

Low Nucleotide Diversity at the *pall* Locus in the Widely Distributed *Pinus sylvestris*

Volodymyr Dvornyk,¹ Anu Sirviö, Merja Mikkonen, and Outi Savolainen

Department of Biology, University of Oulu, Finland

Nucleotide polymorphism in Scots pine (*Pinus sylvestris*) was studied in the gene encoding phenylalanine ammonia-lyase (Pal, EC 4.3.1.5). Scots pine, like many other pine species, has a large current population size. The observed levels of inbreeding depression suggest that Scots pine may have a high mutation rate to deleterious alleles. Many Scots pine markers such as isozymes, RFLPs, and microsatellites are highly variable. These observations suggest that the levels of nucleotide variation should be higher than those in other plant species. A 2,045-bp fragment of the *pall* locus was sequenced from five megagametophytes each from a different individual from each of four populations, from northern and southern Finland, central Russia, and northern Spain. There were 12 segregating sites in the locus. The synonymous site overall nucleotide diversity was only 0.0049. In order to compare *pall* with other pine genes, sequence was obtained from two alleles of 11 other loci (total length 4,606 bp). For these, the synonymous nucleotide diversity was 0.0056. These estimates are lower than those from other plants. This is most likely because of a low mutation rate, as estimated from between-pine species synonymous site divergence. In other respects, Scots pine has the characteristics of a species with a large effective population. There was no linkage disequilibrium even between closely linked sites. This resulted in high haplotype diversity (14 different haplotypes among 20 sequences). This could also give rise to high per locus diversity at the protein level. Divergence between populations in the main range was low, whereas an isolated Spanish population had slightly lower diversity and higher divergence than the remaining populations.

Introduction

Mating system and life history have major influences on the patterns of genetic variation (Hamrick and Godt 1996). Such patterns are well known from studies of isozymes. Most data on nucleotide variation in plants concern annuals and short-lived perennials, partly cereals or other domesticated crops, or predominantly selfing species, such as *Arabidopsis thaliana* (Gaut and Clegg 1993; Innan et al. 1996; Huttley et al. 1997; Kawabe, Miyashita, and Terauchi 1997; Kawabe et al. 1997; Stephan and Langley 1998; Kawabe and Miyashita 1999; Small, Ryburn, and Wendel 1999). Only a few studies compare patterns of nucleotide diversity in plant species with different mating systems (Liu, Zhang, and Charlesworth 1998; Liu, Charlesworth, and Kreitman 1999; Savolainen et al. 2000). So far, outcrossing long-lived plants have not been studied for nucleotide variation.

In this study, we examine nucleotide variation in Scots pine (*Pinus sylvestris* L.). It is a long-lived predominantly outcrossing perennial with continuous distribution that ranges from Scotland to eastern Siberia and from northern Finland to Turkey (fig. 1). It has high variability at enzyme loci, RFLPs, and microsatellites (Gullberg et al. 1985; Muona and Harju 1989; Müller-Stark, Baradat, and Bergmann 1992; Karvonen and Savolainen 1993; Goncharenko, Silin, and Padutov 1994; Karhu et al. 1996). In 10 allozyme studies, the average heterozygosity values per population ranged from 0.25

to 0.39 for Scots pine (Müller-Stark, Baradat, and Bergmann 1992), among the highest for the European tree species compared. These values are high compared with many other groups, such as *Drosophila* allozyme heterozygosity with an average of 0.14. The differentiation between populations at marker loci is low within Scandinavia ($F_{ST} = 0.02$) (Karhu et al. 1996) and even between Scandinavia and the eastern part of the range (Wang, Szmidi, and Lindgren 1991). There is much recombination in the genome of pines: the genetic genome size of *Pinus taeda* is 1,700–1,800 cM, likely to be similar to that in many pines (Remington et al. 1999). No linkage disequilibrium was found between enzyme loci either in pollen grains or female gametes (Muona and Szmidi 1985). Scots pine has very high early and late inbreeding depression (Koski 1971; Kärkkäinen, Koski, and Savolainen 1996). In a partially selfing species, most deleterious alleles should be rapidly eliminated. The most reasonable way to account for high inbreeding depression in a partially selfing species is a high genome-wide mutation rate to deleterious alleles (Lande, Schemske, and Schultz 1994), and Scots pine seems to conform to this model (Koelewijn, Koski, and Savolainen 1999).

This background generates predictions with respect to nucleotide diversity. First, as the effective population size (N_e) and mutation rate (μ) (to deleterious genes) are known to be high, we could predict high levels of neutral nucleotide diversity, $\theta = 4N_e\mu$ (Kimura 1983, pp. 194–252). The earlier findings of enzyme gene variation also predict high diversity. Second, as the populations at molecular markers have only low differentiation, the differentiation at the nucleotide level could be low as well, if both are predominantly governed by gene flow and drift.

However, earlier data do not allow us to make a well-justified prediction on the range of linkage dis-

¹Present address: Osteoporosis Research Center, Omaha, Nebraska, USA.

Key words: nucleotide polymorphism, SNP, *Pinus sylvestris*, codon bias, linkage disequilibrium.

Address for correspondence and reprints: Outi Savolainen, Department of Biology, University of Oulu, P.O. Box 3000, FIN-90014 Oulu, Finland. E-mail: outi.savolainen@oulu.fi.

Mol. Biol. Evol. 19(2):179–188. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

