Molecular Ecology (1998) **7**, 307–316

# Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait.

C. S. ECHT,* L. L. DEVERNO,† M. ANZIDEI‡ and G. G. VENDRAMIN‡
*U.S.D.A. Forest Service, North Central Forest Experiment Station, 5985 County Road K, Rhinelander, Wisconsin 54501, USA, †Canadian Forest Service – Atlantic Forestry Centre, PO Box 4000, Fredericton, New Brunswick, E3B 5P7, Canada, ‡Istituto Miglioramento Genetico Piante Forestali, C.N.R., Via Atto Vannucci 13, 50135 Firenze, Italy

**Abstract**

**Variation in paternally inherited chloroplast microsatellite (cpSSR) DNA was used to study population genetic structure in red pine (*Pinus resinosa* Ait.), a species characterized by morphological uniformity, no allozyme variation, and limited RAPD variation. Using nine cpSSR loci, a total of 23 chloroplast haplotypes and 25 cpSSR alleles were found among 159 individuals surveyed in seven widely separated populations. The total genetic diversity, $H_T$, was 0.618, but haplotype differentiation among populations was low ($G_{ST}$ = 0.121). All populations were distinguished from each other by their haplotype compositions, and only one haplotype was common among all populations. Based on average squared composite cpSSR length differences (stepwise haplotypes), within-population diversity was relatively high for only one population ($D^2_{SH}$ = 0.443). Frequency distributions of pairwise SSR differences among individuals within different populations, as well as branch length differences in neighbour-joining dendrograms, indicated recovery from one or more population bottlenecks, and may be explained by metapopulation dynamics.**

*Keywords:* biodiversity, cpSSRs, forests, haplotypes, pines, simple sequence repeats

*Received 24 July 1997; revision received 22 October 1997; accepted 22 October 1997*

## Introduction

Red pine (*Pinus resinosa* Ait.) is a forest tree species endemic to eastern North America that appears genetically uniform by several measures. Although average allozyme genetic diversity for gymnosperms is the same as that for monocotyledonous species ($H_{ES} \approx 0.18$) (Hamrick & Godt 1990), no allozyme diversity in red pine has been detected (Fowler & Morris 1977; Allendorf *et al*. 1982; Simon *et al*. 1986; Mosseler *et al*. 1991). RAPD markers, which are usually highly polymorphic in conifers, have revealed very limited genetic diversity among red pines (Mosseler *et al*. 1992; DeVerno & Mosseler 1997). The high level of homozygosity observed in red pine could have resulted from one, or a series of, population bottleneck(s) (Nei *et al*. 1975). The most recent drastic decrease in population size to have affected the species is thought to have occurred during the last Pleistocene glaciation 20 000 years ago, when red pine was restricted to refugial populations in the Appalachian highlands of present day West Virginia (Fowler & Morris 1977). The disjunct, dispersed population structure found within the species' current range promotes inbreeding which further increases homozygosity (Fowler & Lester 1970; Mosseler 1992). This type of population structure, and low genetic diversity, are characteristics of a metapopulation, in which subpopulations have restricted gene flow and undergo periodic and localized colonizations, bottlenecks and extinctions (Pimm *et al*. 1989; Hedrick & Gilpin 1997). A nonequilibrium metapopulation structure of red pine, characterized by an excess of local extinctions over local colonizations, has been proposed by Mosseler (1992).

In contrast to most pines, red pine also exhibits very little inbreeding depression (Fowler 1964, 1965a,b,c), suggesting that most deleterious alleles were purged through founder effects, successive generations of inbreeding, and selection. Some variation of growth and reproductive traits among provenances has been reported, but it is much less than that observed for other northern pines such as *P. strobus* and *P. banksiana*, and the heritability of the variation has not been established (Fowler 1965c; Wright *et al*. 1972; Ager *et al*. 1983). Even though a few novel mutant forms are known (Mosseler 1992), morphological and phenological uniformity are considered general

Correspondence: Correspondence: C. S. Echt. Fax: +01-(715)-362-1166; E-mail: cecht@newnorth.net

characteristics of the species (Fowler 1964; Fowler & Lester 1970). A narrow genetic base puts red pine at risk of extinction from exotic pests, disease epidemics, or rapid climate change. Thus there is a need to efficiently identify sources of genetic diversity that may exist so that divergent germplasm may be preserved *in situ*, as well as in seed orchards where it could be utilized in tree improvement programs.

Because of the limited genetic variation reported for red pine, we sought a highly sensitive DNA marker approach for measuring population diversity among red pine stands. Microsatellites, or simple sequence repeats (SSRs), are among the most polymorphic marker systems in current use because of their high mutation rates (Dallas 1992; Weber & Wong 1993; Di Rienzo *et al*. 1994; Ellegren 1995). SSR markers also generally undergo stepwise mutations, where the unit of variation is one repeat unit, thus providing a linear relationship between genetic distance based on allelic length differences and time of divergence (Di Rienzo *et al*. 1994; Slatkin 1995; Goldstein *et al*. 1995a,b; Feldman *et al*. 1997). Most mononucleotide repeats in the chloroplast genomes of conifer species are highly polymorphic within and among populations (Powell *et al*. 1995; Vendramin *et al*. 1996; Vendramin & Ziegenhagen 1997; Morgante *et al*. 1997), although information on the specific rate of mutation of the repeats is not available.

For the special case of detecting genetic variation in populations that have passed through population bottlenecks, DNA markers for haploid, uniparentally inherited genomes are more sensitive indicators of large reductions in population size because they have a twofold smaller effective population size than diploid nuclear genomes in monecious species (Mitton 1994). Haploid genomes that are transmitted through only one parent also retain a clonal record of new mutations, whereas this record is obscured in genomes that recombine the genetic contributions of both parents in each generation.

