Molecular Ecology (2001) **10**, 103–111

Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations

R. WALTER and B. K. EPPERSON
126 Natural Resources Building, Michigan State University, East Lansing, MI 48824, USA

Abstract

Genetic diversity is low in natural populations of red pine, *Pinus resinosa*, a species that has a vast range across north-eastern North America. In this study, we examined 10 chloroplast microsatellite or simple sequence repeats (cpSSR) loci in 136 individuals from 10 widespread populations. Substantial variation for the cpSSR loci was observed in the study populations. The contrast with red pine's lack of variation for other types of loci is likely to be due to the higher mutation rates typical of SSR loci. The amount of variation is lower than that generally found for cpSSR loci in other pine species. In addition, the variation exhibits a striking geographical pattern. Most of the genetic diversity is among populations, with little within populations, indicating substantial isolation of and genetic drift within many populations in the southern half of the species distribution. The greatest diversity now occurs in the north-eastern part of New England, which is especially intriguing because this entire area was glaciated. Thus the centre of diversity cannot be the origin of postglacial populations, rather it is likely caused by admixture, most probably because of influences from two separate refugia. Furthermore, the pattern indicates that the spread of red pine since the last glaciation is rather more complex than usually described, and it likely includes more than one refugia, complex migration routes, and postglacial-retreat isolation and genetic drift among shrinking populations in regions of the present southern range.

Keywords: chloroplast DNA, genetic bottlenecks, geographical genetics, microsatellites, *Pinus resinosa*, population genetics

Received 8 May 2000; revision received 2 September 2000; accepted 2 September 2000

Introduction

Red pine (*Pinus resinosa* Ait.) has a broad natural range across the north-central and north-eastern United States and south-eastern Canada (Fig. 1). However, little is known about how genetic variation is distributed within and among populations, despite the species' considerable ecological importance (Spencer *et al*. 1988). It is known that red pine has unusually low genetic diversity for commonly assayed markers, and until recently no Mendelian polymorphisms had been detected. The fact that red pine has extremely low amounts of variation seems to conflict with the fact that the species consists of millions of individuals occupying a vast range. A number of studies has shown that there is some morphological variation in red pine, but it is much less than that for almost all other conifers studied. Among populations, generally ≈ 10% of morphological variation is determined genetically by source population, as determined in provenance tests (common garden experiments) (Fowler & Lester 1970; Wright *et al*. 1972). Traits studied include growth rates (Wright *et al*. 1972; Yao 1974; Guries & Ager 1980), phenology, wood quality and photoperiod (Fowler & Lester 1970), foliage phenols (Thielges 1972), branching angle (Mosseler *et al*. 1992) and seedling shoot traits (Dhir 1973), and some of these traits are correlated with latitude. Individuals of a small isolated population in Illinois (Fig. 1) produce white pollen (Brenneman 1956), a unique condition in red pine, and a rare, usually recessive, genetic condition in pines in general (Johnson & Critchfield 1974).

The low level of genetic variation in red pine has presumably been caused by one or a series of population size bottlenecks, and there may have been severe bottlenecks

Correspondence: R. Walter. Fax: 1 517 432 1143; E-mail: walterro@msu.edu

© 2001 Blackwell Science Ltd

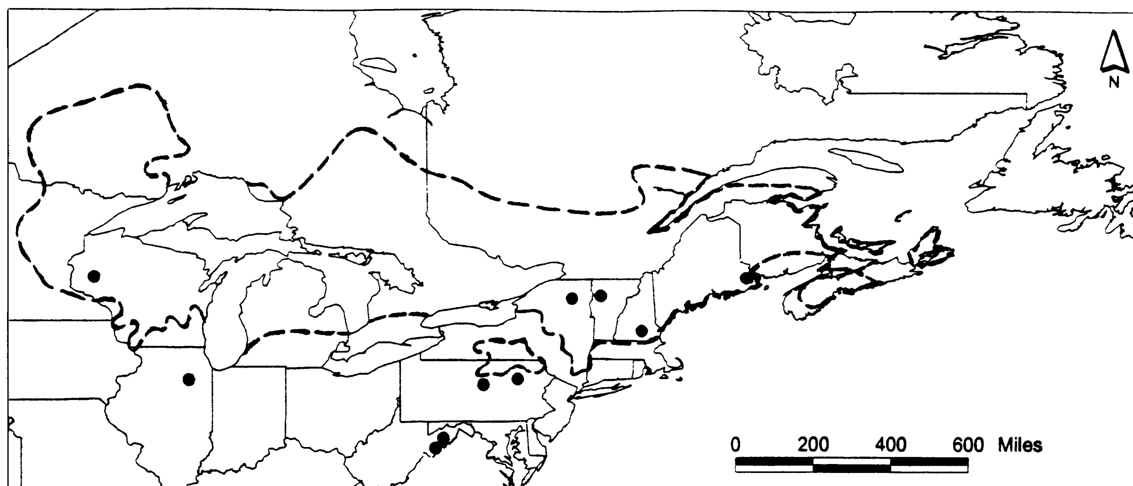

Fig. 1 Distribution map of *Pinus resinosa*; adapted from Little (1971). The dashed line represents the boundary of the main body of the range of red pine. Locations of the 10 sampled populations are represented by black dots. Dots occurring outside the main range represent (sampled) isolated populations.