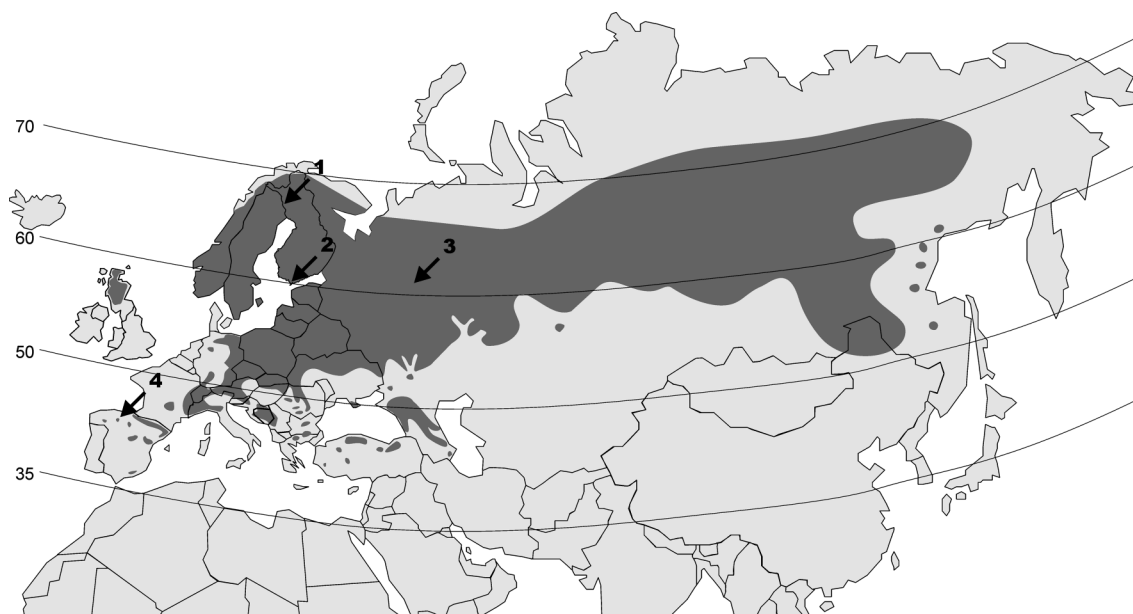


FIG. 1.—The distribution of *P. sylvestris* natural populations has been marked in gray. The numbers indicate the locations sampled in this study: 1, Kolari (Finland); 2, Bromarv (Finland); 3, Kirov (Russia); 4, Puebla de Lillo (Spain).

equilibrium in the genome. The level of disequilibrium between neutral genes is governed by the product $4N_c c$, where c is the recombination rate between loci (Hill and Robertson 1968). Close linkage or inbreeding will restrict effective recombination. The larger this product, the less disequilibrium we expect between neutral loci. The current population size is very large and there is much recombination between genes, but we might still observe effects of historical bottlenecks between closely linked nucleotide sites. The current population structure of Scots pine in Europe appears to have been formed mainly by the postglacial migration of the species from refugia located in southwestern, southern, and southeastern parts of its present range. This expansion took place about 7,000–9,000 years ago (Hyvärinen 1987). Scots pine mitochondrial DNA variants provide further evidence for multiple origins after the last glaciation from the south: through western Europe via France, Germany, and Denmark and from the northeast via Finland (Sinclair, Morman, and Ennos 1999; Soranzo et al. 2000). Are we able to detect traces of the glacial history at the nucleotide level, or has ample gene flow because of pollen migration (Koski 1970) and abundant recombination already eliminated these traces? Are the populations at mutation drift equilibrium, or do they still show traces of the possible glacial bottlenecks, in disequilibria between closely linked nucleotide sites or in frequency spectra that are predicted in expanding populations (Harpending 1994).

We studied nucleotide polymorphism in the gene encoding phenylalanine ammonia-lyase (Pal, EC 4.3.1.5), a key enzyme in the secondary metabolism of higher plants catalyzing the nonoxidative conversion of L-phenylalanine into transcinamate, the initial substrate of the phenylpropanoid pathway in wood formation. In Scots pine, Pal is believed to be related to ozone tolerance (Rosemann, Heller, and Sandermann Jr. 1991), de-

fense against pathogens (Lange et al. 1994), and metabolism of exogenous compounds (Laukkanen and Sarjala 1997). cDNA sequences of *pal* have been reported for a number of angiosperms (e.g., Yamada et al. 1992). In angiosperms this gene usually consists of two exons and one intron. In gymnosperms, *pal* was first reported from loblolly pine, *P. taeda*, as a single-copy gene with no introns (Whetten and Sederoff 1992). Later, Butland, Chow, and Ellis (1998) sequenced a 366-bp-long conserved region of *pal* from *Pinus banksiana* and discovered that *pal* is actually a multigene family of at least 8–10 loci.

Our results first show that Scots pine has rather low nucleotide diversity at synonymous sites. Following this finding, we examined whether other aspects of the nucleotide data are consistent with the predictions for a species with a large effective population size.

Materials and Methods

Plant Materials

Seeds of Scots pine, *P. sylvestris*, were obtained from four natural populations: two Finnish (Bromarv and Kolari), one Russian (Kirov), and one Spanish (Puebla de Lillo in the province of Leon) (fig. 1). Seeds were stored at 4°C. For DNA extraction, the seeds were kept on paper towels moistened with distilled water for 4–5 h at room temperature to make tissue preparation easier. Only haploid tissue (megagametophyte) of the seed was used for DNA isolation. Five megagametophytes from different trees from each population were analyzed. We also obtained two sequences from 11 other loci. The two alleles were from two widely separated trees in southern Finland.

DNA Isolation, Amplification, and Sequencing

Total genomic DNA from the megagametophytes was purified either by CTAB method (Doyle and Doyle

1990) or by FastDNA® Kit (BIO101). On the basis of the *P. taeda pal*-gene nucleotide sequence (one exon of 2,265 bp, GenBank U39792), we designed PCR-primers that would amplify the *pal* gene in two overlapping parts in *P. sylvestris*. To amplify the other *P. sylvestris* genes used in this study, we designed primers on the basis of *P. sylvestris* and *P. taeda* gene sequences from GenBank (table 1). Short fragments were directly sequenced, first using AmpliTaq Gold (PE Applied Biosystems) for the PCR. The PCR products were separated in 1% agarose gels (0.5 TBE), from which they were purified by centrifugation through aquarium filter fibers. Direct sequencing was done from both strands with BigDye-kit (PE Applied Biosystems). For cloning of longer parts of the genes, PCR was done with Dynazyme EXT (with proofreading property, Finnzyme). The amplification products were first cloned with the TOPO TA cloning kit (version J, Invitrogen) and then sequenced (BigDye, PE Applied Biosystems). The PCR-program for amplification of the *pal* gene reactions with AmpliTaq Gold-enzyme was: 94°C 12 min; 94°C 1.10 min, 53°C 1 min, 72°C 1.50 min, 30 cycles; 72°C 10 min, and for Dynazyme-enzyme: 94°C 4 min; 94°C 1.10 min, 53°C 1 min, 72°C 1 min, 30 cycles; 72°C 10 min. Three cloned products were sequenced in order to avoid errors made by the DNA polymerase. If nucleotide differences between the three clones were detected, then the majority nucleotide was taken to be the correct one in that position. In both sequencing methods, we used several sequencing primers designed on the basis of the *P. taeda* sequence and located every 300–400 bp. Newly determined *P. sylvestris* sequences are deposited in the GenBank database and the accession numbers are listed in table 1.

Data Analyses

We obtained estimates of the scaled mutation parameter, θ , based on nucleotide diversity, π (Nei and Li 1979) and on the proportion of segregating sites, S (Watterson 1975). Differences between populations were measured by F_{ST} , estimated with the analysis of molecular variance (AMOVA) (Excoffier, Smouse, and Quattro 1992) as implemented in the Arlequin 2.000 software (Schneider, Roessli, and Excoffier 2000). We also estimated the number of recombinations (Hudson 1987) and the linkage disequilibrium. The degree of codon bias was measured by the effective number of codons (ENC; Wright 1990). The ENC values can range from 20 (strong codon bias) to 61 (no codon bias). Unless otherwise stated, we used DnaSP (Rozas and Rozas 1999) for all analyses.