As chloroplasts have haploid genomes that are paternally inherited in pines (Neale & Sederoff 1989; Wagner *et al*. 1992; Dong & Wagner 1994; Cato & Richardson 1996; Watano *et al*. 1996), we used pine chloroplast SSR (cpSSR) markers (Powell *et al*. 1995; Cato & Richardson 1996; Vendramin *et al*. 1996) in a preliminary survey of seven red pine populations distributed across the natural range of the species. Here we report the first evidence of molecular variation in a red pine organellar genome, and demonstrate population genetic diversity that is consistent with a metapopulation structure.

## Materials and methods

### *Plant material*

Seeds collected in bulk from seven natural populations of red pine (Table 1, Fig. 1) were obtained from the National Tree Seed Centre in Fredericton, New Brunswick, Canada. For each population, DNA from 21 to 24 randomly sampled seedlings was isolated using a modification of the procedure described by Rogers & Bendich (1985). Cotyledons and hypocotyls were ground in 2× CTAB extraction buffer [2% hexadecylcetyltrimethylammonium bromide (w/v), 100 mM Tris-HCL (pH 8.0), 25 mM EDTA, 1.4 mM $MgCl_2$, 1% polyvinyl pyrrolidone (44 000 MW)], plus a small amount of aluminium oxide to aid grinding. After homogenization, 1/10 volume of 10% *N*-lauryl-sarcosine was added, the mixture was heated to 65 °C for 15 min, then extracted with an equal volume of chloroform:isoamyl alcohol (24:1). The DNA was precipitated with ethanol from the cleared aqueous supernatant, then resuspended in TE [10 mM Tris-HCl, 1 mM EDTA (pH 8.0)], plus 10 ng/μl RNase.

### *cpSSR markers*

Because the chloroplast genome does not genetically recombine, or exist in a heterozygous state, it may be viewed as a single 'locus' and all sequence variation interpreted as giving rise to different haplotypes of the genome. The chloroplast genome may alternatively be viewed as a circular haploid chromosome wherein sequence variation generates different alleles within individual, nonrecombining loci. In either case, for ease of presentation and discussion, we use the terms 'locus' to refer to a cpSSR site (defined by the termini of a PCR primer pair), and 'alleles' to refer to length variants at a cpSSR site.

**Table 1** Locations of *Pinus resinosa* populations

| Code | Provenance | Seedlot no. | Latitude/longitude | Elevation (m) |
|---|---|---|---|---|
| A | Nova Scotia, Beaver Lake | 7010280 | 44.14/65.20 | 140 |
| B | Ontario, Eldridge Township | 7030250 | 47.00/79.30 | 310 |
| D | Ontario, Sioux Lookout | 6830060 | 50.04/91.57 | 370 |
| E | Quebec, Norway Bay | 7023040 | 45.32/76.26 | 80 |
| F | Michigan, Delta Co. | 5780350 | 46.00/87.00 | NA |
| G | New Brunswick, Tracy | 7010310 | 45.43/66.42 | 60 |
| H | Ontario, Macdiarmid | 7030260 | 49.18/88.05 | 370 |

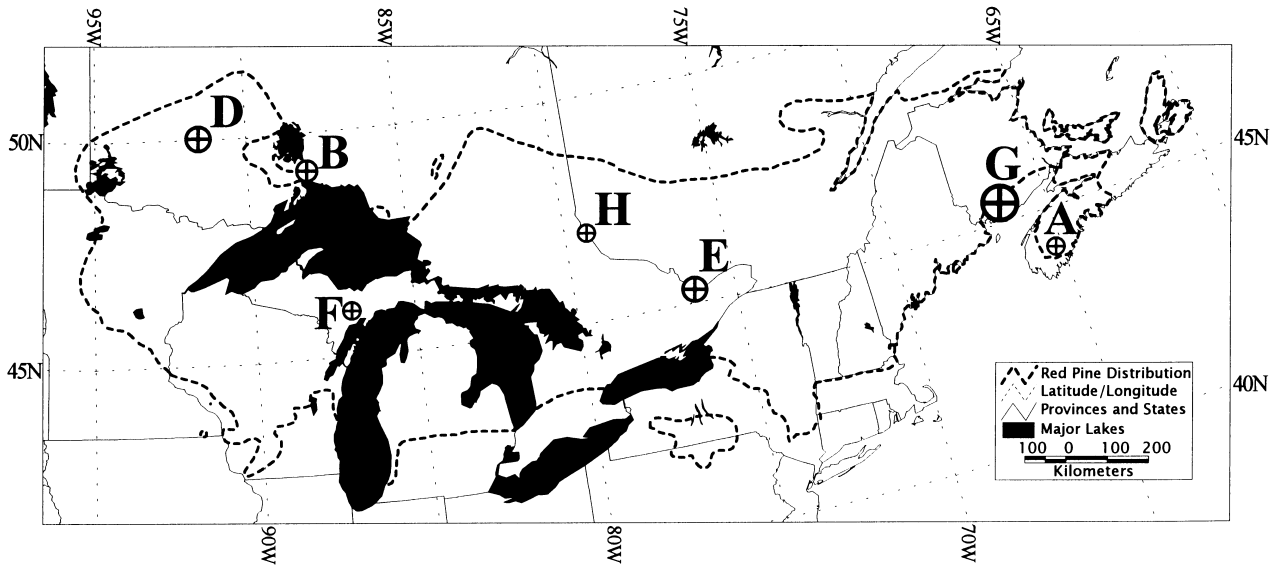


**Fig. 1** Map showing the extent of *Pinus resinosa* in North America, along with locations of surveyed populations. (See Table 1 for key to population codes.) Refugial populations are not shown. The diameters of the site markers are proportional to the unbiased estimate of diversity, $H_E$, found in each population.