during the most recent glaciation. Glaciers covered almost all of the present range of red pine. The finding of macrofossils, dating from ≈16 000 and 18 000 years BP, in northern Georgia (Jackson *et al.* 1997, 2000; S. T. Jackson personal communication) shows that red pine persisted in the southern Appalachian mountains. A second population may have existed much further north, on a series of then-exposed and nonglaciated large islands and extensions of the mainland, off the present coastline of the eastern seaboard (Grant 1977; Ives 1978; Brookes 1982; Pielou 1992). It is also possible that red pine may have persisted elsewhere, for example in the Midwest. Although the fossil pollen evidence suggests that it did not (e.g. Yeatman 1967; Delcourt & Delcourt 1981; Davis 1983; Watts 1983; Webb *et al.* 1983), it cannot preclude the possibility of relatively small populations of red pine, in part because researchers rarely distinguish between pollen from *P. banksiana* and *P. resinosa* (for an exception see Birks 1981). Even one small genetically distinct refugial population could expand greatly in size as new habitat opened up, and this could have great consequences for the succeeding patterns of genetic variation. Different scenarios lead to quite different expected genetic distributions at present, depending on the origins of ancestors and how they spread. It is possible, for example, that red pine in the north-eastern US is derived largely from the Appalachians, and the Lake States, Manitoba and far-western Ontario from a Midwest refugia. Refugial populations during the Wisconsin glaciation play the critical initial role in the distribution of genetic variation in red pine. Each of these refugia must have experienced considerable genetic drift. This means that each refugia both lost genetic variation (compared with the entire species before glaciation), and at the same time drifted away genetically (differentiated genetically) from any other refugia. Thus the species-wide genetic variation that exists today is a direct outcome of the amount of variation within and among the refugia populations. The geographical *pattern* or distribution of genetic variation today is determined by the above process combined with how far and by when the descendants expanded out of the different refugia.

The level of genetic variation for most standard types of markers is extremely low in red pine. Extensive searches for isozyme genetic markers in red pine found no polymorphism (Fowler & Morris 1977; Allendorf *et al.* 1982; Simon *et al.* 1986; Mosseler *et al.* 1991). In contrast, among studied gymnosperms (mostly conifers) an average of 71% of allozyme loci are polymorphic at the species level, with 58% polymorphic loci within populations (Hamrick & Godt 1990). In addition, Mosseler *et al.* (1992) found virtual monomorphism for a number of randomly amplified polymorphic DNAs (RAPDs), although, intriguingly, they did find a few bands that differed in Newfoundland, and these were verified by restriction fragment length polymorphism (RFLP)–RAPD analyses (DeVerno & Mosseler 1997). Very recently, polymorphism for chloroplast microsatellite or simple sequence repeat (cpSSR) loci was found (Echt *et al.* 1998). This was an important finding, but is not inconsistent with previous studies, because cpSSRs mutate at greater rates than normal loci (e.g. Di Rienzo *et al.* 1994; Provan *et al.* 1999), and hence are typically highly variable in conifers (e.g. Powell *et al.* 1995; Vendramin *et al.* 1996). Echt *et al.* (1998) studied the patterns of variation in the northern half of the present distribution range of the species and found that the highest relative levels of genetic variation occurred in New Brunswick, a location intermediate between the

Appalachian and the north-east glacial refugia area. This suggests that separate bottlenecks occurred, but red pine persisted and subsequently spread from the latter two locations, and hence the population in the area around New Brunswick is admixed.

In this study we assayed a set of populations in the southern half of the present distributional range, including remnants of refugial populations in West Virginia and an especially unique isolated tiny natural population in Illinois. We assayed 10 cpSSR loci for these. Because the chloroplast genome does not recombine and is inherited paternally in pines (e.g. Neale & Sederoff 1989; Wagner *et al*. 1992; Smith & Devey 1994), distinct lineages may be represented in the data. The genome is passed on through both pollen and seed and therefore includes dispersal of both.

Materials and methods

Sampling

Vegetative buds were collected from 136 individuals representing 10 original populations of red pine. Seven of these were from widely separated southern sources available from holdings in Michigan State University provenance (common garden) plantations (Table 1, Fig. 1). Three additional populations were sampled directly, Illinois, West Virginia/Hardy and West Virginia/Pendleton. For each population 14–20 trees were sampled, except for the two West Virginia populations (eight and ten individuals), and the Pennsylvania/Wyoming (four individuals) and Illinois populations (four individuals).

