DOI: 10.1534/genetics.106.065102

Multilocus Patterns of Nucleotide Diversity, Linkage Disequilibrium and Demographic History of Norway Spruce [*Picea abies* (L.) Karst]

Myriam Heuertz,[*,†,1,2] **Emanuele De Paoli,**[‡,1] **Thomas Källman,**[*,1] **Hanna Larsson,**[*] **Irena Jurman,**[‡] **Michele Morgante,**[‡] **Martin Lascoux**[*,3] **and Niclas Gyllenstrand**[*,1]

[*]*Program in Evolutionary Functional Genetics, Evolutionary Biology Centre, Uppsala University, 75326 Uppsala, Sweden,*
[†]*Centre de Recherche Public-Gabriel Lippmann, L-4422 Belvaux, Luxembourg and* [‡]*Dipartimento di Scienze Agrarie ed Ambientali, Università di Udine, 33100 Udine, Italy*

Manuscript received August 20, 2006
Accepted for publication October 2, 2006

ABSTRACT

DNA polymorphism at 22 loci was studied in an average of 47 Norway spruce [*Picea abies* (L.) Karst.] haplotypes sampled in seven populations representative of the natural range. The overall nucleotide variation was limited, being lower than that observed in most plant species so far studied. Linkage disequilibrium was also restricted and did not extend beyond a few hundred base pairs. All populations, with the exception of the Romanian population, could be divided into two main domains, a Baltico–Nordic and an Alpine one. Mean Tajima's D and Fay and Wu's H across loci were both negative, indicating the presence of an excess of both rare and high-frequency-derived variants compared to the expected frequency spectrum in a standard neutral model. Multilocus neutrality tests based on D and H led to the rejection of the standard neutral model and exponential growth in the whole population as well as in the two main domains. On the other hand, in all three cases the data are compatible with a severe bottleneck occurring some hundreds of thousands of years ago. Hence, demographic departures from equilibrium expectations and population structure will have to be accounted for when detecting selection at candidate genes and in association mapping studies, respectively.

LEVEL of nucleotide polymorphism, extent and pattern of linkage disequilibrium (LD), and degree of population differentiation are fundamental population genetics parameters that are strongly influenced by evolutionary forces that acted in the past. Their analysis can therefore be used to infer past demographic history and selection events. Solid reconstructions of past demographic events based on a large number of loci are needed to detect genomic areas that are under selection since, if the population departs from the standard neutral model, current neutrality tests that compare the observed polymorphism pattern to that expected under the standard neutral model cannot be used (see, for example, Thornton and Andolfatto 2005). In a few intensively studied species, the availability of extensive genomic data and powerful coalescent-based estimation methods are enabling such reconstructions, thereby greatly facilitating the detection of loci under selection in genome scans (*e.g.*, Akey *et al.* 2002; Schaffner *et al.* 2005; Wright *et al.* 2005). In other organisms, while such fine-tuned reconstructions are still out of reach, more limited surveys of nucleotide variation, coupled to coalescent simulations still do allow the evaluation of different demographic models. For example, Haddrill *et al.* (2005) used multilocus neutrality tests, measures of linkage disequilibrium, and coalescent simulations to show that simple bottleneck models were sufficient to account for most, if not all, polymorphism features of *Drosophila melanogaster.* Such approaches have not yet been applied to conifer species, although they may be the key to the understanding of some of the intriguing patterns of nucleotide polymorphism that have emerged from initial surveys. Estimates of nucleotide diversity reported so far in conifers have been much lower than expected on the basis of their life-history traits and the high heterozygosity levels observed at allozyme loci for these species (Hamrick and Godt 1996). The average π_{silent} was 0.0064 in *Pinus taeda* (Brown *et al.* 2004) and ~0.0041 in *P. sylvestris* (Dvornyk *et al.* 2002; García-Gil *et al.* 2003). In Norway spruce, nucleotide diversity seems also low ($\pi_s = 0.0041$ for 21 EST loci sequenced across 12 individuals; S. degli Ivanissevich and M. Morgante, unpublished data). In *P. taeda*, Brown *et al.* (2004) concluded that the low nucleotide diversity could be the result of a particularly low mutation rate (on the order of 1.7×10^{-10}/bp/year, *i.e.*, an order of

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AM267499–AM268023 and AM268895–AM269360.

[1] These authors contributed equally to this study.