Nine cpSSR primer pairs derived from the *Pinus thunbergii* chloroplast genome sequence were used for this study; PT9383, PT15169, PT26081, PT30204, PT36480, PT63718, PT71936, PT87268, and PT110048 (Vendramin *et al*. 1996; http://s27w007.pswfs.gov/Data/chloroplast.html). PCR amplifications were performed by the method of Vendramin *et al*. (1996), and the PCR-amplified fragments were assayed on an automated laser fluorescence (ALF) DNA sequencer (Pharmacia). Amplified fragments from three cpSSR primer pairs producing fragments in three different size ranges were pooled and loaded along with internal size standards (50, 100, 150 and 200 bp). Markers were separated on a 6% denaturing (7 M urea) polyacrylamide gel (0.35 mm thick) at 35 W constant power for ≈ 80 min. Fragment sizes were calculated using the computer program FRAGMENT MANAGER version 1.1 (Pharmacia) by comparison with internal as well as external standards. Amplification reactions and sample mixtures were prepared using a Biomeck 2000 robotic workstation (Beckman).

### *Statistical analyses*

Within-population diversity based on microsatellite data was estimated by the $D^2_{SH}$ measure, which is based on the stepwise mutation model distance measure of Goldstein *et al*. (1995a), but applied to the specific case of plastid microsatellites (Morgante *et al*. 1997). For this diversity measure, the chloroplast genome was regarded as a single nonrecombining locus and repeat length differences between individuals were summed over all cpSSR sites. We will refer to this as the stepwise haplotype approach. $D^2_{SH}$ is the average squared sum of all length differences at cpSSR loci among all pairs of individuals in a population,

$$D^2_{SH} = 1/m \sum_{i<i'} \left[ \sum_{k=1}^{m} \text{abs}(a_{ik} - a_{i'k}) \right]^2$$

where $m$ is the number of polymorphic SSR loci and $a$ is the basepair length of the marker at SSR locus $k$ in the $i$-th and $i'$-th individuals.

Genetic diversity measures were also calculated from haplotype frequency data. For recording haplotype frequencies, each unique combination of cpSSR alleles, across all cpSSR loci, was scored as a different haplotype. SSR repeat lengths were thus not a factor in genetic distance measures based on haplotype frequencies, as they were in the $D^2_{SH}$ measure. Haplotype variation within populations was calculated by estimating the effective number of haplotypes ($n_e$), computed as $n_e = 1/(\sum p_i^2)$, where $p$ is the frequency of the $i$-th haplotype in a population, and by estimating the unbiased genetic diversity ($H_E$), which accounts for small population sizes, computed as $H_E = (n/(n-1))\,(1 - \sum p_i^2)$, where $n$ is the number of individuals analysed in a population and $p$ is the frequency of the $i$-th haplotype in a population (Nei 1987). The amount of genetic differentiation among populations was estimated using the $G_{ST}$ measure (Nei 1987). Genetic differentiation based on stepwise SSR differences at individual SSR loci was expressed using $R_{ST}$ (Slatkin 1995).

Genetic distances among populations were determined using two measures, $D_a$ (Nei *et al*. 1983; Takezaki & Nei 1996) and $(\delta\mu)^2$ (Goldstein *et al*. 1995b). $D_a$ distances

are independent of the mutation model (Nei 1987; Takezaki & Nei 1996), while $(\delta\mu)^2$ distances are based on a stepwise mutation model of SSRs. For the $D_a$ estimates, frequency data were scored both for alleles of individual cpSSR loci and for individual haplotypes, while for $(\delta\mu)^2$ only cpSSR allele data were used. Genetic distances were calculated using the programs MICROSAT 1.5 (Eric Minch, Stanford University, USA) and NJBAFD (Naoko Takezaki, National Institute of Genetics, Japan). Neighbour-joining analyses (Saitou & Nei 1987) were conducted with the NJBAFD program and the resulting dendrograms were redrawn using the DRAWTREE and DRAWGRAM modules of the PHYLIP package (Felsenstein 1995).

## Results

### *Gene diversity*

The nine cpSSR loci examined were all variable, giving a total of 25 cpSSR alleles and 23 chloroplast haplotypes among the 159 individuals surveyed (Table 2). A total of six private alleles, i.e. alleles existing only in one of the populations sampled, were found in populations A (PT110048:94 bp), B (PT30204:142 bp), D (PT26081:114 bp, PT36480:142 bp), and G (PT87268:164 bp and PT9383:93 bp). The most frequent haplotype was haplotype II, which was also the only haplotype found in all populations (Table 3).

Population G, from Tracy, New Brunswick, was the most variable and most divergent population, as revealed by all estimated parameters (Tables 4 and 5). Compared to the other six populations, the effective number of haplotypes, $n_e$, was three to six times greater for population G, while $D^2_{SH}$ was three to eight times greater (Table 4). When considering chloroplast haplotype frequencies, the mean total genetic diversity, $H_T$, was 0.618, while the mean diversity within populations, $H_S$, was 0.543, and the mean diversity among populations, $D_{ST}$, was 0.075. The proportion of gene (haplotype) differentiation residing among populations, $G_{ST}$, was 12%, leaving 88% of the total haplotype variation common to all populations. No two populations were composed of identical haplotypes. When considering cpSSR allele frequencies, the proportion of genetic differentiation residing among populations, $R_{ST}$, was 6.8%.

