties are separated through much of their combined range by the Great Basin in the south and by the northern Rockies of Idaho, Wyoming and Montana. But they come into contact in a small region of west-central Montana (Conkle & Critchfield 1988). The two varieties are inter-fertile, producing viable progeny and hybrids are observed in the field (Latta & Mitton 1999),

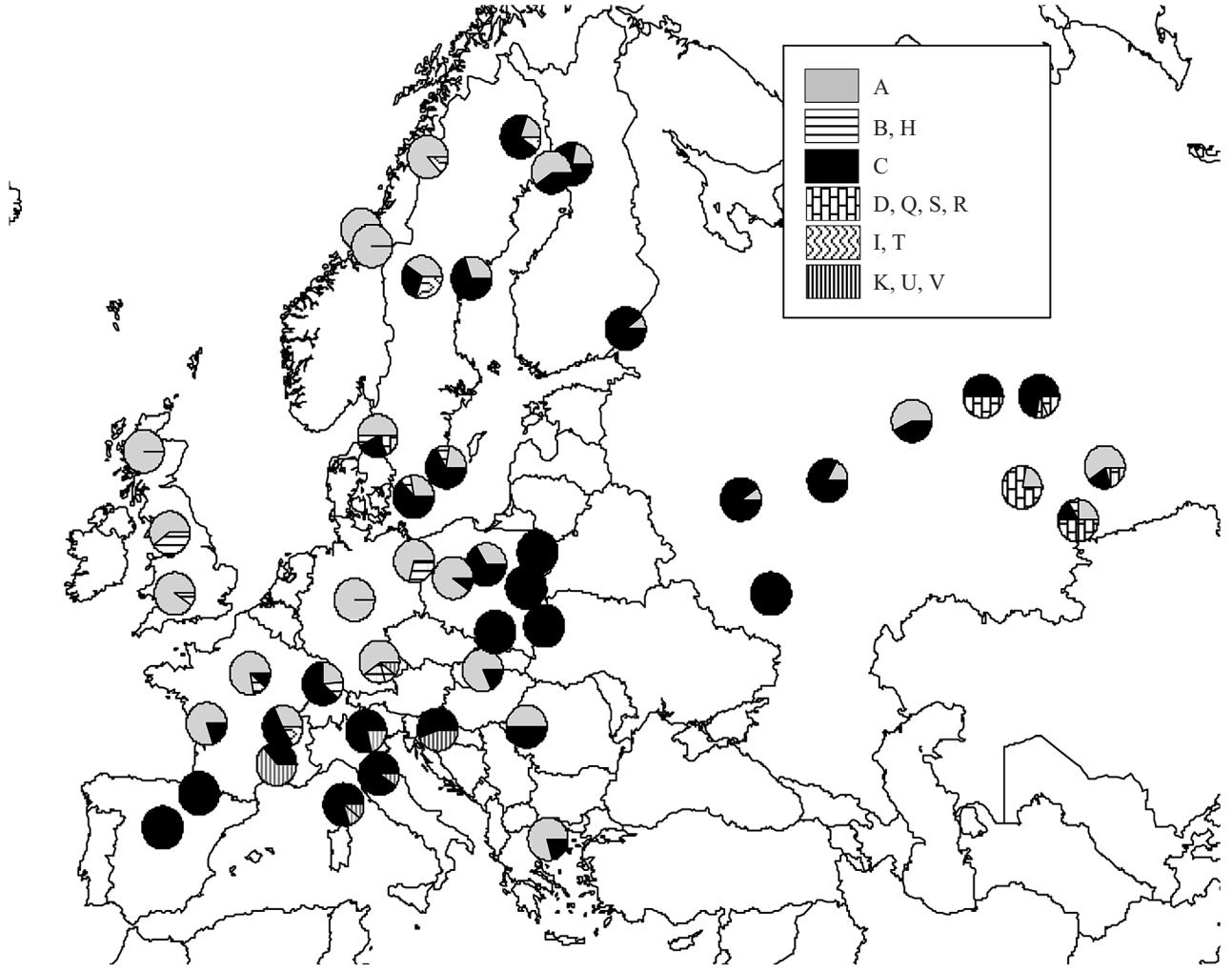

Figure 6. Geographical distribution of PCR–RFLP cp haplotypes in *B. pendula*. The pie charts give the relative frequencies of each haplotype present in the population. The location of the pie charts generally corresponds to the location of the population they represent, but when the populations were very close to each other the positions of the charts have been modified to avoid overlap.