Results

Structure and Divergence of the *pal* Locus of *P. sylvestris*

The *pal* locus in Scots pine is similar to the *pal* locus of jack pine, *P. banksiana* (Butland, Chow, and Ellis 1998), and to the *pal* reported in loblolly pine, *P. taeda* (Whetten and Sederoff 1992). In all these species, the *pal* locus is represented by a single exon without

introns. Within the region studied, there were no insertions or deletions. There were, on an average, only 12.2 nonsynonymous substitutions between the species (K_a nonsynonymous = 0.0079).

Among the sequenced 2,045 sites in 20 alleles, there were 12 segregating sites (table 2). Nine of the polymorphisms were synonymous and three were nonsynonymous. A transition of C to T at site 243 resulted in a change of alanine to valine in two individuals of the Kolari population. In the Kirov population, a transversion of T to G at site 762 resulted in a change of valine into glycine in one individual, and a transversion of A to T at site 915 alters histidine to leucine. These amino acid changes are considered to be conservative for protein function (Graur and Li 2000, p. 20).

The species-wide estimate of synonymous diversity was 0.0049. For comparison we obtained nucleotide sequences from 11 other pine genes (table 1). The average synonymous nucleotide diversity at these loci was 0.0056, with a range of 0.0000–0.0232. The comparable nucleotide diversity at the *pal* locus in the Bromarv population in southern Finland was 0.0056 (table 3). Thus, based on these very preliminary comparisons, *pal* seems representative of other genes as well.

Within-population Nucleotide Diversity

Scots pine has a low level of within-population nucleotide diversity in the *pal* locus (table 3). For synonymous sites, the mean within-population nucleotide diversity π_s was 0.0045, with the highest value found in the populations from Bromarv and Kirov, 0.0056, and the lowest one in the Spanish population from Puebla de Lillo, 0.0024. At nonsynonymous sites, polymorphism was only detected in two of the populations, Kolari and Kirov, at a very low level ($\pi_a = 0.0004$ and 0.0005, respectively) with an overall average of populations of 0.0002. The lowest total nucleotide polymorphism value was observed in the Spanish population ($\pi = 0.0006$), and the highest was detected in the Russian population ($\pi = 0.0018$). The three populations from the main range of Scots pine distribution (Bromarv, Kolari, and Kirov) had significantly higher levels of total nucleotide variation than the geographically isolated population of Puebla de Lillo from Spain ($P < 0.05$, t -test).

All polymorphic sites had only two alternative nucleotides, and four of the sites were singletons. As the populations showed so little divergence, we combined them into one set for Tajima's test. A significant negative Tajima's test could suggest either purifying selection or recent population expansion. However, the value of -0.56 was not statistically significantly different from zero.

Differentiation Between Populations

The genetic differentiation between the populations as estimated with AMOVA (table 4) was very low. If the population from Spain was excluded, the value of F_{ST} was only 0.017, but if it was included in the calculations, the F_{ST} increased to 0.110. Statistically sig-

Table 1
The Genes and Primers Used and the GenBank Accession Numbers of the Sequences Determined in this Study

Gene Name and Function	GenBank Sequences the Primers Were Based on	Primers for PCR Amplification and Sequencing ^a	Accession Numbers of New Sequences
<i>palI</i> , phenylalanine ammonia-lyase	<i>P. taeda</i> U39792	5F 5'-TAGCAGCAGCAGAAAATAAC-3' 127R 5'-TCTGAGAAATTGAGCGAACAC-3' 910F 5'-ACAGATCCATTAACTCACA-3' 2088R 5'-ACTGATAGCGTCGTAAACCA-3' 176F 5'-GTAGGCAGGCTGTGAGT-3' 637R 5'-TACCCACACATCAGTCCAGTC-3' 283F 5'-GGAGCTAGTGAACGGGAGGTG-3' 672R 5'-GAAACGGGCCCTTCTGTGA-3' 7264F 5'-ACCTTCCACCGATAAAACATT-3' 7812R 5'-AGGATAAAGGCACAAACATAC-3' 64F 5'-TGCAATTACGGTTACTGAGAG-3' 350R 5'-ATTTCTTGACACCCGAAATAGG-3' 14F 5'-TTGTAGTTGGAGAGAAAGACT-3' 362R 5'-CAATATCCAGCATTTTATGT-3' 343F 5'-CATATCCGTGATTTGGTGCTT-3' 511R 5'-CAGTGGCATCCAGTTTTTCTT-3' 17F 5'-GCAAAATGCAAAAACCATAC-3' 427R 5'-AAAAATAAGCAGGCAAAAATGAC-3' 86F 5'-TACTATGCTGATGGCGAGGAT-3' 474R 5'-GTTTCAACCGGATTTACAG-3' 224F 5'-TTCTGCCGATTTCCATAGAG-3' 460R 5'-TTAATAAGCGATTCCTTGATG-3' 93F 5'-GCTTCCACCGCACGAACAGT-3' 532R 5'-TCTTCCGCATCATTTCTCTGG-3' 173F 5'-TCAGATTTCCGAGTGTCTATG-3' 499R 5'-AACTATTTCCCAAGGATTTGTA-3'	AF353967–AF353986
<i>LHCA2</i> , LHC I type II chlorophyll-binding protein	<i>P. syhvestris</i> X58516		AF359117
<i>LHCA4</i> , LHC I type IV chlorophyll-binding protein	<i>P. syhvestris</i> Z17226		AF359118
<i>GapCp</i> , NAD-dependent glyceraldehyde-3-phosphate dehydrogenase	<i>P. syhvestris</i> AJ001706		AF359129
<i>Fructok</i> , putative fructokinase	<i>P. taeda</i> AA556405		AF359130
<i>Epoxhyd</i> , putative epoxide hydrolase	<i>P. taeda</i> AA556165		AF359119
<i>Maldehy</i> , Putative malate dehydrogenase	<i>P. taeda</i> AA557074		AF359131
<i>Phosfru</i> , putative phosphofructokinase	<i>P. taeda</i> AA556465		AF359132
<i>ARD1</i> , putative N-terminal acetyltransferase	<i>P. taeda</i> AA557106		AF359115
<i>DTDPglu</i> , putative DTDP-glucose	<i>P. taeda</i> AA557089		AF359116
<i>Dehydri</i> , putative dehydrin	<i>P. taeda</i> AA556801		AF359113
<i>Peroxi</i> , putative peroxidase	<i>P. taeda</i> AF028073		AF359114

^a The number in the primer's name refers to its 5' position in the GenBank sequence file the primers were based on (the 5' end of the positive strand), and the letter F or R identifies whether it is a forward or a reverse primer.