[2] *Present address:* Laboratoire d'Eco-Ethologie Evolutive - CP 160/12, Université Libre de Bruxelles. 50, Av. F.D. Roosevelt, B-1050 Bruxelles, Belgium.

[3] *Corresponding author:* Program in Evolutionary Functional Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 75326 Uppsala, Sweden. E-mail: martin.lascoux@ebc.uu.se

magnitude lower than in angiosperms) combined with a low effective population size (5.6×10^5) due to population fluctuations during the late Pleistocene and the Holocene. This low effective population size was derived from the relationship $\theta = 4N_e\mu$, using the mutation rate per generation, and hence a standard neutral model was assumed. An alternative explanation of the low nucleotide diversity could be the presence of repeated selective sweeps but this seems unlikely in conifers as current estimates suggest that LD does not extend beyond a few hundred or thousand base pairs (Neale and Savolainen 2004). However, LD is known to vary extensively along the genome, and at different scales (*e.g.*, Myers *et al.* 2005), and current estimates in conifers are based only on a handful of loci in a few species, so this picture might change drastically as data accumulate.

In this study, we surveyed DNA polymorphism at 22 loci in an average of 47 haplotypes from seven populations representative of the Norway spruce natural range. Eleven loci were candidate genes for seasonal growth cessation and the remaining ones were randomly chosen from an EST database. The latter are *a priori* not related to seasonal growth cessation, a trait showing strong clinal variation (Ekberg *et al.* 1979). The aim of this study was to assess nucleotide diversity, population structure, and LD and address the following questions:

i. Are nucleotide polymorphism and LD in spruce as limited as in other conifer species and do the patterns indicate departure from the standard neutral model?
ii. Do some of the candidate genes depart from the average pattern?
iii. Do nucleotide polymorphisms display patterns of population structure similar to allozymes and cytoplasmic markers? Those markers distinguished two main domains, one covering northeastern Russia and Scandinavia (Baltico–Nordic domain) and the other centered on the Alps and extending into Poland (Alpine and Central European domain, hereafter called the Alpine domain) (Lagercrantz and Ryman 1990; Bucci and Vendramin 2000; Vendramin *et al.* 2000; Sperisen *et al.* 2001). The domains mirror the natural distribution of the species into two main geographical areas with smaller, more isolated pockets in the Carpathians and the Balkans.
iv. Is it indeed so that the Alpine domain has a lower level of diversity than the Baltico–Nordic domain and went through a bottleneck as suggested by Lagercrantz and Ryman (1990) while the Baltico–Nordic domain is closer to an equilibrium neutral model?
v. If populations went through a bottleneck, what were its characteristics (time of occurrence and overall severity) and could a bottleneck help explain the particularly low level of nucleotide polymorphism?

TABLE 1

Geographical coordinates of the *Picea abies* populations analyzed in this study

Population	Latitude	Longitude
North Sweden	66°50′N	22°40′E
South Sweden	58°22′N	13°10′E
Russia	60°49′N	34°18′E
Germany	47°23′N	12°23′E
Switzerland	46°13′N	07°24′E
Romania	46°49′N	25°07′E
Italy	46°15′N	09°45′E

To address these questions various population growth and bottleneck models were evaluated through coalescent simulations.

MATERIALS AND METHODS

Plant material: *Picea abies* seeds were collected from nonadjacent maternal trees in seven natural populations or artificial populations representing local gene resources (Table 1). Seedlots were partitioned between the Uppsala and the Udine laboratories. Seeds were soaked in water overnight and haploid DNA was extracted from megagametophytes using a CTAB method (Doyle and Doyle 1990). In each population, both laboratories mostly used megagametophytes from the same individuals for sequencing but in a few cases additional megagametophytes from different individuals were included.