although hand pollinations suggest that some barrier to gene exchange may be present (Critchfield 1984*b*). Pine pollen can be difficult to identify to species, so the pollen record for the western USA is not complete, and exact locations of refugia are still not known with certainty. However, the combination of pollen and packrat middens data suggests that the contact zone was established very recently after the Wisconsin glaciation. Barnosky *et al.* (1987) report ponderosa pine pollen appearing *ca*. 7 ka in the sediment cores taken from lakes in southwest Washington. Bettancourt *et al.* (1990) and Thompson (1990) report that ponderosa pine macrofossils are not found in packrat middens north of Arizona aged more than *ca*. 10 ka. Virtually no inter-mixing of the two varieties by seed movement has taken place, even within the contact zone. Maternally inherited mtDNA reveals no shared haplotypes between Eastern and Western varieties outside of an extremely narrow contact zone only 10 km wide (Johansen & Latta 2003). Further regional subdivisions within each variety are seen, with similarly sharp boundaries (Johansen & Latta 2003). Some gene exchange by pollen movement has occurred within the contact zone, with analysis of *paternally* inherited cpDNA showing a cline *ca*. 100 km long, 10-fold that of mtDNA (Latta & Mitton 1999). However, although pollen movement is demonstrably the main agent of gene flow in ponderosa pine, this does not extend to most of the species' range, and the organellar DNA clearly captures the major historical division in this species (figure 7). Nuclear loci, alternatively, although showing significantly different allele frequencies between the varieties at several loci (figure 7), do not reveal nearly as great a level of genetic divergence (Niebling & Conkle 1990; Latta & Mitton 1999). This clearly cannot be ascribed to gene flow through pollen because of the sharp divergence of paternally inherited cpDNA. The nuclear allozyme patterns are also consistent with a historical division. Coalescence theory predicts that if a few independent populations are present, there should be a high variability among loci in the degree of subdivision they reveal (Robertson 1975). This prediction is easy to test empirically. Allozyme data were available for two regions within the range of the western variety (*P. p.* var. *ponderosa*), as well as from the interior variety (*P. p.* var. *scopulorum*) (Niebling & Conkle 1990). The level of divergence is greater between the varieties than within either (both comparisons involved approximately the same geographical distance), but, more importantly, it is clearly more variable among loci (Latta & Mitton 1999), as