DNA isolation

DNA was isolated using a modified protocol of the method developed by Guillemaut & Maréchal-Drouard (1992). Bud tissue (≈40 mg) was disrupted using a plastic pestle in a solution of 450 μL extraction buffer [100 mM sodium acetate, 50 mM EDTA, pH 8.0, 500 mM sodium chloride, 0.5% polyvinylpyrrolidone (M_r 360 000), 1.4% SDS, 60 mM cysteine, pH 5.3]. After homogenization, the samples were incubated at 55 °C for 30 min (lysis step). The DNA was extracted with 0.5 vol. of 3.5 M potassium acetate, pH 5.3, and incubated on ice for 20 min. After centrifugation for 10 min at room temperature, the supernatant was transferred to a new microfuge tube and the DNA was precipitated with 0.6 vol. isopropanol. The DNA was redissolved in 10 mM Tris–HCl, 1 mM EDTA, pH 8.0. Quantification of the DNA was determined by spectrophotometry, and the average yield was 7 μg per extraction.

Table 1 Sample origin locations of red pine populations (state/county)

Provenance	Population code	Latitude/ Longitude
Wisconsin/St Croix	776	45°15′/92°02′
Maine/Washington	758	45°00′/67°45′
New Hampshire/Merrimack	759	43°10′/71°47′
Vermont/Chittenden	739	44°38′/73°08′
New York/Franklin	705	44°29′/74°17′
Pennsylvania/Wyoming	742	41°31′/76°09′
Pennsylvania/Clinton	732	41°16′/77°48′
West Virginia/Hardy	WVH	39°00′/78°45′
West Virginia/Pendleton	WVP	38°45′/79°15′
Illinois/LaSalle	ILL	41°32′/88°42′

Chloroplast microsatellite markers

We used 10 chloroplast mononucleotide SSR primers (PT9383, PT15169, PT30204, PT36480, PT41093, PT48210, PT71936, PT87268, PT107148 and PT110048) designed from known sequences in black pine, *Pinus thunbergii* (Vendramin *et al*. 1996). Many of these primers were used in the study by Echt *et al*. (1998); except that they included one that this study did not (PT26081) and this study added two (PT48210 and PT107148). The forward strand of each forward primer pair was labelled with one of the three phosphoramidites: 6-FAM, HEX and TET (ABI) for detection of alleles, and this permitted multiplexing, simultaneous analysis, of polymorphic loci which may overlap in size.

PCR amplification and multiplexing

PCR amplifications were carried out in a total volume of 10 μL containing 5 ng of template DNA, 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 3.5 mM magnesium chloride, 200 μM each dNTP, 200 nM of each primer, and 0.4 units of *Taq* polymerase (Perkin–Elmer). PCRs were performed on a MJ Research PTC-100 thermocycler with a touchdown amplification protocol: two cycles with a denaturing step at 94 °C for 1 min, an annealing step at 65 °C for 1 min, and an extension step at 70 °C for 35 s; followed by 18 cycles with denaturation at 93 °C for 45 s, primer annealing at 64 °C for 45 s (progressive decrease of the temperature by 0.5 °C every cycle until 55.5 °C was reached), and primer extension at 70 °C for 45 s. The last 20 cycles were with denaturation at 92 °C for 30 s, primer annealing at 55 °C for 30 s, primer extension at 70 °C for 60 s, and a final extension cycle at 70 °C for 5 min. Individual PCR products with different fluorescent labels and sizes were combined in a single lane on a 0.40-mm, 6% denaturing polyacrylamide gel (7 M urea). Samples were electrophoresed at 2500 V/30 W for 5.5–6 h to allow the 350 bp labelled size standard to be detected. In addition, an external standard was applied to standardize fragment

sizes and correct for possible gel-to-gel variations. Labelled DNA fragments were analysed on a 373 ABI sequencer using GENESCAN. Under these conditions 1 bp differences were resolved. PCRs and gel runs for all individuals which had variants from the common allele were repeated at least once, and these and all other replications gave completely consistent values for fragment sizes.

Statistical analysis

Of the 10 primers, three revealed polymorphism in our samples (PT71936, PT87268, PT110048). For these, diversity measures and *F*-statistics (Wright 1965) were computed using the Lewis & Zaykin (2000) GDA program (version 1.0 d15), calculated directly on the haplotype frequencies. We cannot formally combine our data set with that published earlier for the northern populations (Echt *et al*. 1998), because we used some different primers, and because PCR and gel conditions were considerably different, although we can make a number of strong qualitative comparisons and contrasts between the two studies. It is also worth noting that some special diversity statistics designed to exploit differences in the sizes (i.e. numbers of repeats) of alleles (e.g. R_{ST}, Goldstein *et al*. 1995; Slatkin 1995) are inappropriate for our data, because loci either had only two alleles, or at most three alleles with at least one of those in very low frequency. Thus, there cannot be substantial additional information to exploit, and hence it is not helpful or necessary to assume a specific ('stepwise') mutation model, unknown violations of which may be misleading. The following diversity statistics for the haplotypes in each population sample were calculated: number of haplotypes and effective 'heterozygosity', H_E (Weir 1996). We also calculated genetic distances between pairs of populations using the measures of Nei's (1978), and Weir's (1996) θ estimate of the theoretical parameter F_{ST}.