Table 2
Summary of Nucleotide Polymorphism in the Phenylalanine Ammonia-lyase (*pall*) Locus in *P. sylvestris*

CONSENSUS	NUCLEOTIDE POSITIONS											
	112	139	243 ^a	427	628	762 ^a	915 ^a	1156	1456	1687	2029	2038
	C	T	C	C	T	T	A	C	G	C	T	T
<i>Populations^b</i>												
Bro1.....	—	—	—	—	—	—	—	—	—	—	—	G
Bro2.....	—	—	—	—	—	—	—	—	A	—	—	—
Bro3.....	—	—	—	—	C	—	—	—	—	T	—	G
Bro4.....	—	—	—	T	—	—	—	—	—	—	—	—
Bro5.....	—	—	—	—	—	—	—	—	—	T	G	—
Kol1.....	—	—	—	—	C	—	—	G	—	—	—	G
Kol2.....	—	—	—	—	—	—	—	—	—	T	—	—
Kol3.....	—	—	T	—	—	—	—	—	—	T	—	—
Kol4.....	—	—	T	—	—	—	—	G	—	—	—	—
Kol5.....	—	—	—	—	C	—	—	—	—	—	—	—
Kir19.....	—	C	—	—	—	G	—	—	—	—	—	—
Kir21.....	—	—	—	T	—	—	—	—	—	—	—	—
Kir23.....	T	—	—	—	—	—	—	—	—	—	—	—
Kir27.....	—	C	—	—	—	—	—	G	—	—	—	—
Kir30.....	T	—	—	—	—	—	T	—	—	T	—	G
Lillo1.....	—	—	—	—	C	—	—	—	—	T	—	G
Lillo3.....	—	—	—	—	C	—	—	—	—	T	—	G
Lillo17.....	—	—	—	—	C	—	—	—	—	—	—	—
Lillo19.....	—	—	—	—	C	—	—	—	—	—	—	—
Lillo22.....	—	—	—	—	C	—	—	—	—	—	—	—

^a Nonsynonymous sites.

^b Bro, Bromarv, Finland; Kol, Kolari, Finland; Kir, Kirov, Russia; Lillo, Puebla de Lillo in the province of Leon, Spain.

nificant pairwise differentiation was found only between the most distant populations from Puebla de Lillo and Kirov ($F_{ST} = 0.310$).

Linkage Disequilibrium and Recombination

As the level of differentiation was so low, we analyzed the total set of 20 sequences jointly for linkage disequilibrium and for recombination parameter *R* (Hudson 1987). There was significant pairwise correlation by chi-square test between two pairs of sites (112 and 915, 139 and 762), but this is what is to be expected by chance, and after Bonferroni correction there was no evidence for linkage disequilibrium. Because of the low sample sizes the power of test is low, but table 2 shows no haplotype structure. The high recombination observed in the locus results in high haplotype diversity. A total of 14 haplotypes were detected among the 20 sequences analyzed. The overall haplotype diversity, *H* (Nei 1987, pp. 259–260), was 0.95 ± 0.03 . The smallest

number of the haplotypes, two, occurred in the Spanish population from Lillo. The other populations had five haplotypes each.

Discussion

Low Level of Within-population DNA Variability in Scots Pine

Our data do not support the predictions of high nucleotide polymorphism in Scots pine. Species-wide silent nucleotide diversity, π_s , was 0.0049. Table 5 shows that this is lower than the diversity in the short-lived plants, e.g., in *A. thaliana*, *Leavenworthia stylosa*, or *Zea mays*. Some plants do have lower estimates of diversity, such as *Arabidopsis lyrata* at *Adh*, or *Dioscorea tokoro* (Terauchi, Terachi, and Miyashita 1997). It is higher than the value in humans (Cargill et al. 1999) but lower than that in *Drosophila melanogaster* and *Drosophila simulans* (Moriyama and Powell 1996). The organisms with the longest generation times, pines and

Table 3
Nucleotide Polymorphism in *pall* Locus of Scots Pine

POPULATIONS	SUBSTITUTIONS ^a					
	Synonymous		Nonsynonymous		Total	
	π_s	θ_n	π_a	θ_n	π (S.D)	θ_n (SD)
Bromarv.....	0.0056	0.0058	0	0	0.0014 (0.0002)	0.0014 (0.0006)
Kolari.....	0.0044	0.0039	0.0004	0.0003	0.0014 (0.0003)	0.0012 (0.0005)
Kirov.....	0.0056	0.0058	0.0005	0.0006	0.0018 (0.0004)	0.0019 (0.0007)
Lillo.....	0.0024	0.0019	0	0	0.0006 (0.0002)	0.0005 (0.0003)
All.....	0.0049	0.0058	0.0003	0.0006	0.0014 (0.0001)	0.0017 (0.0005)

^a π_s from synonymous sites; π_a from nonsynonymous sites; θ_n theta from the number of mutations.

Table 4
Intra- and Interpopulation Variability (AMOVA)

Source of Variation	df	Sum of Squares	Variance Components	Percentage of Variation
Among populations	3	6.31	0.160 (V_a)	11
Within populations	16	20.83	1.302 (V_b)	89
Total	19	27.13	1.462	

NOTE.— F_{ST} (fixation index) = 0.11. V_a and F_{ST} : $P < 0.04$.

humans, have the lowest nucleotide diversity. Our estimate of synonymous nucleotide diversity in *P. sylvestris* at 11 other loci, with just two sequences for each locus was $\pi_s = 0.0056$ (table 5). Thus, the *pal* gene is not exceptional among genomic areas of *P. sylvestris*.

However, the level of replacement nucleotide polymorphism in the *pal* locus ($\pi_a = 0.0003$) was about seven times lower than in the other genes of Scots pine ($\pi_a = 0.0022$) and equal to the π_a in humans (table 5). This suggests a very high functional constraint on the protein.

Our results on rather low diversity could be because of several factors that we can examine. First, the effective population size (at least for this locus) could be lower than that assumed. Second, the mutation rate could be lower than that suggested by the inbreeding depression data. Third, the synonymous sites may in fact not be fully neutral, but under directional selection for major codons, e.g., for reasons of translational efficiency (see Akashi 1997). Further, the low variability could be due to selective forces at other loci that reduce levels of variation, such as hitchhiking (Maynard Smith and

Haigh 1974; Kaplan, Hudson, and Langley 1989), background selection (Charlesworth, Morgan, and Charlesworth 1993), or to pseudohitchhiking due to alternating direction of selection in large populations (Gillespie 2000). The high generation time and short time after glaciation bottlenecks may also mean that nucleotide polymorphism is not at equilibrium, even if our small data set did not show evidence of this (see below).