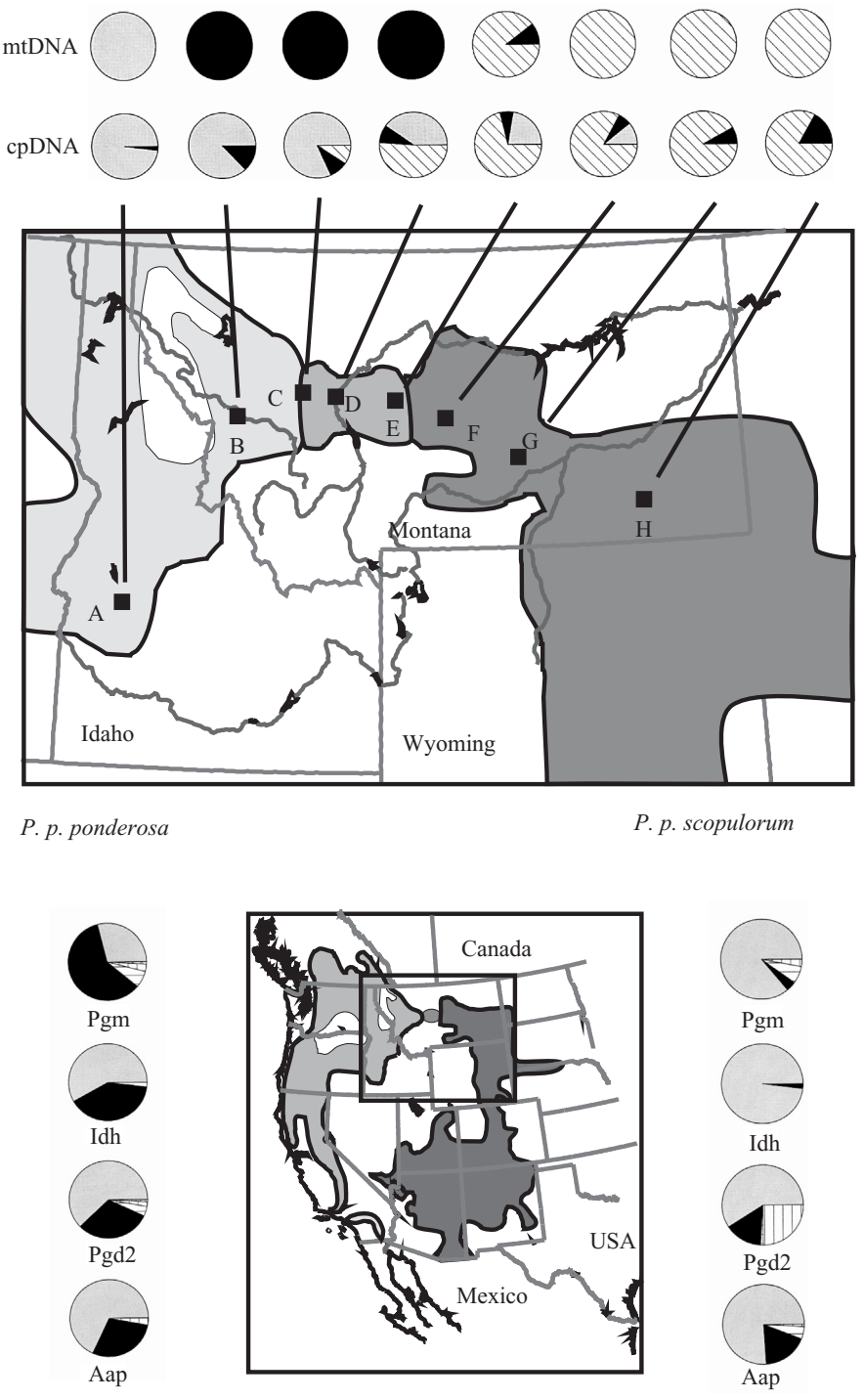

Figure 7. Distribution of genetic variation in *P. ponderosa*. The upper map indicates the sampling sites for mtDNA and cpDNA variation through the secondary contact zone. Haplotypes A, B and C as surveyed by Latta & Mitton (1999) are indicated in grey, black and hatched sections of the diagrams, respectively. The lower map locates the transect within the range of ponderosa pine, and gives the allele frequencies for four (out of 16) representative allozyme loci (from Niebling & Conkle 1990), showing the range of results from Pgm (F_{ST} = 0.282) to Aat (F_{ST} = 0.007).

predicted by coalescence theory. Integrating the organellar with the nuclear findings can be accomplished by using coalescent simulations (Hudson 1990). The fairly simple species history suggested by the pollen record provides a framework for the simulations; an ancestral population was separated into two daughter populations at time τ in the past. For all combinations of τ (measured in generations) and effective population size (N_e), the distribution of allelic or haplotype frequencies expected in the sample can be derived quickly (Latta *et al.* 2004). For $\tau < 2N_e$, the fixed differences seen at organellar haplotypes fall in the tail of the simulated distribution. Therefore we can reject any hypothesis with $\tau < 2N_e$ at less than 5% probability. On the other hand, the distribution of Wright fixation index, F_{ST}, across allozyme loci is significantly different from the simulated distribution for $\tau > 2N_e$. It would therefore appear that the two varieties of ponderosa pine have been separate for a time, in

generations, roughly equal to twice their effective population size. Simulations assuming smaller effective population sizes (less than 10 000) were only consistent with observations by assuming an implausibly high mutation rate for allozymes. These simulations suggest that ponderosa pine did not go through a pronounced bottleneck during the Wisconsin glaciation, and, since $\tau \approx 2N_e$, the varieties appear to have been separated for over 10 000 generations (more than 250 kyr).