It was not possible to construct meaningful UPGMA dendrograms of the genetic distances for the set of populations, because some pairs had undefined distances, owing to the occurrence of fixations for opposite or identical haplotypes, and there were also many 'ties'. It is worth noting that treating the data as independent allele frequencies for alleles among specific loci would be inappropriate. All variant haplotypes observed differed from the most common by a single variant at one of the three polymorphic loci, thus, virtually all of the information on alleles is contained in the haplotypes. Moreover, alleles at different loci are completely dependent with respect to dispersal and drift, because there is no recombination. Treating haplotypes as 'alleles' is proper under the infinite alleles assumptions, and, for example, Nei's measure is linear with time, under the standard model of mutation and drift but no migration (the latter is technically invalid in most studies of genetic variation within species).

Table 2 Haplotype definitions, with fragment sizes for polymorphic cpSSRs, and measures of differentiation (θ) in the study populations

	cpSSR locus PT No.*			
Haplotype	71936	87268	110048	θ
I	149	164	95	0.57
II	148	164	95	0.39
III	149	164	96	0.26
IV	149	164	94	−0.03
V	149	162	95	−0.03
VI	149	163	95	1.00
Combined				0.56

*As identified by Vendramin *et al*. (1996).

Table 3 Frequencies and diversities of haplotypes within study populations

	Numbers of each haplotype							
Population	I	II	III	IV	V	VI	A*	H_E†
Wisconsin	19	0	0	0	1	0	2	0.097
Maine	7	3	6	0	0	0	3	0.653
New Hampshire	17	1	1	1	0	0	4	0.277
Vermont	12	2	0	0	0	0	2	0.254
New York	20	0	0	0	0	0	1	0
Pennsylvania (W.)	0	4	0	0	0	0	1	0
Pennsylvania (C.)	19	1	0	0	0	0	2	0.097
West Virginia (H.)	0	0	0	0	0	10	1	0
West Virginia (P.)	8	0	0	0	0	0	1	0
Illinois	4	0	0	0	0	0	1	0
Total/mean	106	11	7	1	1	10	1.8	0.138

*Number of different haplotypes.
†Effective heterozygosity of haplotypes.

Results

Three (for primer pairs PT71936, PT87268 and PT110048) of the 10 (30%) SSR loci were polymorphic in our samples, having either two or three alleles (Table 2). The relative sizes of fragments scored for the seven monomorphic loci are 92, 123, 140, 142, 78, 91 and 123 bp for PT9383, PT15169, PT30204, PT36480, PT41093, PT48210 and PT107148, respectively. For the total sample, the average number of alleles was 1.5 for all loci and 2.7 for polymorphic loci.

A total of six haplotypes was found, of a theoretical maximum of 18 (3 × 3 × 2) based on the numbers of alleles (Table 2). One of these, haplotype I, is particularly common, having a total frequency of 0.78 for the combined samples (Table 3). Haplotype I was the predominant haplotype in every population sample except two (742 and WVH), both of which were small samples fixed for another haplotype. All other haplotypes differ from

Table 4 Genetic distances* based on haplotype frequencies

	776	758	759	739	705	742	732	WVH	WVP	ILL
776	—	0.297	0.000	0.008	0.000	∞	0.000	∞	0.000	0.000
758		—	0.224	0.236	0.297	1.144	0.275	∞	0.297	0.297
759			—	−0.002	0.000	2.834	−0.003	∞	0.000	0.000
739				—	0.008	1.800	−0.001	∞	0.008	0.008
705					—	∞	0.000	∞	0	0
742						—	2.944	∞	∞	∞
732							—	∞	0.000	0.000
WVH								—	∞	∞
WVP									—	0
ILL										—

*Pairs denoted by ∞ were fixed for different haplotypes, thus Nei's (1978) measure is undefined. Values of 0 occurred for pairs fixed for identical haplotypes. Elsewhere numerical values are given to the third decimal.

haplotype I by a difference at a single locus, and all except one (the 162 allele of PT87268, which distinguished haplotype V) were caused by a single base pair change. Thus haplotype I is not only clearly the consensus 'sequence' it is also the most common. Haplotype II, the second most frequent (0.08), occurred in five populations centred in New England, and haplotype III (0.05) occurred in the two most easterly populations. Haplotypes IV and V were each present only in single individuals (0.01) in New Hampshire and Wisconsin, respectively. Finally, haplotype VI occurred in only the 10 individuals sampled from one of the West Virginia populations (WVH). The geographical pattern is apparent; haplotype I occurs throughout, haplotype II is centred in New England, and haplotypes III and IV have more sharply limited areas in north-eastern New England, whereas haplotypes V and VI are 'private' haplotypes in Wisconsin and West Virginia.