Sequencing: To identify functional candidate genes, we performed BLAST, BLASTX, and TBLASTX searches (Altschul *et al.* 1997) in the NCBI and the loblolly pine EST (http://pinetree.ccgb.umn.edu/) databases, using published sequences of genes from the photoperiod and vernalization pathways in model organisms, mainly *Arabidopsis thaliana* (Simpson and Dean 2002; Yanovsky and Kay 2003; Hayama and Coupland 2004). A total of 11 growth cessation candidate genes were chosen in *P. abies* (Table 2), showing similarity with the *A. thaliana* genes *co* (*constans*, Putterill *et al.* 1997), *cry1* (*cryptochrome1*, Lin *et al.* 1996), *ebs* (*early bolting in short days*, Piñero *et al.* 2003), *gi* (*gigantea*, Fowler *et al.* 1999), *pat1* (*phytochrome A signal transduction1*, Bolle *et al.* 2000), *phyA* and *phyB* (*phytochrome A* and *phytochrome B*, Sharrock and Quail 1989), and *vip3* (*vernalization independence 3*, Zhang *et al.* 2003). Consistent with the three phytochrome gene lineages reported in gymnosperms (*phyN*, *phyO*, and *phyP*; Schmidt and Schneider-Poetsch 2002), we identified multiple gene copies and pseudogenes within the *P. abies* phytochrome gene family by cloning (data not shown). Therefore, nonoverlapping regions of *phyN* and *phyP* genes were treated as different loci. PCR primers were designed from loblolly pine or Scots pine EST sequences or from *P. abies* specific sequences, obtained through RT–PCR and RACE reactions, using the Primer3 software (Rozen and Skaletsky 2000). Control loci *a priori* not involved in the photoperiod or vernalization pathways (*se121*, *se129*, *se1100*; *se1151*, *se1358*, *se1364*, *se1368*, *se1390*, *se1391*, *xy225*; *xy1420*) were selected from a pilot resequencing survey of 21 EST-based loci across 12 *P. abies* individuals (S. degli Ivanissevich and M. Morgante, unpublished data). Selection criteria included ease of amplification and sequencing with the amplification primers and the presence of at least two polymorphic sites detected in the pilot survey. All genes were amplified from haploid megagametophyte DNA with the Phusion DNA Polymerase (Finnzymes,

TABLE 2

Nucleotide variation, haplotypic diversity, and neutrality tests in 22 *Picea abies* loci sequenced across seven populations

		Nucleotide diversity															
		Total				Nonsynonymous sites				Silent sites				Haplotype diversity		Neutrality tests	
Gene	n	L	S (singl.)	θ_{Wt}	π_t[a]	L	S	θ_{Wa}	π_a	L	S	θ_{Ws}	π_s	N_h (SD)	H_e (SD)	D[a]	H
col1	46	3,196	76 (28)	5.41	3.03	881	1	0.26	0.47	2263	74	7.44	4.08	36	0.985 (0.009)	−1.57**	−5.47
cry	52	918	4 (2)	0.97	0.57	595	4	1.49	0.87	321	0	0	0	4	0.418 (0.076)	−0.93	−4.88
ebs	50	730	16 (8)	4.89	2.26	317	2	1.41	0.25	407	14	7.68	3.86	12	0.481 (0.087)	−1.67*	—
gi	48	772	7 (3)	2.04	1.28	243	2	1.85	1.56	521	5	2.16	1.17	7	0.546 (0.073)	−1.00	−1.21
pat1	40	420	3 (1)	1.69	1.95	162	1	1.45	2.37	256	2	1.84	1.70	3	0.396 (0.077)	0.35	−0.04
phynrI	54	759	8 (5)	2.33	1.22	585	4	1.50	0.43	171	4	5.14	3.94	8	0.619 (0.050)	−1.27	0.56
phynrII	35	689	2 (1)	0.71	0.24	535	1	0.45	0.11	152	1	1.60	0.73	3	0.165 (0.082)	−1.28*	0.06
phyo	44	1,776	19 (8)	2.47	1.58	1016	5	1.13	1.35	759	14	4.24	1.88	20	0.910 (0.027)	−1.16	0.21
phyP	49	794	4 (1)	1.13	1.15	599	1	0.37	0.66	193	3	3.49	2.67	5	0.509 (0.073)	0.04	−0.30
phyP2	53	273	5 (2)	4.08	2.05	211	2	2.09	0.53	62	3	10.64	7.20	6	0.440 (0.078)	−1.18	−1.42
vip3	54	762	6 (3)	1.73	0.57*	353	1	0.62	0.10	400	5	2.74	0.99	6	0.234 (0.077)	−1.68**	0.41
se121	41	440	4 (3)	2.14	0.55	ND	ND	ND	ND	ND	ND	ND	ND	5	0.230 (0.086)	−1.76**	0.23
se129	49	275	2 (0)	1.64	1.85	ND	ND	ND	ND	ND	ND	ND	ND	3	0.471 (0.068)	0.24	−0.73
se1100	40	346	6 (0)	4.12	3.95	83	0	0	0	263	6	5.36	5.19	7	0.831 (0.031)	−0.09	0.53
se1151	49	480	8 (5)	3.77	2.11	ND	ND	ND	ND	ND	ND	ND	ND	9	0.687 (0.076)	−1.19	0.40
se1358	49	447	8 (3)	4.01	2.88	355	4	2.53	1.54	92	4	9.77	8.04	8	0.684 (0.056)	−0.78	0.32
se1364	47	552	4 (1)	1.64	1.32	228	0	0	0	321	4	2.83	2.28	5	0.423 (0.080)	−0.44	0.59
se1368	47	429	5 (3)	2.66	1.04	87	1	2.59	0.49	340	4	2.67	1.19	4	0.201 (0.048)	−1.49	−2.98
se1390	49	495	13 (4)	5.89	4.83	309	4	2.91	2.88	182	9	11.06	8.25	15	0.922 (0.015)	−0.54	1.33
se1391	47	503	4 (1)	1.80	1.02	ND	ND	ND	ND	ND	ND	ND	ND	4	0.304 (0.083)	−0.99	0.46
xy225	48	209	6 (3)	6.47	3.42	50	0	0	0	155	6	8.74	4.63	7	0.582 (0.074)	−1.21	0.62
xy1420	49	571	20 (4)	7.86	6.81*	400	5	2.80	2.31	169	15	21.24	17.56	21	0.955 (0.011)	−0.56	−4.22
Total	—	15,836	230 (89)	—	—	7109	38	—	—	7288	175	—	—	207	—	—	
Average	47	719	10.5 (4)	3.16	2.08	—	—	1.30	0.88	—	—	5.81	3.99	9 (8)	0.545 (0.254)	−0.92	−0.74