The combination of coalescent simulations, genetic surveys and explicit species history from biogeographical and pollen evidence thus makes it possible to arrive at a fairly complete accounting of the forces shaping genetic variation in ponderosa pine. In spite of the sharp contrast between organellar and nuclear loci, the coalescent model does not invoke selection. This is not to argue for a panneutralist interpretation of genetic variation as adaptation will undoubtedly play a role in shaping genetic variation. We recognize that inferences drawn on the assumption of neutrality must be treated with caution (Ford 2002). However, what the simulations demonstrate is that the patterns can apparently be explained without the *need* to invoke strong selective pressures, and therefore do not present strong evidence for such selection. That it is possible to explain the seemingly opposite inferences about gene flow derived from nuclear versus organellar markers (figure 7) within a simple historical context highlights the following:

(i) the inherent variability of the coalescent (Hudson & Turelli 2003); and
(ii) the different rates of divergence at organellar and nuclear loci.

Because ponderosa pine is a hermaphrodite, the difference in effective population size between nucleus and organelles is only twofold. So at migration/drift equilibrium the difference between the two is expected to be relatively minor. Nevertheless, Palumbi *et al.* (2001) showed that, under complete separation, organelles will reach this equilibrium much faster than nuclear markers, whereas stochasticity further blurs the difference (Hudson & Turelli 2003). More complex histories could of course be simulated, but simplicity of the model is an important criterion when selecting among competing hypotheses (Knowles & Maddison 2002). Fine-scale details, such as the precise location of glacial refugia, are obscured in a simple model, but the ability to integrate multiple lines of evidence lends strength to the interpretation presented here.

5. CONCLUSION

In the present paper we briefly reviewed some recent work on the impact of Ice Ages on the genetic structure of trees and shrubs. We purposely focused on intraspecific genetic variation and on the impact of the LGM because more long-term inferences are so far limited and highly speculative (Tzedakis *et al.* 2002). We also concentrated on the impact on demography as inferred from organellar genomes. This is simply because these have been the markers of choice in most phylogeographical studies, not because we believe that they are more informative or valuable. As we showed in § 1, they do have strong inherent limitations, the main one being that they are equivalent to a single locus. While organellar genomes will remain useful tools in the future, polymorphism at nuclear markers will become increasingly important. Owing to recombination, many realizations of the demographic process can be retrieved with nuclear markers, reducing the variability of the estimates of population parameters. In addition, if polymorphism at many loci is assessed it can serve as 'background' for the detection and analysis of loci of adaptive significance. The impact of the last Ice Ages on adaptation has remained largely unexplored, at the molecular level at least (Davis & Shaw 2001), but there too the impact of genome projects and the accompanying availability of molecular markers for a large range of species will soon be felt.

Regarding the impact of the LGM on neutral genetic structure and inferences on demographics, much progress has been made since the first continental studies of genetic variation at molecular markers in forest trees (Cwynar & MacDonald 1987; Lagercrantz & Ryman 1990). If, in most cases, much remains to be discovered, we now have a good sketch of what happened to a host of woody plant species since the LGM. The most striking feature of the emerging picture is perhaps the diversity of patterns. In other words, species attributes and historical contingencies make general predictions based on simple genetic models largely irrelevant.