Genetic diversity is greatest, by a good margin, in Maine, followed by New Hampshire and Vermont, as measured by H_E (Table 3). Two other populations (Wisconsin and the more western Pennsylvania population, 732) have a single variant ($H_E = 0.097$), and the remaining five population samples are fixed for one haplotype. Similar trends are observed for the actual numbers of haplotypes present in the samples (Table 3). Mean H_E is 0.138.

The overall value of Weir's θ (Weir 1996), which is a standard measure of the theoretical value of Wright's F_{ST} measure of population differentiation under the 'island model', is 0.56, although it varies widely among haplotypes (Table 2). Extremely small values must result when a haplotype occurs in only a single individual (haplotypes IV and V). Extremely large values result for the case in which a sample is fixed for a private haplotype (WVH for haplotype VI). Intermediate values result for haplotypes II and III, because, although rare, they occur in more than one population. These results for haplotypes II–IV indicate that the data do not fit the equilibrium island model, rather there is spatial pattern indicating a lack of equilibrium. The value of θ for haplotype I (0.57) is probably a reasonable estimate of the overall degree of differentiation, arising from the combination of mutation, drift and migration, and it is very close to the estimate for all haplotypes combined.

The same pattern of differentiation of haplotypes is reflected in Nei's (1978) genetic distances based on differences in haplotype frequencies between pairs of populations (Table 4). Three populations (WVP, ILL and 705) were fixed for haplotype I and thus there is no genetic distance among them. Because WVH is fixed for a private haplotype (VI) it has undefined distance with the other populations. There are a few other cases in which the pair is fixed for nonidentical haplotypes. Among the remaining pairs there is a wide range of distances, from slightly negative estimates up to 2.944, and the smaller values tend to occur in those pairs which involve at least one population with relatively high genetic diversity (758, 759 and 739), or pairs in which both are nearly fixed for haplotype I. Most remarkable is the quite large values between polymorphic populations in New England, despite short distances of separation. In total, this pattern indicates admixed areas centred around north-eastern New England, with other populations forming satellites, some of which are fixed or almost so for haplotype I or fixed for haplotype VI (WVH).

Discussion

Substantial variation for chloroplast microsatellite (SSR) loci was observed in the study populations, despite red pine's lack of variation for other types of loci, and the difference is likely to be due to the higher mutation rate typical of SSR loci. However, the amount of variation is much lower than that for cpSSR loci in other pine species (e.g. Powell *et al*. 1995), and our cpSSR data clearly indicate

that the species as a whole has lost variation through bottlenecks. Moreover, the variation indicates strong genetic differentiation with a striking geographical pattern, and this pattern indicates that the spread of red pine since the last glaciation is rather more complex than is usually described. Most importantly, the pattern suggests that there were at least two refugia that contributed substantially to the present variation of red pine. However, our results indicate that the expansion must have involved many processes that affected genetic variation, including population growth, details about the paths of dispersal and exactly how dispersal proceeded, mutations, and likely regionalized or localized population contractions in the south, after expansion. The genetic aspects of expansions could be influenced by many factors. For example, if expansion was typically dominated by low-frequency, long-distance seed dispersal, much greater loss of genetic variation would precipitate compared with the alternative of essentially the edge of a red pine dominated forest moving via massive, although shorter distance, seed dispersal. The point is, as the glaciers retreated, and environmental factors became favourable, precisely how did red pine spread to those areas? The pollen fossil record suggests that the range of red pine (and many other tree species) spread an average of 200–300 m per year (e.g. Davis 1983), and this seems near the limit potential for dispersal. Thus, this pattern of dispersal could be expected to involve events of relatively long-distance (e.g. Epperson & Allard 1989; Epperson 1992, 1993) dispersal of seed and hence subsequent localized founder effects on genetic variation.

Three of the 10 loci were polymorphic in the total of study populations. Over the entire set of study populations, the level of variation is similar to that observed for red pine in the study by Echt *et al*. (1998). For the eight loci that were common to both studies, PT9383, PT15169, PT30204, PT36480, PT41093, PT71936, PT87268 and PT110048, Echt *et al*. (1998) found that all were polymorphic with an average of 2.7 alleles (22/8). However, most of the difference is caused by alleles that are unique or extremely rare. Similarly, based on these same eight loci, we found six haplotypes, compared with their 19. Again most of the additional haplotypes they found are very rare, and for example we observed an H_E value for the total sample of 0.375, which is comparable with their value of 0.570 (excluding the one polymorphic locus we did not assay). In contrast, average H_E within populations was 0.138 in our study populations, and this is much smaller than in the north, 0.543 (Echt *et al*. 1998). In addition, we found a very large amount of differentiation among our study populations, total $\theta = 0.56$, and this is much larger than the value of the closely related estimator (Nei 1987), $G_{ST} = 0.12$, found in the study by Echt *et al*. (1998). Some of the difference may be due to extremely small population sizes ($N = 4$ in the Illinois population) or sample sizes ($n = 4$, in population 742), but heterozygosity is also low within other of our 'southern' populations (e.g. WVH, WVP and 705) and many are highly differentiated from the others. Thus, overall, we see levels of genetic variation in our study populations similar to those in the more northern populations, and it is certainly not greater, although it could be slightly lower. However, the variation is structured differently in the southern populations, it is more differentiated among populations, with less variation within populations. In those cases in which our populations are relatively near the study populations of Echt *et al*. (1998) we find very similar amounts of variation within and between populations. Indeed, this is also observed on an allelic basis (discussed below).