### *Genetic distances and bottleneck effects*

Genetic distances between populations are presented in Table 5. Phenetic trees obtained from $D_a$ distances, using both chloroplast haplotype frequencies and cpSSR allele frequencies, are presented in Fig. 2. As judged by inspection of Figs 1 and 2, there was no discernible correlation between

**Table 2** Haplotypes and marker sizes (bp) found in *Pinus resinosa*

| | cpSSR locus, PT No.* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Haplotype | 9383 | 15169 | 26081 | 30204 | 36480 | 41093 | 71936 | 87268 | 110048 |
| I | 92 | 122 | 113 | 140 | 141 | 78 | 147 | 163 | 95 |
| II | 91 | 122 | 113 | 140 | 141 | 78 | 147 | 163 | 95 |
| III | 91 | 122 | 113 | 140 | 141 | 78 | 147 | 163 | 94 |
| IV | 92 | 122 | 113 | 141 | 141 | 78 | 147 | 163 | 94 |
| V | 91 | 122 | 113 | 140 | 141 | 78 | 148 | 163 | 95 |
| VI | 91 | 122 | 113 | 142 | 141 | 79 | 147 | 163 | 95 |
| VII | 91 | 122 | 113 | 140 | 141 | 78 | 147 | 163 | 96 |
| VIII | 91 | 122 | 112 | 140 | 141 | 78 | 147 | 163 | 95 |
| IX | 91 | 122 | 113 | 140 | 141 | 78 | 146 | 163 | 95 |
| X | 91 | 124 | 113 | 140 | 141 | 78 | 147 | 163 | 95 |
| XI | 91 | 122 | 114 | 140 | 141 | 78 | 147 | 163 | 96 |
| XII | 91 | 122 | 113 | 141 | 141 | 78 | 147 | 163 | 95 |
| XIII | 92 | 122 | 113 | 140 | 141 | 78 | 148 | 163 | 95 |
| XIV | 92 | 122 | 113 | 141 | 141 | 78 | 147 | 163 | 95 |
| XV | 91 | 121 | 113 | 140 | 141 | 78 | 147 | 163 | 95 |
| XVI | 91 | 122 | 113 | 140 | 141 | 78 | 147 | 162 | 95 |
| XVII | 91 | 124 | 112 | 140 | 141 | 78 | 147 | 163 | 95 |
| XVIII | 93 | 122 | 113 | 140 | 141 | 78 | 147 | 164 | 95 |
| XIX | 91 | 124 | 113 | 140 | 141 | 78 | 147 | 163 | 96 |
| XX | 93 | 122 | 113 | 140 | 141 | 78 | 146 | 164 | 95 |
| XXI | 91 | 124 | 112 | 140 | 141 | 78 | 146 | 162 | 95 |
| XXII | 91 | 122 | 112 | 140 | 141 | 78 | 146 | 163 | 95 |
| XXIII | 92 | 122 | 113 | 140 | 142 | 78 | 147 | 163 | 95 |

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

**Table 3** Haplotype frequences and number of individuals (*n*) for seven *Pinus resinosa* populations

| | Population | | | | | | |
|---|---|---|---|---|---|---|---|
| Haplotype | A | B | D | E | F | G | H |
| I | 1 | – | – | 8 | 1 | 2 | – |
| II | 20 | 16 | 12 | 9 | 18 | 5 | 17 |
| III | 1 | – | – | – | – | – | – |
| IV | 1 | – | – | – | – | – | – |
| V | 1 | 1 | 1 | – | 1 | – | – |
| VI | – | 1 | – | – | – | – | – |
| VII | – | 1 | 3 | – | – | – | – |
| VIII | – | 1 | 1 | – | 1 | 2 | – |
| IX | – | 2 | – | – | 1 | 2 | 2 |
| X | – | 2 | – | – | – | 4 | 1 |
| XI | – | – | 2 | – | – | – | – |
| XII | – | – | 1 | 2 | – | – | – |
| XIII | – | – | 1 | – | – | – | – |
| XIV | – | – | – | 1 | – | – | – |
| XV | – | – | – | 1 | 1 | 1 | – |
| XVI | – | – | – | – | 1 | – | 1 |
| XVII | – | – | – | – | – | 2 | – |
| XVIII | – | – | – | – | – | 1 | – |
| XIX | – | – | – | – | – | 1 | – |
| XX | – | – | – | – | – | 1 | – |
| XXI | – | – | – | – | – | 1 | – |
| XXII | – | – | – | – | – | 1 | – |
| XXIII | – | – | 1 | – | – | – | – |
| *n* | 24 | 24 | 22 | 21 | 24 | 23 | 21 |

genetic distance and geographical distance. Spatial autocorrelation analyses were not performed because such tests would have little significance applied to only seven locations. To more clearly reveal bottleneck effects on elongated branch lengths of populations D, E, and G (Savard *et al*. 1993; Takezaki & Nei 1996), the $(\delta\mu)^2$ distance measure, which varies linearly with time, was used to construct a neighbour-joining dendrogram (Fig. 3a).

Bottleneck effects also generate unimodal distributions of pairwise differences between individuals in a population (Slatkin & Hudson 1991; see also von Haeseler *et al*. 1996). Distributions of pairwise cpSSR repeat length differences among individuals, summed over all nine cpSSR loci within an individual, are shown in Fig. 3b. Only population A had what appeared to be a bimodal distribution that could be viewed as characteristic of a panmictic population with low diversity and constant effective population size. Distributions in populations B, F, and H were unimodal, but all had a mode of zero, perhaps the result of declining population sizes. The remaining populations D, E, and G all had unimodal distributions with varying mean differences. Only population G had what approximated a Poisson distribution, indicative of exponential growth from a narrow coalescence (Slatkin & Hudson 1991). Independent measures of changes in the sizes of these populations are not available. When pairwise differences were counted among all 159 individuals, the frequency distribution was unimodal, with a mode of 1, but extending out to seven differences.