Large Effective Population Size and Mutation Drift Equilibrium

Many aspects of the data are consistent with a large effective population size. The differentiation between populations separated by thousands of kilometers was very low (e.g., Kolari in northern Finland and Kirov in Russia). Even the currently isolated Lillo was not much diverged. Further, there was no evidence of linkage disequilibrium in the data set, even between the very closely linked polymorphic sites. Linkage disequilibrium at neutral sites is governed by $1/(4N_e c)$, and as c between

Table 5
Allozyme Heterozygosity, Nucleotide Polymorphism, and Codon Bias (measured as effective number of codons, ENC)

Species	Sequences	Size (bp)	π_s	π_a	H_e	ENC
<i>P. sylvestris</i>	<i>pal</i>	2043	0.0049	0.0003	0.359 ^a	52.7
	11 genes	4606	0.0056	0.0022		
<i>P. radiata</i>					0.098 ^b	53.8
<i>D. tokoro</i>	<i>Pgi</i>	1698	0.0015 ^c	0.0010 ^c		47.8
<i>A. thaliana</i>	<i>Adh</i>	1137	0.0192 ^d	0.0022 ^d	0.148 ^e	52.0
<i>A. lyrata</i> subsp. <i>petraea</i>	<i>Adh</i>	2400	0.0038 ^f		0.288 ^g	52.2
<i>L. stylosa</i>	<i>PgiC</i>	2300	0.0530 ^h	0.0062 ^h	0.524 ⁱ	38.5
<i>Dr. melanogaster</i>	25 genes	23531	0.0135 ^j		0.132 ^k	49.4
<i>Dr. simulans</i>	12 genes	15066	0.0304 ^j		0.077 ^k	43.1
<i>Homo sapiens</i>	106 genes	135823	0.0010 ^l	0.0003 ^l	0.067 ^k	51.7
<i>Z. mays</i>	<i>Adh</i>	1400	0.0055 ^m		0.229 ⁿ	51.56

NOTE.—Calculations of ENC-values for the species are based on the Codon Usage Database (Nakamura, Gojobori, and Ikemura, 2000) and some additional sequences for *P. sylvestris*.

^a Based on Müller-Starck, Paradat, and Bergmann 1992.

^b Based on Moran, Bell, and Elridge 1988.

^c Based on Terauchi, Terachi, and Miyashita 1997.

^d Based on Miyashita et al. 1998.

^e Based on Abbot and Gomes 1989.

^f Based on Savolainen et al. 2000.

^g Based on von Treuren et al. 1997.

^h Based on Filatov and Charlesworth 1999.

ⁱ Based on Charlesworth and Yang 1998.

^j Based on Moriyama and Powell 1996.

^k Based on Nevo 1978.

^l Based on Cargill et al. 1999.

^m Based on Eyre-Walker et al. 1998.

ⁿ Based on Doebley and Goodman 1984.

closely linked sites is very small, N_e must be large to account for the lack of disequilibrium.

We also examined the data for signs of departure from demographic equilibrium, i.e., of postglacial expansion from several refugia, as explained earlier. A recent expansion could show up as an excess of singletons, which can be measured by Tajima's D . The value of -0.56 for the nuclear loci did not provide evidence for recent expansion. However, there are still local traces of possible refugia. The small Spanish populations may be remnants of a larger distribution. The observed lower level of DNA polymorphism in the Spanish population than in the Russian and Finnish populations agree with the previous data obtained with allozymes. The Spanish populations of Scots pine are slightly less variable than the populations from the continuous range of the distribution, with H_e of 0.325 and 0.363, respectively (Prus-Glowacki and Stephan 1994). In all, the data on the Scots pine main range are consistent with a large effective population size and mutation drift equilibrium.

High Mutation Rate in Pines?

Because *P. sylvestris* does not have higher nucleotide diversity compared with other species, we need to examine again the hypothesis of a high mutation rate in pines. As the mutation rate governs the rate of neutral substitution between species, the synonymous divergence between species (K_s) provides an estimate of the mutation rate. Thus, $\mu = K_s/2T$, where T is the time since divergence.

The best estimate of divergence time, based on fossil evidence, for *P. sylvestris* and *P. taeda* is 120 Myr (Millar 1998), and the rate of synonymous substitutions is 0.15×10^{-9} per year for *pal1* and about 0.05×10^{-9} per year for *Adh* (table 6). The *pal1* and the *Adh* genes have similar rates of neutral nucleotide substitutions, and both appear to evolve much more slowly in pines than in the monocots or dicots. Lu, Szmidt, and Wang (1998) have observed a low number of substitutions between *Ginkgo biloba* and *Larix sibirica* at the nuclear *cox1* locus (based on their data, $K_s = 0.016$). This K_s estimate suggests an even lower mutation rate in those lineages than in pines because *Ginkgo* and *Larix* have diverged before the pines (about 200 MYA). A similar low rate (0.24×10^{-9}) has also been found in the chloroplast gene *rbcl* of the chestnut tree (family Fagaceae), another group of long-lived woody perennials (Frascaria et al. 1993). As shown in table 6, the estimates of angiosperm mutation rates are higher by an order of magnitude. Note that even if we use a generation time of 25 years for pine, the per generation mutation rate estimates are not higher than in other plants in either *pal* or *Adh*. Lande, Schemske, and Schultz (1994) have suggested that mutation rates in pines must be much higher than those measured in annual plants to account for the levels of inbreeding depression that pines maintain. It seems that the observed high inbreeding depression in Scots pine fits the model by Lande, Schemske, and Schultz (1994) quite well (Koelewijn, Koski, and Savolainen 1999). Mutation rates per generation at nuclear genes

leading to chlorophyll deficiency have also been estimated to be about 10 times as high in pines and other woody plants as in annuals (Kärkkäinen, Koski, and Savolainen 1996; Klekowski 1998). Thus, the genome-wide per generation mutation rate to deleterious alleles gives a very different picture from the neutral mutation rate per site per year. This difference merits further study.