Among the study populations, far greater variation is found in the north-eastern regions of New England, i.e. in Maine, New Hampshire and Vermont. This is especially intriguing because this entire area was glaciated. Moreover, among the study populations of Echt *et al*. (1998), the one with the greatest diversity was in New Brunswick, the nearest to these three. Throughout the species range, the greatest diversity now occurs in an area that was glaciated. Thus the centre of diversity cannot be the origin of modern populations, as often presumed in genetic studies. For example, in modern humans, the centre of diversity is considered as supportive evidence of the Out of Africa Hypothesis (e.g. Stoneking *et al*. 1997), although such data have recently been interpreted by some as evidence for admixture from a more recent immigration into Africa (e.g. Hammer *et al*. 1998; Epperson 1999). The simplest explanation for the pattern in red pine is that the region has become admixed, from red pine descended from the Appalachians and those descended from a second refugia, most likely somewhere in the glacial refugial area off the north-eastern seaboard, for example just east of New Brunswick.

One of the most striking features of the pattern is the preponderance of haplotype I throughout the study populations, and its near-monomorphism in most, especially in the more southern and western populations (Wisconsin, Illinois, and one each of the populations in Pennsylvania and West Virginia, and that in New York). This pattern suggests a common lineage. Moreover, as discussed below, haplotype I almost certainly corresponds to 'haplotype II' of Echt *et al*. (1998), and the latter haplotype predominates in a north-western region extending throughout Ontario and the Great Lakes States.

Our results reveal sometimes remarkable differentiation on a very small spatial scale. One example is the contrast between the two Pennsylvania populations. More important is the differentiation of the two West Virginia populations. One population is fixed, or almost fixed, for the common haplotype I, the other is fixed, or almost

fixed, for a private haplotype, VI. The two haplotypes differ by a single allele variant. This indicates one of two possibilities. The first is that the populations of the Appalachian refugium area were differentiated, and the individuals that successfully dispersed northward were a subset of those, and more closely related to individuals today in Pendleton county than those in Hardy county, West Virginia. Alternatively, the Hardy county population experienced a recent, essentially postglacial mutation, which has become frequent through genetic drift. In either case, the two populations have had little genetic exchange for a considerable period, either since before the end of the glaciation, or since the mutation. Clearly many of the southern populations have experienced considerable genetic drift with relatively little gene flow among them.

It is noteworthy that the Illinois population has no detected difference from the others, it is fixed for haplotype I. Thus there is no evidence that there was a separate refugia in the central or Midwestern US that had genetic influence on the present distribution. However, it should be pointed out that the Illinois population must have been quite small and isolated for a very long time, probably thousands of years. Although we cannot know the exact size of the population during that time, up to the European settlement, at present the population consists of four trees, separated from any other red pine populations by hundreds of kilometres. Thus, given the widespread nature of haplotype I, even if other signature variants of another refugia were present there would be considerable chance of fixation for haplotype I.

Because we used different gel conditions, in particular different internal size standards (which we know affects the absolute but not the relative sizes of alleles within our study) and binning procedures, we cannot match our alleles with those in the study by Echt *et al*. (1998), based on scored fragment sizes alone. However, both studies showed predominating alleles study-wide, and thus we may presume that our most common alleles are the same. Consequently, our alleles 92, 123, 140, 142, 78, 149, 164 and 95 would match their alleles 91, 122, 140, 141, 78, 147, 163 and 95, for PT9383, PT15169, PT30204, PT36480, PT41093, PT71936, PT87268 and PT110048, respectively. Note that in all cases except one our values are the same (small fragments), or 1 bp larger (large fragments). The exception, for PT71936, has our 149 compared with their 147, and it is possible that our allele 148 is their 147 because our 148 is concentrated in the more north-eastern samples among ours. In any case, this one exception does not affect our conclusions.

Given that haplotype I dominates elsewhere, the high incidence of variant types in and around New Hampshire, Maine and New Brunswick is particularly striking. Moreover, if we assume the allele assignments between the two studies, we find the same allele variants in this region in both studies. Our haplotype III is a +1 bp variant of PT110048 found only in New Hampshire and Maine. Echt *et al*. (1998) found three haplotypes with the same +1 bp variant, one of which was found only in New Brunswick. The other two, which may represent a separate mutation, were found in low frequencies in Wisconsin and nearby in western Ontario. Our haplotype IV, which was only found in the New Hampshire sample, is a –1 bp variant of the same locus, and the only corresponding variant in the other study occurred in two haplotypes that were found only in Nova Scotia. The –1 bp variant of PT71936, which characterized haplotype II, a New England region haplotype, is more difficult to interpret, because it was found in many haplotypes in the study by Echt *et al*. (1998). The –2 bp variant of PT87268 found in a single tree in the Wisconsin sample was not found in the study populations of Echt *et al*. (1998). Finally, the –1 bp variant we found fixed in the Hardy county population of West Virginia corresponds to a variant found in two haplotypes, each in a single tree, in the sample of Echt *et al*. (1998).