n, sample size; L, length in base pairs; S (singl.), number of segregating sites (number of singletons); N_h (SD), number of haplotypes (standard deviation); H_e (SD), Nei's haplotypic diversity (standard deviation); D, Tajima's D-statistic; H, Fay and Wu's H-statistic. Indels are excluded from the estimates. Nucleotide diversity estimates (θ_w and π) are $\times 10^3$. ND: loci for which no similarity was found in Blast searches were considered only in the calculation of the total nucleotide variation.

[a] Values that are significantly different from the average, *i.e.*, falling outside a 95 (99)% confidence interval obtained from standard coalescent simulations with recombination rate 8.5×10^{-3} for *col1* and 5.3×10^{-3} for all other genes (see MATERIALS AND METHODS), are indicated by * (**). Tajima's D was nonsignificant when no recombination was assumed.

Espoo, Finland) or AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA) and directly sequenced from the PCR product either with BigDye v3.1 and run on a ABI 3730 (Applied Biosystems) or with Dyenamic ET terminators and run on a MegaBace 1000 (GE Healthcare, Piscataway, NJ). Most gene regions were covered by two or more reads. Sequences were base called and assembled with PHRED and PHRAP (Ewing and Green 1998; Ewing *et al.* 1998) and visualized and edited with CONSED (Gordon *et al.* 1998). A putative SNP was considered true when PHRED quality scores of the different variants exceeded 25.

Nucleotide diversity analysis: Estimates of standard population genetics parameters and neutrality test statistics were calculated for each locus with the DnaSP v. 4.0 software (Rozas *et al.* 2003). Insertions or deletions are reported, but were excluded from further analyses. The level of polymorphism for each locus was estimated as both haplotype and nucleotide diversities.

Population structure: Population differentiation was first estimated with Wright's fixation index F_{ST} (Wright 1951). F-statistics of each gene were computed from allele frequencies as variance component ratios with the locus-by-locus AMOVA approach (Excoffier *et al.* 1992) implemented in the Arlequin software (Schneider *et al.* 2000). Single-gene F_{ST}'s and overall F_{ST} were obtained by summing variance components over nucleotide loci ($\sum_{\text{loci}} V_a / \sum_{\text{loci}} V_t$) according to Weir and Cockerham (1984). The significance of F_{ST} was tested by comparing the observed value with the distribution of F_{ST} after 10,000 permutations of sequences among populations.