## Discussion

### *Genetic diversity*

Population genetic differences in *Pinus resinosa* (red pine) have been demonstrated using cpSSR markers. The highest

**Table 4** Measures of cpSSR haplotype variation within populations

| Population | $n_e$ | $f_\alpha$ | $H_E$ | $D^2_{SH}$ |
|---|---|---|---|---|
| A | 1.430 | 0.833 | 0.314 | 0.053 |
| B | 2.153 | 0.667 | 0.559 | 0.160 |
| D | 2.998 | 0.545 | 0.698 | 0.131 |
| E | 2.924 | 0.429 | 0.691 | 0.157 |
| F | 1.754 | 0.750 | 0.449 | 0.056 |
| G | 8.397 | 0.217 | 0.920 | 0.443 |
| H | 1.497 | 0.810 | 0.348 | 0.081 |

$f_\alpha$, frequency of the most common haplotype (haplotype II). See the Material and methods for definitions of other measures.

**Table 5** cpSSR genetic distance values among seven *Pinus resinosa* populations

| Population | A | B | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| A | | 0.213 | 0.282 | 0.276 | 0.126 | 0.514 | 0.179 |
| B | 0.007 | | 0.235 | 0.465 | 0.151 | 0.354 | 0.113 |
| D | 0.011 | 0.011 | | 0.451 | 0.273 | 0.593 | 0.336 |
| E | 0.016 | 0.027 | 0.021 | | 0.263 | 0.467 | 0.411 |
| F | 0.002 | 0.007 | 0.009 | 0.019 | | 0.373 | 0.113 |
| G | 0.068 | 0.044 | 0.076 | 0.073 | 0.070 | | 0.398 |
| H | 0.005 | 0.002 | 0.012 | 0.026 | 0.003 | 0.052 | |

$(\delta\mu)^2$ distances among populations, based on cpSSR allele frequencies, appear below the diagonal. $D_a$ distances among populations, based on chloroplast haplotype frequencies, appear above the diagonal.

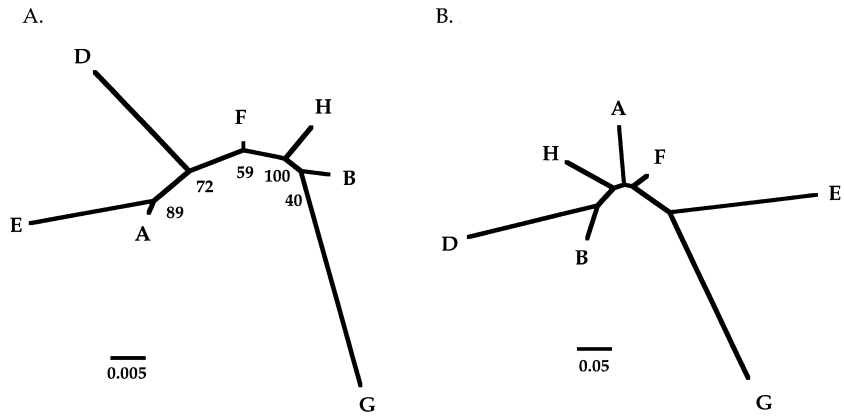


**Fig. 2** Unrooted neighbour-joining trees for *Pinus resinosa* populations, based on $D_a$ distances. The tree on the left (A) was derived from cpSSR allele frequencies, while the tree on the right (B) was derived from chloroplast haplotype frequencies. The scale bars represent distance units. Numbers at the nodes are bootstrap values (in percentages) based on 1000 replications. As only a single 'locus' was considered for building the haplotype tree, no bootstrap value could be estimated.

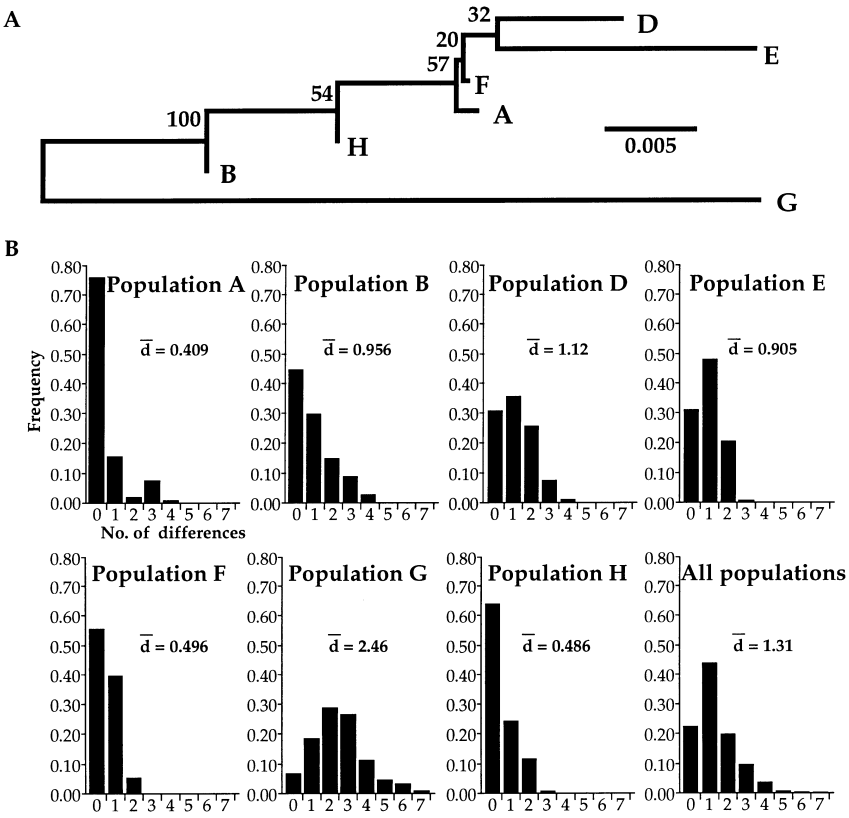


**Fig. 3** Evidence of population bottleneck events. A. Neighbour-joining dendrogram of *Pinus resinosa* populations, based on $(\delta\mu)^2$ distances. Populations B and H had negative branch lengths. The scale bar represents distance units. Numbers are bootstrap values (in percentages) based on 1000 replications. B. Histograms showing cpSSR repeat unit differences, across all nine cpSSR loci, among pairs of individuals within populations. The values indicated above the bars are the means of the differences.