The second important point that emerges from our review is of a more methodological character. Most 'classical' phylogeography makes an implicit use of the fossil record. For example, a high genetic diversity has traditionally been used to identify the location of putative glacial refuges. Using this line of reasoning, the data of the multispecies project presented in § 2 (Petit *et al.* 2003) would have led us to locate the glacial refugia of many species in regions of Central Europe. Although this would be patently silly for thermophilous species like oaks, it may have seemed sensible for more cold-tolerant species such as willows or birches. Instead we chose to interpret the areas of high variability as the result of postglacial merging of populations migrating from different glacial refugia in the Mediterranean region, because palaeoecological data suggest that most species involved in the study were not in those putative refugia during the last glacial maximum. Coalescent-based analysis, by contrast, uses palaeoecological data in a more explicit way: all parameters characterizing the population history are estimated conditional on a given demographic model based on palaeoecological data. So, apart from taking into account the stochasticity of both the evolutionary and sampling processes (the first one being simply ignored in 'classical' phylogeographical methods, see Knowles & Maddison (2002)), a coalescent-based analysis also explicitly reflects the uncertainty attached to our knowledge of past climate. This may be perceived as a weakness, but it is, at any rate, better than avoiding the explicit statement of what is assumed. In addition, as we saw in the case of *Betula*, an uninformative pollen record also curtails the conclusions that can be drawn from the genetic data, independently of the type of analysis that is done.

So, what is the future of phylogeographical studies? Probably, studies based on organellar genomes will

continue unabated: they are relatively easy to do and undoubtedly provide good hints on what went on. In some cases, a more extensive coverage of the natural range is obviously needed to draw meaningful conclusions on the impact of the Ice Ages. For instance, phylogeographical studies on European species often do not include any populations from Eastern Europe and Russia. Whereas the outlines and ages of the LGM's Scandinavian and British ice sheets are well established (Peltier 1994), the extent of glaciation of northwestern Russia and eastern Siberia during the LGM is still debated (Clark & Mix 2002). In northwestern Russia the extent of glaciation was apparently less than predicted earlier, making the areas that could serve as refugia to cold-tolerant species during the LGM larger. The Russian plains have already been suggested as refugia for some forest tree species (e.g. *Picea abies*; Lagercrantz & Ryman 1990) but clearly more studies are needed to assess the part played by this vast area during the LGM. Organellar DNA will, of necessity, also play the main part in ancient DNA studies. In plants these studies are still in their infancy and the first step will be to overcome technical difficulties and establish reproducible methods of DNA extraction and analysis. Hence, while ancient DNA will contribute to qualitative inferences on past demographics, it will not constitute an alternative to the analysis of contemporary DNA for years to come. Furthermore, ancient organellar DNA has the same limitations as contemporary organellar DNA (absence of recombination), leaving no alternative but the analysis of contemporary nuclear DNA if one aims at statistically meaningful estimates of past demographics. This is not an easy task but the rewards are certainly worth the effort. In addition to providing better estimates of population parameters, the analysis of nuclear markers, if combined with the analysis of the sequence of candidate genes, can also help link phylogeography studies to studies of adaptation. Decades of observations have shown that forest trees are characterized by steep latitudinal gradients in life-history traits such as budburst or budset, and we are now in a position to start investigating the molecular basis of this strong pattern of local adaptation and how it relates to the post-glacial history of the species. Never before have we been in a better position to carry out Mayr's programme with the rigor aimed at by Haldane, Wright and Fisher.

The authors thank The Royal Society and the organizers of the meeting 'The evolutionary legacy of the Ice Ages' for inviting them. This study has been done with financial support from the Commission of the European Communities, Agriculture and Fisheries (FAIR) specific RTD programme, CT97-3795, 'CYTOFOR' and the Carl Tryggers Foundation. Thanks to all the participants of the CYTOFOR project and to V. Semerikov for helping M.L. with sampling in Russia. The authors thank J. Wakeley for providing them with a copy of his manuscript before publication and the European Pollen Database and its contributors (see http://www.ngdc.noaa.gov/palaeo/epd/epd_main.html).

REFERENCES

Avise, J. C. 1992 Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos* **63**, 62–76.

Barnosky, C. W., Anderson, P. M. & Bartlein, P. J. 1987 The northwestern U. S. during deglaciation: vegetational history and paleoclimatic implications. In *North America and adjacent oceans during the last deglaciation* (ed. W. F. Ruddiman & H. E. Wright Jr), pp. 289–321. Boulder, CO: Geological Society of America.