The genetic structure of the Norway spruce sample was also investigated with the model-based clustering algorithm implemented in STRUCTURE v. 2.1 (Pritchard *et al.* 2000; Falush *et al.* 2003). We used the admixture model on a subset of the data represented by 105 parsimony informative unlinked loci, that is, loci between which Fisher's exact test with Bonferroni correction (see *Linkage disequilibrium*) was not significant. Ten runs with a burn-in of 100,000 and a run length of 500,000 iterations were performed for a number of clusters from $K = 1$ to $K = 7$, allowing for correlation of allele frequencies between clusters. As individuals are assigned to clusters to achieve linkage and Hardy–Weinberg equilibria, the fact that different loci were sequenced from different megagametophytes from the same individual should not affect the results. When, as occurred in a few cases, loci were sequenced in gametophytes from different mother trees, sequences were not pooled to create haplotypes. Instead other loci were coded as missing values. In any case, it should be emphasized that we use Structure here primarily as a tool to explore the data rather than a first step in an association study, which would clearly require a stricter control of the population structure. More generally, and as pointed out by Setakis *et al.* (2006), the notion of subpopulation is a theoretical construct that will only imperfectly reflect reality and therefore the resulting clusters should not be interpreted too literally.

Linkage disequilibrium: The level of linkage disequilibrium between parsimony-informative sites within genes was estimated as r^2, the mean squared correlation in allelic state between pairs of SNPs, using DnaSP. Significance of the associations between SNPs was determined with Fisher's exact test with Bonferroni correction. The overall decay of LD with physical distance within genes was evaluated by nonlinear regression of r^2 on distance between sites in base pairs (Remington *et al.* 2001). We used the Hill and Weir (1988) expectation of r^2 between adjacent sites,

$$E(r^2) = \left[\frac{10 + C}{(2 + C)(11 + C)}\right]\left[1 + \frac{(3 + C)(12 + 12C + C^2)}{n(2 + C)(11 + C)}\right],$$

where C is the population recombination parameter ($\rho = 4N_e r$) and n the sample size, and replaced C by $C \times$ distance in base pairs when fitting the formula to our data using the nonlinear regression (nls) function in the R software (R Development Core Team 2005).

Statistical test of neutrality and evaluation of alternative models: To test for departure from the standard neutral model the mean value of Tajima's D and Fay and Wu's H over loci was compared with the distribution of mean values from coalescent simulations using code kindly provided by P. Andolfatto and described in Haddrill *et al.* (2005). Both test statistics compare two estimates of θ: Tajima's D measures the standardized difference between π and θ_W (Tajima 1989) while Fay and Wu's H (Fay and Wu 2000) measures the difference between π and θ_H. The former is most sensitive to an excess of rare variants whereas the latter is most sensitive to an excess of high-frequency-derived variants. Both D and H are expected to be close to zero under the standard neutral model. All tests were carried out with recombination, as the lack of recombination makes the tests overly conservative. An estimate of the population recombination parameter $\rho = 4N_e r$ was obtained with the composite-likelihood method of Hudson (2001) adapted to finite-sites models as implemented in the software LDHat v. 2.0 (McVean *et al.* 2002), for *col1* (this study) and for two other genes, *ft1* and *toc1* (our unpublished data). Estimates of ρ were 8.5×10^{-3}, 4.7×10^{-3}, and 2.6×10^{-3}/bp, respectively. For *col1* we used the estimated ρ-value for that gene, while for all other genes we used the average of the three estimates, 5.3×10^{-3}. The ancestral state of nucleotides, required for Fay and Wu's *H*-test, was inferred by using a single sequence of *P. mariana*, *P. glauca*, *P. sitchensis*, or, in the case of *col1*, *P. sylvestris* as an outgroup.