level of haplotypic diversity was concentrated in population G from Tracy, New Brunswick, which should therefore deserve particular attention in genetic conservation programs for red pine (see the Conclusions). The mean proportion of chloroplast haplotype diversity that we found to exist among populations was nearly twice the average value reported for wind-pollinated conifers using nuclear allozyme markers, where $G_{ST}$ = 6.8% (Hamrick & Godt 1990). However, the level of genetic differentiation found in red pine is lower than that generally reported for other pine species having disjunct population structures, as measured by both chloroplast and mitochondrial markers (Moran *et al*. 1988; Gibson & Hamrick 1991; Hong *et al*. 1993; Strauss *et al*. 1993; Wang & Szmidt 1994; Powell *et al*. 1995; Szmidt *et al*. 1996). Genetic differentiation values are generally higher using mitochondrial markers because they are maternally inherited in conifers, and seeds typically disperse shorter distances than pollen. It must be stressed that direct comparisons of $G_{ST}$ values between different species, or between different marker

systems, might not be informative in some cases because $G_{ST}$ is a function of effective population size, mutation rate and other factors (Birky *et al*. 1989; Hong *et al*. 1993; Petit *et al*. 1993). Considering the limited variation in the red pine nuclear genome, the level of haplotype differentiation found in this study is consistent with disjunct populations having limited gene flow.

The level of within-population stepwise haplotypic diversity of red pine (mean $D^2_{SH} = 0.154$) was much lower than that observed for cpSSRs in *Pinus halepensis* (mean $D^2_{SH} = 3.58$) (Morgante *et al*. 1997). *P. halepensis* (Aleppo pine) is another species characterized by very low allozyme diversity (Schiller *et al*. 1986), and is also found in disjunct populations throughout its range in the Mediterranean region. Most of the cpSSR diversity in *P. halepensis* is attributed to just two populations in Greece, which is thought to be the centre of diffusion for this, and other Mediterranean species (Morgante *et al*. 1997). This suggests the possibility that high levels of cpSSR diversity may also reside in relict red pine populations found in West Virginia, USA, an area from which the species is thought to have expanded following the last Ice Age.

The presence of a single, common, and often frequent, haplotype in all populations (Table 3) suggests a major ancient population bottleneck event for red pine. Evidence for rapid expansion from a population bottleneck can be seen in the unimodal distribution of pairwise differences among all sampled individuals (Fig. 3b) (Slatkin & Hudson 1991; von Haeseler *et al*. 1996). However, differences in the levels and distributions of diversity found among the populations (Table 4, Fig. 3) indicate that other process affecting population genetic structure have been at work since this initial expansion. Indeed, if a simple, continuous expansion of red pine into its present day range from a single Ice Age refugium had occurred, then a 'star-shaped' dendrogram would be expected (Slatkin & Hudson 1991; von Haeseler *et al*. 1996). Instead, the dendrograms had a number of internodes (Fig. 2), indicating diverse origins of present-day populations. Together with what is know about the life history of red pine, these cpSSR marker data provide population genetic evidence for a metapopulation structure of red pine (Pimm *et al*. 1989; Mosseler 1992; Hedrick & Gilpin 1997). Present-day distribution of cpSSR diversity is an expected result of localized colonizations with variable numbers of migrants, coupled with restricted gene flow between disjunct populations contributing to localized declines and extinctions.

There was no obvious correlation between genetic and geographical distances to suggest a general pattern of dispersal of the species. But then, spatial autocorrelations would not be evident with a metapopulation structure given our sampling strategy. More dense sampling throughout the range of red pine will be needed to discern geographical lineages and routes of migration, provided that rates of extinction of local populations have not been great enough to obscure this record.

### *Mutational mechanisms*

A relatively constant rate of mutation among cpSSR loci was evident, as seven polymorphic cpSSR loci had three alleles, and two had two alleles (Table 2). The two loci having two alleles, PT36480 and PT41093, also had the shortest repeat lengths (Vendramin *et al*. 1996). This is consistent with reports that short nuclear dinucleotide repeats mutate more slowly than longer repeats (Jin *et al*. 1996; Weber 1990; Weber & Wong 1993). Although the rate of cpSSR mutation is not known, a lower rate would be expected because mutation of nuclear SSRs can be enhanced by heterozygosity (Amos *et al*. 1996), and heterozygosity is absent in chloroplast genomes. Whether the mutational mechanisms for SSRs are the same between plastid and nuclear genomes remains to be determined; therefore assumptions of cpSSR variability drawn from information known about nuclear SSRs should be used with caution.

Our data support a stepwise model for cpSSR mutations within a species, as all but one allelic length difference that we observed involved a single nucleotide change (Table 2). The 2-nucleotide length difference between allele PT15169:124 bp and PT15169:122 bp does not necessarily contradict a strict stepwise model. The longer allele was present in only two populations, B and G, where its frequency was 0.083 and 0.348, respectively. As these two populations are separated by about 1700 km, the missing allele of intermediate size (123 bp) may exist in populations not included in our survey.