Bennett, K. D. 1997 *Evolution and ecology. The pace of life*. Cambridge University Press.

Bettancourt, J. L., Van Devender, T. R. & Martin, P. S. 1990 *Packrat middens: the last 40 000 years of biotic change*. Tucson, AZ: University of Arizona Press.

Birks, H. H. 2003 The importance of plant macrofossils in the reconstruction of Late glacial vegetation and climate: examples from Scotland, western Norway and Minnesota, USA. *Quatern. Sci. Rev.* **22**, 453–473.

Brewer, S., Cheddadi, R., de Beaulieu, J. L. & Reille, M. & Data contributors 2002 The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecol. Mngmt* **165**, 27–48.

Clark, P. U. & Mix, A. C. 2002 Ice sheets and sea level at the Last Glacial Maximum. *Quatern. Sci. Rev.* **21**, 1–7.

Conkle, M. T. & Critchfield, W. B. 1988 Genetic variation and hybridization of ponderosa pine. In *Ponderosa pine: the species and its management* (ed. D. M. Baumgartner & J. E. Lotan), pp. 27–43. Pullman, WA: Washington State University Press.

Critchfield, W.B. 1984*a* Impact of the Pleistocene on the genetic structure of North American conifers. In *Proceedings of the Eighth North American Forest Biology Workshop*, 30 July–1 August 1984 (ed. R.M. Lanner), pp. 70–118. Logan, UT: Utah State University Press.

Critchfield, W. B. 1984*b* Crossability and relationships of Washoe pine. *Madrono* **31**, 144–170.

Cwynar, L. C. & MacDonald, G. M. 1987 Geographical variation of lodgepole pine in relation to population history. *Am. Nat.* **129**, 463–469.

Davis, M. B. & Shaw, R. G. 2001 Range shifts and adaptive responses to quaternary climate change. *Science* **292**, 673–679.

Ewens, W. J. 1990 Population genetics theory—the past and the future. In *Mathematical and statistical developments of evolutionary theory* (ed. S. Lessard), pp. 177–227. Amsterdam: Kluwer.

Ewens, W. J. 1993 Beanbag genetics and after. In *Human population genetics* (ed. P. P. Majunder), pp. 7–29. New York: Plenum.

Felsenstein, J. 1999 From population genetics to evolutionary genetics: a view through the trees. In *Evolutionary genetics: from molecules to morphology*, vol. 1 (ed. R. S. Singh & C. B. Krimbas), pp. 609–627. New York: Cambridge University Press.

Ford, M. J. 2002 Applications of selective neutrality tests to molecular ecology. *Mol. Ecol.* **11**, 1245–1262.

Hewitt, G. 2000 The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. In *Oxford surveys in evolutionary biology*, vol. 7 (ed. D. Futuyma & J. Antonovics), pp. 1–43. Oxford University Press.

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.

Huntley, B. & Birks, H. J. B. 1983 *An atlas of past and present pollen maps for Europe: 0–13 000 years ago*. Cambridge University Press.

Johansen, A. D. & Latta, R. G. 2003 Mitochondrial haplotype distribution, seed dispersal and patterns of post glacial expansion of ponderosa pine. *Mol. Ecol.* **12**, 293–298.

Kingman, J. F. C. 1982 The coalescent. *Stochast. Proc. Appl.* **13**, 235–248.