Coalescent simulations were also used to evaluate two types of alternative models: exponential growth models and bottleneck models followed by exponential growth. Briefly, it is now well established that the ranges of tree taxa went through cycles of contraction and expansion in response to climate changes during the late Quaternary (Bennett 1997). In Norway spruce, as in most species, however, the severity of the contractions, the size and location of the refugia, and the rate of the ensuing growth are still poorly characterized and, consequently, we modeled bottlenecks of various severity and ages and considered models with different growth rates. We also assessed the effects of repeated bottlenecks (data not shown). Importantly, because all times are in units of effective population size for which we do not have any independent estimate, the age and severity of the bottleneck cannot be defined exactly. So our primary aim in this study was to test whether the data could be better explained by a bottleneck than by the standard neutral model in the first place rather than to obtain a fine characterization of that bottleneck. The difficulty in characterizing a bottleneck is compounded by the fact that the effect of a bottleneck on the frequency distribution of mutations segregating in a population depends on the time at which the bottleneck ended and its strength, which is approximately a function of the ratio of its severity (the magnitude of the reduction in population size) to its duration. Hence different combinations of the three parameters can lead to the same nucleotide frequency spectrum (Fay and Wu 1999; Voight *et al.* 2005; Wright *et al.* 2005). The bottleneck models were tested over a grid of parameter values: the severity varied between 0.0004 and 0.001 and the time at which the bottleneck ended (*t*_end) varied between 0.001 and 0.0095. The length of the bottleneck was fixed to 0.0015 in all bottleneck models. Time measures are in units of $4N_0$ generations from the present and the severity of the bottleneck is in units of the current population size. The coalescent simulations were run with recombination estimated as above. Because we would actually need to run the simulation using the recombination rate in the ancestral population, for which we have no estimates, to assess the robustness of the results, we also ran a subset of demographic scenarios with twice that value and without recombination. Details on the methods can be found in supplemental material at http://www.genetics.org/supplemental/.

RESULTS

Nucleotide variation: Sequence variation was obtained for all 22 loci (supplemental Table 1 at http://www.genetics.org/supplemental/) in an average of 47 megagametophytes, ~7 from each of seven *P. abies* populations. A total of 16,161 bp were aligned over the 22 genes, of which more than half was coding sequence, resulting in a total of ~760 kb of sequence information across individuals. Insertions/deletions (indels) covered 130 bp. They comprised seven microsatellites with a motif length of 1, 3, 4, or 9 bp in, respectively, *ebs*, *phyo*, and *se1390* (1 bp); *se1364* and *xy225* (3 bp); *ebs* (4 bp); and *se1100* (9 bp). The microsatellites were located in noncoding regions, except in the case of *se1364*, where a 3-bp repeat in the coding region produced no shift in the reading frame. The remaining indels were located in noncoding regions, namely an 11- and a 54-bp stretch in *col1*, a 27-bp stretch in *se1100*, and an 8-bp stretch in each *pat1* and *se1368*. Indels were excluded from further analyses.

We identified a total of 230 segregating sites, of which 89 were singletons and 141 were parsimony-informative sites (Table 2). This corresponds to 1 SNP every 69 bp. One parsimony-informative site with three variants was found in *se1420*; it was excluded from further analyses. Forty (17.4%) SNPs were amino acid replacement substitutions. Statistics of sequence variation are summarized in Table 2. Total nucleotide diversity π_t was between 0.0002 and 0.0068 (average $\pi_t = 0.0021$) and silent nucleotide diversity, including synonymous and noncoding positions, varied between 0 and 0.0176 (average $\pi_s = 0.0039$; loci for which no similarity was found in Blast searches were considered only in the calculation of the total nucleotide variation). Nonsynonymous nucleotide diversity was on average 4.2 times smaller than

TABLE 3

Pairwise measures of population differentiation

	Northern Sweden	Southern Sweden	Germany	Switzerland	Italy	Romania
Russia	0.003 (0.410)	0.030 (0.064)	0.136 (<0.001)	0.186 (<0.001)	0.173 (<0.001)	0.262 (<0.001)
Northern Sweden		0.040 (0.023)	0.125 (<0.001)	0.122 (<0.001)	0.115 (<0.001)	0.234 (<0.001)
Southern Sweden			0.009 (0.297)	0.085 (<0.001)	0.038 (0.063)	0.212 (<0.001)
Germany				0.021 (0.135)	−0.032 (0.914)	0.194 (<0.001)
Switzerland					0.044 (0.036)	0.233 (<0.001)
Italy						0.222 (<0.001)

F_{ST}-values are given above the diagonal with their *P*-values in parentheses.

silent nucleotide diversity and varied from 0 to 0.0029 (average $\pi_a = 0.0009$). Control loci were more polymorphic than candidate loci [π_t(controls) $= 0.0027 \pm 0.0019$ (SD) *vs.* π_t(candidates) $= 0.0014 \pm 0.0008$ (SD)]. It is difficult to speculate on the cause of this difference since (i) it might result simply from differences in sampling approach for the two groups of genes and (ii) different species were used as outgroups for the different genes, making estimates of the average mutation rate in the two groups complicated. With these caveats in mind, considering only the genes for which the same outgroup, *P. taeda*, was used, the average divergence at silent sites among control loci was larger than the average divergence among the candidate loci but the standard deviations were very large [0.1777, SD = 0.115 ($n = 6$) *vs.* 0.077, SD = 0.065 ($n = 6$)].