### *Phenetic relationships*

Takezaki & Nei (1996) used simulations to show that for species, such as red pine, that have survived bottlenecks, $D_a$ distances and neighbour-joining analysis give the most accurate phenetic tree topologies when compared to other distance measures, including those based on stepwise mutation models. As $D_a$ distances are independent of mutation model, we generated neighbour-joining dendrograms from haplotype data, which fit an infinite allele mutation model, and cpSSR allele data, which fit a stepwise mutation model. However, in this case, there was very good congruence between the $D_a$ distance and $(\delta\mu)^2$ distance neighbour-joining trees based on cpSSR allele data (Figs 2a and 3a). In contrast, there were notable differences between the allelic tree and the haplotype tree based on $D_a$ distances (Fig. 2). Perhaps the most striking are the smaller differences among external branch lengths found in the haplotype tree (Fig. 2b). This indicates that

cpSSR allele frequency data provided greater discrimination than haplotype data of genetic distance differences among populations. When cpSSR repeat length differences were considered, as with $(\delta\mu)^2$ distances, even greater differences among populations were evident (Fig. 3a). Fewer differences among branch lengths were to be expected with haplotype data (Fig. 2b) because the haplotypes under-represented mutational differences at individual cpSSR loci. For example, haplotypes I and II differed by one repeat unit length at a single locus, while haplotypes IV and V differed by repeat unit length changes at each of four cpSSR loci (Table 2). In genetic distance measures based on haplotype frequencies, the single SSR difference between haplotypes I and II would be treated the same as multiple SSR differences between haplotypes IV and V. The different placement of various populations on the two trees, most notably populations B and E, may have resulted from this homoplastic effect of haplotype designations. As a final consideration, it may be that some of the differences between the dendrograms are artefacts of the rather small sample sizes and few loci used (see Takezaki & Nei 1996; Zhivotovsky & Feldman 1995).

## Conclusions

We have shown that red pine chloroplast genome diversity exists in contrast to a largely homozygous nuclear genome. Moreover, variation in the distributions of pairwise haplotype differences among populations has increased our knowledge of the evolutionary history of this species, and provided the first population genetic evidence for a metapopulation structure of red pine. These results do not imply that the chloroplast genome is evolving faster than the nuclear genome. Rather, they derive from the paternally inherited chloroplast genome of pines retaining more evolutionary information than the nuclear genome, and from cpSSR markers being more variable than either isozymes or RAPD markers. Whether red pine nuclear SSR loci mutate at the same rate as the cpSSR loci, and whether nuclear genome SSR marker data would support these findings, remain to be seen. Studies are in progress to obtain nuclear SSR genotypes for red pine using SSR primer pairs that were developed from related pine species.

More extensive cpSSR surveys should identify additional populations that have high levels of chloroplast genetic diversity and differentiation. This information could prove to be important for genetic resource conservation programs of red pine. With proper sampling, cpSSR marker data have the potential to trace paths of post-Pleistocene migrations of the species, and thus efficiently identify populations that are the most evolutionarily divergent. It is reasonable to assume that such populations would also be the most divergent for nuclear genomic traits. This assumption could be tested by using nuclear SSR markers, and provenance tests for adaptive traits, to evaluate populations that have been selected based on their maximum cpSSR distances. As the chloroplast microsatellite approach revealed population genetic differences in a species characterized by no detectable allozyme variation, it might also be considered for studying population structures and evolutionary histories of other conifer species that have low nuclear genome diversity, such as Torrey pine, (*Pinus torreyana* Parry ex Carr) (Ledig & Conkle 1983) or western red cedar (*Thuja plicata* Donn ex E. Don) (Copes 1981).

## Acknowledgements

We thank Eric Minch of Stanford University, and Naoko Takezaki of the National Institute of Genetics (Japan), for assistance with the programs they provided. The ArcView expertise of John Wright of the North Central Forest Experiment Station is much appreciated. We also thank Jean Bousquet of the University of Laval, Quebec, Bryan Epperson of Michigan State University, USA, Chris Lucarotti of the Canadian Forest Service, Om Rajora of the University of Alberta, Canada, Outi Savolainen of the University of Oulu, Finland, and two anonymous reviewers for their helpful comments. This project was partly supported by the National Research Council of Italy.

## References

Ager A, Guries R, Lee C-H (1983) Genetic gains from red pine seedling seed orchards. *Proceedings of the 28th Northeastern Forest Tree Improvement Conference*, pp. 175–194. Institute of Natural and Environmental Resources, University of New Hampshire, Durham, NH.

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.

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. *Nature Genetics*, **13**, 390–391.

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

Cato SA, Richardson TE (1996) Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. *Theoretical and Applied Genetics*, **93**, 587–592.

Copes DL (1981) Isoenzyme uniformity in western red cedar seedlings from Oregon and Washington. *Canadian Journal of Forest Research*, **11**, 451–453.

Dallas JF (1992) Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammalian Genome*, **3**, 452–456.

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

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 Sciences, USA*, **91**, 3166–3170.

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

Ellegren H (1995) Mutation rates at porcine microsatellite loci. *Mammalian Genome*, **6**, 376–377.

Feldman MW, Bergman A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range constraints: Analytic description and problems of estimation. *Genetics*, **145**, 207–216.

Felsenstein J (1995) PHYLIP (Phylogeny Inference Package) version 3.572. Department of Genetics, University of Washington, Seattle.

Fowler DP (1964) Effects of inbreeding in red pine, *Pinus resinosa* Ait. I. Natural variation. *Silvae Genetica*, **13**, 170–177.

Fowler DP (1965a) Effects of inbreeding in red pine, *Pinus resinosa* Ait. II. Pollination studies. *Silvae Genetica*, **14**, 12–23.

Fowler DP (1965b) Effects of inbreeding in red pine, *Pinus resinosa* Ait. III. Factors affecting natural selfing. *Silvae Genetica*, **14**, 37–46.