There seem to be only two plausible explanations for the observations of relatively high variation in a constrained area centred near Maine and much lower levels of variation everywhere else. One is that there was a separate red pine population in the north-eastern seaboard refugial area, and genetic variation is the result of admixture between that and the spreading populations descended from the refugial area in the Appalachians. The other possibility is that in one way or another genetic variation was lost elsewhere but for some reason maintained in the area around Maine. It is somewhat likely that some of the other populations, particularly those in our study that are highly isolated today in the south, could have lost genetic variation due to small population sizes, but it is more difficult to understand patterns in some of the other populations, e.g. New York and Wisconsin. It becomes even more complicated to explain the lack of variation throughout Ontario and the Great Lakes States observed in the study by Echt *et al*. (1998), and more so if one considers that the types in Ontario are almost certainly the same as those in the southern populations. Today, there are very large populations of red pine throughout much of Ontario and the Great Lakes States, which now must experience negligible genetic drift. Moreover, even if these populations lost genetic variation as they moved westward from New England, chances are that they would not have become nearly fixed for the same haplotypes as found in one of the West Virginia populations and elsewhere in the southern range today. The pattern of genetic distances among pairs of study populations seems to favour the admixture hypothesis. The set of values is multimodal and wide-ranging, and this is opposite to expectations from models of population expansion or contractions (Slatkin & Hudson 1991).

Acknowledgements

The authors thank Jim Colbert for collecting the two West Virginia samples, Paul Bloese, Greg Kowalewski and John Vigneron for field collections, Craig Echt for providing primers and for helpful comments on the manuscript, Anne Plovanich-Jones, Shari Tjugum-Holland and Tom Newman for help on various molecular analysis, and Eric Myers for help with statistical analyses. The work was supported in part by McIntire-Stennis project #1774, the Michigan State University Biotechnology Center, and MSU Plant Breeding and Genetics program, and the USDA Forest Service North Central Research Station (23–99–01-RJVA).

References

Allendorf FW, Knudsen KL, Blake GM (1982) Frequencies of null alleles at enzyme loci in natural populations of ponderosa and red pine. *Genetics*, **100**, 497–504.

Birks HJB (1981) Late Wisconsin vegetational and climatic history at Kylen Lake, northeastern Minnesota. *Quaternary Research*, **16**, 322–355.

Brenneman WS (1956) Red pine indigenous to Illinois? *Journal of Forestry*, **54**, 775.

Brookes IA (1982) Ice marks in Newfoundland: a history of ideas. *Geographie physique et quarternaire*, **36**, 139–163.

Davis MB (1983) Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden*, **70**, 550–563.

Delcourt PA, Delcourt HR (1981) Vegetation maps for eastern North America: 40,000 B.P. to the present. In: *Geobotany I* (ed. Romans R), pp. 123–166. Plenum, New York.

DeVerno LL, Mosseler A (1997) Genetic variation in red pine (*Pinus resinosa*) revealed by RAPD and RAPD-RFLP analysis. *Canadian Journal of Forest Research*, **27**, 1316–1320.

Dhir NK (1973) Geographic variation and relationships among shoot characteristics of red pine seedlings. *Minnesota Forestry Research Notes*, **244**, 1–4.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Science of the USA*, **91**, 3166–3170.

Echt CS, DeVerno LL, Anzidei M, Vendramin GG (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology*, **7**, 307–316.

Epperson BK (1992) Spatial structure of genetic variation within populations of forest trees. *New Forests*, **6**, 257–278.

Epperson BK (1993) Recent advances in correlation studies of spatial patterns of genetic variation. *Evolutionary Biology*, **27**, 95–155.

Epperson BK (1999) Gene genealogies in geographical genetics. *Genetics*, **152**, 797–806.

Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics*, **121**, 369–377.

Fowler DP, Lester DT (1970) *Genetics of red pine*. U.S. Forest Service, Washington Office Research Paper WO-8.

Fowler DP, Morris RW (1977) Genetic diversity in red pine: evidence for low genic heterozygosity. *Canadian Journal of Forest Research*, **7**, 343–347.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.

Grant DR (1977) Glacial style and ice limits, the Quaternary stratigraphic record, and changes of land and ocean level in the Atlantic provinces Canada. *Geographie physique et quarternaire*, **31**, 247–260.