High Diversity at Many Marker Loci in Pines

Several measures of population diversity show pines to be quite variable. Scots pine has high levels of morphological, isozyme, RFLP, and microsatellite polymorphism (e.g., Karhu et al. 1996). Also, finding variable markers for mapping has been easy in many pine species (e.g., Sewell, Sherman, and Neale 1999). Enzyme gene markers have been used in several studies to quantify genetic variation at the species level in Scots pine (see review by Müller-Stark, Baradat, and Bergmann 1992). For example, in a study of 10 enzyme loci (2–5 alleles per locus) heterozygosity was between 0.075 and 0.482, with mean heterozygosity of 0.322 (Muona, Harju, and Kärkkäinen 1988). In 10 studies, the average heterozygosity values per population ranged from 0.25 to 0.39 for Scots pine (Müller-Stark, Baradat, and Bergmann 1992). How can the low nucleotide polymorphism at *pal1* be reconciled with the previous data? First, RFLPs measure both nucleotide polymorphism and length variation. Large direct repeats seem to be widespread in introns of conifer genes, at least at the *Adh* gene (Perry and Furnier 1996). We have also observed that pine genes often have indel variation, sometimes involving these repeats. These could lead to the high RFLP diversity. For microsatellites and random amplified polymorphic DNA (RAPDs), high mutation rates at repetitive areas of the genome occur in many organisms. Allozyme diversity is based on the number of alleles that give rise to electrophoretically different proteins. If we examine the *pal1* data, we find that the low per nucleotide diversity results in a high haplotype diversity of 0.95. Much of this diversity is because of silent sites. If we just take into account the three non-synonymous sites (243, 762, and 915), we observe five different haplotypes, with frequencies 0.8, 0.1, 0.05, and 0.05, with results in expected per locus heterozygosity of 0.35. Thus, it is possible that at least a part of the high locus-level diversity in pines is because of the lack of within-locus disequilibrium. This suggestion, of course, needs to be further studied at a larger number of loci.

Table 5 shows that *Dr. melanogaster*, like Scots pine, also has high enzyme variation compared with other species, but low nucleotide diversity compared with *Dr. simulans*. Enough comparisons are not yet available to account for these differences. We cannot exclude the possibility that the high variability at allozyme loci is partly maintained by selection, and thus not directly comparable with the patterns at presumably neutral sites.

Are Synonymous Sites Neutral: Codon Usage in *P. sylvestris*

Mutations occurring in the third positions of codons should be selectively neutral because they do not

Table 6
Rates of Synonymous Nucleotide Substitutions Per Year in *pal* and *Adh* Genes

	DIVERGENCE TIME (Myr)	<i>pal</i>		<i>Adh</i>	
		K_s	$r = \mu/\text{year}^a$	K_s	$r = \mu/\text{year}^a$
Pines					
<i>P. sylvestris/P. banksiana</i>	120 ^b	0	0	0.0125	0.052
<i>P. sylvestris/P. taeda</i>	120 ^b	0.0357	0.149		
Monocots					
<i>Oryza sativa/Z. mays</i>	50 ^c	0.1830	1.830	0.5147	5.147
Dicots					
<i>A. thaliana/Brassica oleracea</i>	25 ^d			0.6935	13.87
<i>A. thaliana/A. lyrata</i> subsp. <i>petrea</i>	5 ^d			0.1883	18.83
<i>Solanum tuberosum/Lycopersicon</i> <i>esculentum</i>	10 ^e	0.251	12.6		

^a The rates are in units of substitutions per site per 10^{-9} years ($r = K_s/2T$). Only exons have been used in these calculations. The calculations for the rates of synonymous nucleotide substitutions per year were based on the following sequences: *Adh2*-gene: *Arabidopsis lyrata* subsp. *petrea* AJ251278, *Arabidopsis thaliana* D48248, *Brassica oleracea* AF110434, *Oryza sativa* X16297, *Pinus banksiana* U48373, *Pinus sylvestris* (unpublished data), and *Zea mays* X01965; *pal*-gene: *Lycopersicon esculentum* M90692, *Oryza sativa* X16099, *Pinus banksiana* AF013481, *Pinus sylvestris* AF353967, *Pinus taeda* AF013481, *Solanum tuberosum* X63103, and *Zea mays* M95077.

^b Based on Millar 1998.

^c Based on Gaut et al. 1996.

^d Based on Koch, Haubold, and Mitchell-Olds 2000.

^e Based on Moniz de Sá and Drouin 1996.

cause amino acid changes (Kimura 1968, 1983). However, unequal use of synonymous codons points to selection at these sites also. Selection may favor some synonymous codons over the others, especially in highly expressed genes. There is a negative correlation between synonymous substitution rate and codon usage bias in many species (e.g., Sharp and Li 1986, 1989). Selection for alternative codons is expected to be rather weak, and thus efficient only in large populations. *Drosophila simulans* has high nucleotide diversity and codon bias, and *Dr. melanogaster* has lower diversity and less codon bias (see table 5). This is presumably because of the larger effective population size of *Dr. simulans* (Akashi 1997). A further point of evidence for the large effective population size of Scots pine could be found by examining codon bias. As Scots pine has very large populations, even weak selection could be efficient. Table 5 shows a comparison of codon bias in Scots pine and *P. radiata*, a species with very limited population size and less genetic diversity than Scots pine (Moran, Bell, and Elridge 1988). There is no evidence for increased codon bias in Scots pine relative to *P. radiata*, and thus at least not for the kind of selection at third positions because of major codon bias in large populations.

Selection at Linked Loci

Selection at linked loci has proved to have a major influence on patterns of diversity, especially in areas of low recombination (Aguadé, Miyashita, and Langley 1989). This could be because of selective sweeps (Kaplan, Hudson, and Langley 1989), background selection (Charlesworth, Morgan, and Charlesworth 1993), or alternating selection pressures (pseudohitchhiking) (Gillespie 2000). These selective explanations require that there be linkage disequilibrium between the neutral site and the target of selection, and then make additional

assumptions. We do not yet know anything about the distribution of the rate of recombination in the *P. sylvestris* genome. Scots pine has high levels of deleterious mutations (as required by background selection), selection at many loci is efficient in large populations (as required by hitchhiking), and the direction of selection at these many loci can vary (as assumed in the pseudohitchhiking model). However, the similarity of the nucleotide diversity at *pal* gene to the 11 other loci and the low level of disequilibrium within the gene suggest that disequilibrium will also be low between the locus itself and adjacent areas. Thus, selection at linked loci is not the main cause for the low diversity. More genomic areas need to be studied.

Acknowledgments

We gratefully acknowledge the financial support from the Research Council for Environment and Natural resources and the Center for International Mobility (CIMO). The Finnish Forest Research Institute, Dr. Igor Yakovlev, Mari-El State Technical University, Yoshkar-Ola, Russia, and Dr. Ricardo Alia, INIA (Instituto Nacional de Investigación Agropecuaria) kindly provided the seed samples. We thank Dr. Helmi Kuittinen and Prof. Pekka Pamilo for comments on the manuscript.

LITERATURE CITED

- ABBOT, R. J., and M. F. GOMES. 1989. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**:411–418.
- AGUADÉ, M., N. T. MIYASHITA, and C. H. LANGLEY. 1989. Reduced variation in the yellow-achaete-acute region in natural populations of *Drosophila melanogaster*. *Genetics* **122**: 607–616.
- AKASHI, H. 1997. Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. *Gene* **205**:269–278.