Knowles, L. L. & Maddison, W. P. 2002 Statistical phylogeography. *Mol. Ecol.* **11**, 2623–2635.
Lagercrantz, U. & Ryman, N. 1990 Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution* **44**, 38–53.
Latta, R. G. & Mitton, J. B. 1999 Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution* **53**, 769–776.
Latta, R. G., Boudreau, M. E. R. & Zeleneitz, S. A. 2004 Resolving discordant levels of population divergence across loci in ponderosa pine. *Mol. Ecol.* (Submitted.)
Mayr, E. 1959 Where are we? *Cold Spring Harbor Symp. Quantitative Biol.* **24**, 1–14.
Mayr, E. 1963 *Animal species and evolution*. Cambridge, MA: Harvard University Press.
Nei, M. 1987 *Molecular evolutionary genetics*. New York: Columbia University Press.
Niebling, C. R. & Conkle, M. T. 1990 Diversity of Washoe pine and comparisons with allozymes of ponderosa pine races. *Can. J. Forest Res.* **20**, 298–308.
Nordborg, M. 2001 Coalescent theory. In *Handbook of statistical genetics* (ed. D. Balding, M. Bishop & C. Cannings), pp. 179–212. Chichester: Wiley.
Palmé, A. E., Semerikov, V. & Lascoux, M. 2003*a* Absence of geographical structure of chloroplast DNA in sallow, *Salix caprea* L. *Heredity* **91**, 465–474.
Palmé, A. E., Su, Q., Rautenberg, A., Manni, F. & Lascoux, M. 2003*b* Postglacial recolonisation and cpDNA variation of silver birch, *Betula pendula*. *Mol. Ecol.* **12**, 201–212.
Palumbi, S. R., Cipriano, F. & Hare, M. P. 2001 Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* **55**, 859–868.
Peltier, W. R. 1994 Ice Age paleotopography. *Science* **265**, 195–201.
Petit, R. J., Bialozyt, R., Brewer, S., Cheddadi, R. & Comps, B. 2001 From spatial patterns of genetic diversity to postglacial migration processes in forest trees. In *Integrating ecology and evolution in a spatial context* (ed. J. Silvertown & J. Antonovics), pp. 295–318. Oxford: Blackwell Science.
Petit, R. J. (and 16 others) 2003 Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565.
Ptak, S. E. & Przerworski, M. 2002 Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet.* **18**, 559–563.
Rendell, S. & Ennos, R. A. 2002 Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. *Mol. Ecol.* **11**, 69–78.
Richardson, B. A., Brunsfeld, J. & Klopfenstein, N. B. 2002 DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). *Mol. Ecol.* **11**, 215–227.
Robertson, A. 1975 Gene frequency distributions as a test of selective neutrality. *Genetics* **81**, 775–785.
Soltis, D. E., Gitzendanner, M. A., Strenge, D. D. & Soltis, P. A. 1997 Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Pl. Syst. Evol.* **206**, 353–373.
Stephens, M. 2001 Inference under the coalescent. In *Handbook of statistical genetics* (ed. D. Balding, M. Bishop & C. Cannings), pp. 213–238. Chichester: Wiley.
Stewart, J. R. & Lister, A. 2001 Cryptic northern refugia and the origins of the modern biota. *Trends Ecol. Evol.* **16**, 608–613.
Taberlet, P. & Cheddadi, R. 2002 Quaternary refugia and persistence of biodiversity. *Science* **297**, 2009–2010.
Thompson, R. S. 1990 Late Quaternary vegetation and climate in the great Basin. In *Packrat middens: the last 40 000 years of biotic change* (ed. J. L. Bettancourt, T. R. Van Devender & P. S. Martin), pp. 200–239. Tucson, AZ: University of Arizona Press.
Tzedakis, P. C., Lawson, I. T., Frogley, M. R., Hewitt, G. M. & Preece, R. C. 2002 Buffered tree population changes in Quaternary refugium: evolutionary implications. *Science* **297**, 2044–2047.
Wakeley, J. 2004 Inferences about the structure and history of populations: coalescent and intraspecific phylogeography. In *The evolution of population biology—modern synthesis* (ed. R. Singh, M. Uyenoyama & S. Jain). Cambridge University Press. (In the press.)
Willis, K. J., Rudner, E. & Sümegi, P. 2000 The full-glacial forests of Central and South-eastern Europe. *Quatern. Res.* **53**, 203–213.

GLOSSARY

cp: chloroplast
LGM: last glacial maximum
MRCA: most recent common ancestor
mt: mitochondrial
PCR–RFLP: polymerase chain reaction–restriction fragment length polymorphism