Guillemaut P, Maréchal-Drouard L (1992) Isolation of plant DNA: a fast, inexpensive, and reliable method. *Plant Molecular Biology Report*, **10**, 60–65.

Guries R, Ager A (1980) Red pine seedling seed orchards: 10 year results. *Forestry Research Notes, University of Wisconsin*, **242**, 1–4.

Hammer MF, Karafet T, Rasanayagam A *et al.* (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Molecular Biology and Evolution*, **15**, 427–441.

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates, Sunderland, MA.

Ives JD (1978) The maximum extent of the Laurentide ice sheet along the east coast of North America during the last glaciation. *Arctic*, **31**, 24–53.

Jackson ST, Overpeck JT, Webb T, Keattch SE, Anderson KH (1997) Mapped plant-macrofossils and pollen records of late Quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, **16**, 1–70.

Jackson ST, Webb RS, Anderson KH, Overpeck JT, Webb T, Williams JW, Hansen BCS (2000) Vegetation and environment in eastern North America during the last glacial maximum. *Quaternary Science Reviews*, **19**, 489–508.

Johnson LC, Critchfield WB (1974) A white-pollen variant of bristle cone pine. *Journal of Hereditry*, **65**, 123.

Lewis PO, Zaykin D (2000) *Genetic Data Analysis: Computer program for the analysis of allelic data, Version 1.0 (d15)*. Free program distributed by the authors over the internet from the GDA home page http://alleyn.eeb.uconn.edu/gda/

Little EL (1971) *Atlas of United States Trees, Vol. 1. Conifers and Important Hardwoods*. USDA Forest Service Miscellaneous Publication, 1146.

Mosseler A, Egger KN, Hughes GA (1992) Low levels of genetic diversity in red pine confirmed by random amplified polymorphic DNA markers. *Canadian Journal of Forest Research*, **22**, 1332–1337.

Mosseler A, Innes DJ, Roberts BA (1991) Lack of allozymic variation in disjunct Newfoundland populations of red pine (*Pinus resinosa*). *Canadian Journal of Forest Research*, **21**, 525–528.

Neale DB, Sederoff RR (1989) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics*, **77**, 2112–2216.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Pielou EC (1992) *After the Ice Age. The Return of Life to Glaciated North America*. University of Chicago Press, Chicago.

Powell W, Morgante M, McDevitt R, Vendramin G, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences of the USA*, **92**, 7759–7763.

Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999) A low mutation rate for chloroplast microsatellites. *Genetics*, **153**, 943–947.

Simon J-P, Bergeron Y, Gagnon D (1986) Isozyme uniformity in red pine (*Pinus resinosa*) in the Abitibi Region, Quebec. *Canadian Journal of Forest Research*, **16**, 1133–1135.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequency. *Genetics*, **139**, 457–462.

Slatkin M, Hudson R (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome*, **37**, 977–983.

Spencer JS Jr, Smith WB, Hahn JT, Raile GK (1988) *Wisconsin's Fourth Forest Inventory, 1983 USDA-Forest Service*. Resource Bulletin NC-107.

Stoneking M, Fontius JJ, Clifford SL *et al.* (1997) *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Research*, **7**, 1061–1071.

Thielges BA (1972) Intraspecific variation in foliage polyphenols of *Pinus* (Subsection *Sylvestris*). *Silvae Genetica*, **2**, 114–119.

Vendramin GG, Lelli LR, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology*, **5**, 595–598.

Wagner DB, Nance WL, Nelson CD *et al.* (1992) Taxonomic patterns and inheritance of chloroplast DNA variation in a survey of *Pinus echinata*, *Pinus elliottii*, *Pinus palustris*, and *Pinus taeda*. *Canadian Journal of Forest Research*, **22**, 683–689.

Watts WA (1983) Vegetational history of the eastern United States 25 000–10 000 years ago. In: *Late-Quaternary Environments of the United States, Vol. 1. The Late Pleistocene* (ed. Wright HE Jr), pp. 294–310. University of Minnesota Press, Minneapolis.

Webb T, Cushing EJ, Wright HE (1983) Holocene changes in vegetation of the Midwest. In: *Late-Quaternary Environments of the United States, Vol. 2. The Holocene* (ed. Wright HE Jr), pp. 142–165. University of Minnesota Press, Minneapolis.

Weir BS (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Wright JW, Read RA, Lester DT, Merritt C, Mohn C (1972) Geographic variation in red pine: 11-year data from the North Central states. *Silvae Genetica*, **21**, 205–222.

Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.

Yao YN (1974) Improved red pine for Michigan. *Quarterly Journal of Chinese Forestry*, **7**, 19–29.

Yeatman CW (1967) Biogeography of jack pine. *Canadian Journal of Botany*, **45**, 2201–2211.

Rosemarie Walter, a research associate in the Department of Forestry, Michigan State University, develops and applies molecular markers for forest genetics. Bryan Epperson, from the same institution, studies theoretical and statistical aspects of geographical genetics, and uses molecular markers to study the spatial population genetics of trees and other species.