- BUTLAND, S. L., M. L. CHOW, and B. E. ELLIS. 1998. A diverse family of phenylalanine ammonia-lyase genes expressed in pine trees and cell cultures. *Plant Mol. Biol.* **37**:15–24.
- CARGILL, M., D. ALTSHULER, J. IRELAND et al. (17 co-authors). 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22**:231–238.
- CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**:1289–1303.
- CHARLESWORTH, D., and Z. YANG. 1998. Allozyme diversity in *Leavenworthia* populations with different inbreeding levels. *Heredity* **81**:453–461.
- DOEBLEY, J. F., and M. M. GOODMAN. 1984. Isoenzymatic variation in *Zea* (Gramineae). *Syst. Bot.* **9**:203–218.
- DOYLE, J. J., and J. L. DOYLE. 1990. Isolation of plant DNA from fresh tissue. *BRL Focus*. **12**:13–15.
- EXCOFFIER, L., P. E. SMOUSE, and J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479–491.
- EYRE-WALKER, A., R. B. GAUT, H. HILTON, D. L. FELDMAN, and B. S. GAUT. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. USA* **95**:4441–4446.
- FILATOV, D. A., and D. CHARLESWORTH. 1999. DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia PgiC* locus. *Genetics* **153**:1423–1434.
- FRASCARIA, N., L. MAGGIA, M. MICHAUD, and J. BOUSQUET. 1993. The *rbcL* gene sequence from chestnut indicates a slow rate of evolution in the *Fagaceae*. *Genome* **36**:668–671.
- GAUT, B. S., and M. T. CLEGG. 1993. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (*Poaceae*). *Genetics* **135**:1091–1097.
- GAUT, B. S., B. R. MORTON, B. C. McCAIG, and M. T. CLEGG. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc. Natl. Acad. Sci. USA* **93**:10274–10279.
- GILLESPIE, J. H. 2000. Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* **155**:909–919.
- GONCHARENKO, G. G., A. E. SILIN, and V. E. PADUTOV. 1994. Allozyme variation in natural populations of Eurasian pines: III. Population structure, diversity, differentiation and gene flow in central and isolated populations of *Pinus sylvestris* L. in Eastern Europe and Siberia. *Silvae Genet.* **43**:119–132.
- GRAUR, D., and W.-H. LI. 2000. *Fundamentals of Molecular Evolution*. Second edition. Sinauer Associates, Inc., Sunderland.
- GULLBERG, U., R. YAZDANI, D. RUDIN, and N. RYMAN. 1985. Allozyme variation in Scotch pine (*Pinus sylvestris*) in Sweden. *Silvae Genet.* **34**:193–201.
- HAMRICK, J. L., and M. J. W. GODT. 1996. Effects of life history traits on genetic diversity in plant species. *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* **351**:1291–1298.
- HARPENDING, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**:591–600.
- HILL, W. G., and A. ROBERTSON. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**:226–231.
- HUDSON, R. R. 1987. Estimating the recombination parameter of a finite population model without selection. *Genet. Res.* **50**:245–250.
- HUTTLEY, G. A., M. L. DURBIN, D. E. GLOVER, and M. T. CLEGG. 1997. Nucleotide polymorphism in the chalcone synthase-A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. *Mol. Ecol.* **6**:549–558.
- HYVÄRINEN, H. 1987. History of forests in northern Europe since the last glaciation. *Ann. Acad. Sci. Fenn. Ser. A III. Geol. Geogr.* **145**:7–18.
- INNAN, H., F. TAJIMA, R. TERAUCHI, and N. T. MIYASHITA. 1996. Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. *Genetics* **143**:1761–1770.
- KAPLAN, N. L., R. R. HUDSON, and C. H. LANGLEY. 1989. The “hitchhiking effect” revisited. *Genetics* **123**:887–899.
- KARHU, A., P. HURME, M. KARJALAINEN, P. KARVONEN, K. KÄRKKÄINEN, D. NEALE, and O. SAVOLAINEN. 1996. Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? *Theor. Appl. Genet.* **93**:215–221.
- KÄRKKÄINEN, K., V. KOSKI, and D. SAVOLAINEN. 1996. Geographical variation in inbreeding depression of Scots pine. *Evolution* **50**:111–119.
- KARVONEN, P., and O. SAVOLAINEN. 1993. Variation and inheritance of ribosomal DNA in *Pinus sylvestris* L. (Scots pine). *Heredity* **71**:614–622.
- KAWABE, A., H. INNAN, R. TERAUCHI, and N. T. MIYASHITA. 1997. Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. *Mol. Biol. Evol.* **14**:1303–1315.
- KAWABE, A., and N. T. MIYASHITA. 1999. DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. *Genetics* **153**:1445–1453.
- KAWABE, A., N. T. MIYASHITA, and R. TERAUCHI. 1997. Phylogenetic relationship among the section *Stenophora* in the genus *Dioscorea* based on the analysis of nucleotide sequence variation in the phosphoglucose isomerase (*Pgi*) locus. *Genes Genet. Syst.* **72**:253–262.
- KIMURA, M. 1968. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**:247–269.
- KIMURA, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, U.K.
- KLEKOWSKI, E. 1998. Mutation rates in mangroves and other plants. *Genetica* **102/103**:325–331.
- KOCH, M. A., B. HAUBOLD, and T. MITCHELL-OLDS. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* **17**:1483–1498.
- KOELEWIJN, H. P., V. KOSKI, and O. SAVOLAINEN. 1999. Magnitude and timing of inbreeding depression in Scots pine (*Pinus sylvestris* L.). *Evolution* **53**:758–768.
- KOSKI, V. 1970. A study of pollen dispersal a mechanism of gene flow in conifers. *Commun. Inst. For. Fenn.* **70**(4):1–78.
- KOSKI, V. 1971. Embryonic lethals of *Picea abies* and *Pinus sylvestris*. *Commun. Inst. For. Fenn.* **75**:1–30.
- LANDE, R., D. W. SCHEMSKE, and S. T. SCHULTZ. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. *Evolution* **48**:965–978.
- LANGE, B. M., M. TROST, W. HELLER, C. LANGEBARTELS, and H. SANDERMANN JR. 1994. Elicitor-induced formation of free and cell-wall-bound stilbenes in cell-suspension cultures of Scots pine (*Pinus sylvestris* L.). *Planta* **194**:143–148.
- LAUKKANEN, H., and T. SARJALA. 1997. Effect of exogenous polyamines on Scots pine callus *in vitro*. *J. Plant Physiol.* **150**:167–172.
- LIU, F., D. CHARLESWORTH, and M. KREITMAN. 1999. The effect of mating system differences on nucleotide diversity at