Fowler DP (1965c) Effects of inbreeding in red pine, *Pinus resinosa* Ait. IV. Comparison with other northeastern *Pinus* species. *Silvae Genetica*, **14**, 76–81.

Fowler DP, Lester DT (1970) The genetics of red pine. U.S.D.A. Forest Service Research Paper W-8.

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

Gibson JP, Hamrick JL (1991) Genetic diversity and structure in *Pinus pungens* (Table Mountain pine) populations. *Canadian Journal of Forest Research*, **21**, 635–642.

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

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences, USA*, **92**, 6723–6727.

von Haeseler A, Sajantila A, Paabo S (1996) The genetical archaeology of the human genome. *Nature Genetics*, **14**, 135–140.

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

Hedrick PW, Gilpin ME (1997) Genetic effective size of a metapopulation. In: *Metapopulation Biology; Ecology, Genetics and Evolution*. (eds Hanski I, Gilpin ME), pp. 165–181. Academic Press, San Diego.

Hong YP, Hipkins VD, Strauss SH (1993) Chloroplast DNA diversity among trees, populations and species in the California closed-cone pines (*Pinus radiata*, *Pinus muricata* and *Pinus attenuata*). *Genetics*, **135**, 1187–1196.

Jin L, Macaubas C, Hallmayer J, Kimura A, Mignot E (1996) Mutation rate varies among alleles at a microsatellite locus: Phylogenetic evidence. *Proceedings of the National Academy of Sciences, USA*, **93**, 15285–15288.

Ledig FT, Conkle MT (1983) Gene diversity and genetic structure in a narrow endemic, Torrey pine (*Pinus torreyana* Parry ex. Carr.). *Evolution*, **37**, 79–85.

Mitton JB (1993) Molecular approaches to population biology. *Annual Review of Ecology and Systematics*, **25**, 45–69.

Moran GF, Bell JC, Eldridge KG (1988) The genetic structure and the conservation of the five natural populations of *Pinus radiata*. *Canadian Journal of Forest Research*, **18**, 506–514.

Morgante M, Felice N, Vendramin GG (1997) Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. In: *Molecular Tools for Screening Biodiversity: Plants and Animals*. (eds Karp A, Isaac PG, Ingram DS), Chapman and Hall, London.

Mosseler A (1992) Life history and genetic diversity in red pine: implications for gene conservation in forestry. *The Forestry Chronicle*, **68**, 701–708.

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**, 212–216.

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

Nei M, Maruyana T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

Petit RJ, Kremer A, Wagner DG (1993) Finite island model for organelle and nuclear genes in plants. *Heredity*, **71**, 630–641.

Pimm SL, Gittleman JL, McCracken GF, Gilpin M (1989) Plausible alternatives to bottlenecks to explain reduced genetic diversity. *Trends in Ecology and Evolution*, **4**, 176–178.

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, USA*, **92**, 7759–7763.

Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology*, **5**, 69–76.

Saitou N, Nei M (1987) The neighbor-joining method: a new model for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Savard L, Michaud M, Bousquet J (1993) Genetic diversity and phylogenetic relationships between birches and alders using *rbcL*, 18S and ITS rRNA gene sequences. *Molecular Phylogenetics and Evolution*, **2**, 112–118.

Schiller G, Conkle MT, Grunwald C (1986) Local differentiation among Mediterranean populations of Aleppo pine in their isoenzymes. *Silvae Genetica*, **35**, 11–19.

Simon J-P, Bergeron Y, Gagnon D (1986) Isozyme uniformity in populations of 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.

Strauss SH, Hong YP, Hipkins VD (1993) High levels of population

differentiation for mitochondrial DNA haplotypes in *Pinus radiata*, *muricata*, and *attenuata*. *Theoretical and Applied Genetics*, **86**, 605–611.

Szmidt AE, Wang XR, Changtragoon S (1996) Contrasting patterns of genetic diversity in two tropical pines: *Pinus kesiya* (Royle ex Gordon) and *P. merkusii* (Jungh et De Vriese). *Theoretical and Applied Genetics*, **92**, 436–441.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

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

Vendramin GG, Ziegenhagen B (1997) Characterization and inheritance of polymorphic plastid microsatellites in *Abies*. *Genome*, **40**, 857–864.

Wagner DB, Nance WL, Nelson CD, Li T, Patel RN, Govindaraju DR (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 Genome Research*, **22**, 683–689.

Wang XR, Szmidt AE (1994) Hybridization and chloroplast DNA variation in a *Pinus* species complex from Asia. *Evolution*, **48**, 1020–1031.

Watano Y, Imazu M, Shimizu T (1996) Spatial distribution of cpDNA and mtDNA haplotypes in a hybrid zone between *Pinus pumila* and *P. parviflora* var. *pentaphylla* (Pinaceae). *Journal of Plant Research*, **109**, 403–408.

Weber JL (1990) Informativeness of human (dC–dA)n: (dG–dT)n polymorphisms. *Genomics*, **7**, 524–530.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**, 1123–1128.

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.

Zhivotovsky LA, Feldman MW (1995) Microsatellite variability and genetic distances. *Proceedings of the National Academy of Sciences, USA*, **92**, 11549–11552.

---

Craig Echt, a Research Geneticist with the U.S.D.A. Forest Service, develops and applies molecular markers for the study of genetic systems of forest trees. Linda DeVerno, a molecular biologist with the Canadian Forest Service, uses molecular markers for the analysis of biodiversity of forest trees, insects and pathogens. Maria Anzidei is a PhD. student at the University of Florence and works with Giovanni Vendramin, a research plant population geneticist with the National Research Council. G. Vendramin develops and uses molecular markers to study population and evolutionary processes of forest trees at the Forest Tree Breeding Institute in Florence, Italy.

---