- the phosphoglucose isomerase locus in the plant genus *Leavenworthia*. *Genetics* **151**:343–357.
- LIU, F., L. ZHANG, and D. CHARLESWORTH. 1998. Genetic diversity in *Leavenworthia* populations with different inbreeding levels. *Proc. R. Soc. Lond. Biol. Sci.* **265**:293–301.
- LU, M. Z., A. E. SZMIDT, and X.-R. WANG. 1998. RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *coxI* gene. *Plant Mol. Biol.* **37**:225–234.
- MAYNARD SMITH, J., and J. HAIGH. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**:23–35.
- MILLAR, C. I. 1998. Early evolution of pines. Pp. 69–91 in D. M. RICHARDSON, ed. *Ecology and biogeography of Pinus*. Cambridge University Press, Cambridge, U.K.
- MIYASHITA, N. T., A. KAWABE, H. INNAN, and R. TERAUCHI. 1998. Intra- and interspecific DNA variation and codon bias of the alcohol dehydrogenase (*Adh*) locus in *Arabidopsis* and *Arabidopsis* species. *Mol. Biol. Evol.* **15**:1420–1429.
- MONIZ DE SÁ, M., and G. DROUIN. 1996. Phylogeny and substitution rates in angiosperm actin genes. *Mol. Biol. Evol.* **13**:1198–1212.
- MORAN, G. F., J. C. BELL, and K. ELRIDGE. 1988. The genetic structure and conservation of five natural populations of *Pinus radiata*. *Can. J. For. Res.* **18**:506–514.
- MORIYAMA, E. N., and J. R. POWELL. 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**:261–277.
- MÜLLER-STARCK, G., P. H. BARADAT, and F. BERGMANN. 1992. Genetic variation within European tree species. Pp. 23–47 in W. T. ADAMS, S. H. STRAUS, D. L. COPES, and A. R. GRIFFIN, eds. *Population genetics of forest trees*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- MUONA, O., and A. HARJU. 1989. Effective population sizes, genetic variability, and mating system in natural stands and seed orchards of *Pinus sylvestris*. *Silvae Genet.* **38**:221–228.
- MUONA, O., A. HARJU, and K. KÄRKKÄINEN. 1988. Background pollination in *Pinus sylvestris* seed orchards. *Scand. J. For. Res.* **4**:513–520.
- MUONA, O., and A. E. SZMIDT. 1985. A multilocus study of natural populations of *Pinus sylvestris*. Pp. 226–240 in H. R. GREGORIUS, ed. *Population genetics in forestry*. Lecture notes in biomathematics, Vol. 60. Springer Verlag, Berlin.
- NAKAMURA, Y., T. GOJOBORI, and T. IKEMURA. 2000. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**:292.
- NEI, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- NEI, M., and W.-H. LI. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**:5269–5273.
- NEVO, E. 1978. Genetic variation in natural populations: patterns and theory. *Theor. Popul. Biol.* **13**:121–177.
- PERRY, D. J., and G. R. FURNIER. 1996. *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proc. Natl. Acad. Sci. USA* **93**:13020–13023.
- PRUS-GLOWACKI, W., and B. R. STEPHAN. 1994. Genetic variation of *Pinus sylvestris* from Spain in relation to other European populations. *Silvae Genet.* **43**:7–14.
- REMINGTON, D. L., R. W. WHETTEN, B. H. LIU, and D. M. O' MALLEY. 1999. Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theor. Appl. Genet.* **98**:1279–1292.
- ROSEMANN, D., W. HELLER, and H. SANDERMANN JR. 1991. Biochemical plant responses to ozone: II. Induction of stilbene biosynthesis in Scots pine (*Pinus sylvestris* L.) seedlings. *Plant Physiol.* **97**:1280–1286.
- ROZAS, J., and ROZAS, R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**:174–175.
- SAVOLAINEN, O., C. H. LANGLEY, B. P. LAZZARO, and H. FREVILLE. 2000. Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Mol. Biol. Evol.* **17**:645–655.
- SCHNEIDER, S., D. ROESSLI, and L. EXCOFFIER. 2000. Arlequin Version 2.000: a software for population genetics analysis.
- SEWELL, M. M., B. K. SHERMAN, and D. B. NEALE. 1999. A consensus map for loblolly pine (*Pinus taeda* L.) I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* **151**:321–330.
- SHARP, P. M., and W.-H. LI. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**:28–38.
- SHARP, P. M., and W.-H. LI. 1989. On the rate of DNA sequence evolution in *Drosophila*. *J. Mol. Evol.* **28**:398–402.
- SINCLAIR, W. T., J. D. MORMAN, and R. A. ENNOS. 1999. The postglacial history of Scots pine (*Pinus sylvestris* L.) in Western Europe: evidence from mitochondrial DNA variation. *Mol. Ecol.* **8**:83–88.
- SMALL, R. L., J. A. RYBURN, and J. F. WENDEL. 1999. Low levels of nucleotide diversity at homologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Mol. Biol. Evol.* **16**:491–501.
- SORANZO, N., R. ALIA, J. PROVAN, and W. POWELL. 2000. Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Mol. Ecol.* **9**:1205–1211.
- STEPHAN, W., and C. H. LANGLEY. 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**:1585–1593.
- TERAUCHI, R., T. TERACHI, and N. T. MIYASHITA. 1997. DNA polymorphism in the *Pgi* locus of a wild yam, *Dioscorea tokoro*. *Genetics* **147**:1899–1914.
- VON TREUREN, R., H. KUITTINEN, K. KÄRKKÄINEN, E. BAENAGONZALEZ, and O. SAVOLAINEN. 1997. Evolution of microsatellites in *Arabidopsis petrea* and *Arabidopsis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol. Biol. Evol.* **14**:220–229.
- WANG, X. R., A. E. SZMIDT, and D. LINDGREN. 1991. Allozyme differentiation among populations of *Pinus sylvestris* (L.) from Sweden and China. *Hereditas* **114**:219–226.
- WATTERSON, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**:256–276.
- WHETTEN, R. W., and R. R. SEDEROFF. 1992. Phenylalanine ammonia-lyase from loblolly pine: purification of the enzyme and isolation of complementary DNA clones. *Plant Physiol.* **98**:380–386.
- WRIGHT, F. 1990. The “effective number of codons” used in a gene. *Gene* **87**:23–29.
- YAMADA, T., Y. TANAKA, P. SRIPRASERTSAK, H. KATO, T. HASHIMOTO, S. KAWAMATA, Y. ICHINOSE, H. KATO, T. SHIRAI, and H. OKU. 1992. Phenylalanine ammonia-lyase genes from *Pisum sativum*: structure, organ-specific expression and regulation by fungal elicitor and suppressor. *Plant Cell Physiol.* **33**:715–725.

WOLFGANG STEPHAN, reviewing editor

Accepted September 27, 2001