Population structure: F_{ST}-values varied between 0 and 0.289 among loci and revealed substantial differentiation among the seven populations, with an overall value of $F_{ST} = 0.117$ over all 229 SNPs (supplemental Table 2 at http://www.genetics.org/supplemental/). Romania was the most differentiated population in pairwise comparisons ($0.194 \leq F_{ST} \leq 0.262$, Table 3).The two-level variance partitioning revealed a high differentiation ($F_{ST} = 0.147$) between the Baltico–Nordic domain (Northern Sweden, Southern Sweden, Russia), the Alpine domain (Italy, Switzerland, Germany), and the Carpathian domain (Romania) (data not shown). Populations within the Baltico–Nordic domain were significantly, though weakly differentiated ($F_{ST} = 0.025$, $P \leq 0.05$), whereas populations from the Alpine domain were not significantly differentiated ($F_{ST} = 0.015$, data not shown).

The STRUCTURE program revealed the highest likelihood for $K = 4$ clusters (average log probability of data Ln $P(D) = -1452.01 \pm 10.39$, SD); however, biologically meaningful genetic structure was already detected at $K = 3$ with Ln $P(D) = -1518.26 \pm 34.16$ (SD). With $K = 3$, the Baltico–Nordic, the Alpine, and the Carpathian domains were essentially distinguished (Figure 1). The Romanian population was the most distinct with 92.7 ± 0.32% (SD) of ancestry in cluster 1. All other populations were fairly admixed. Populations from the Baltico–Nordic domain had their largest proportion of ancestry, 59.0 ± 0.69% (SD) in cluster 2 while populations from the Alpine domain had theirs in cluster 3 (68.0 ± 0.66%). The populations from southern Sweden (41.0% in cluster 2, 36.8% in cluster 3) and Italy (33.9% in cluster 1, 42.8% in cluster 2) were even more admixed. With $K = 4$, the structure of $K = 3$ was confirmed and a fourth cluster accounted for 20–27% of the ancestry of populations from southern Sweden, Germany, Switzerland, and Italy.

The results on among-population differentiation suggest different evolutionary histories for the Baltico–Nordic *vs.* the Alpine part of the Norway spruce range and a particular situation for Romania. Diversity estimates were lowest in Romania, with $\pi_T = 0.0012$ lower than $\pi_T \geq 0.0016$ in other populations (unilateral paired *t*-tests: $P \leq 0.05$ except for Germany where $P = 0.053$ and Switzerland where $P = 0.091$). Mean genetic diversities in the Baltico–Nordic and in the Alpine domains were very close ($\pi_T = 0.0022$ *vs.* $\pi_T = 0.0017$).

Linkage disequilibrium: A low level of linkage disequilibrium was observed within genes, with an average $r^2 = 0.115$ and 75 significant exact tests after Bonferroni

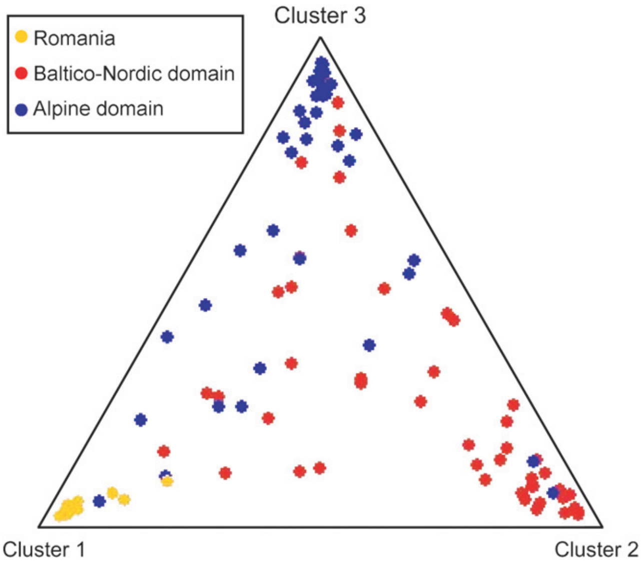

FIGURE 1.—Structure analysis of the seven populations when $K = 3$ clusters are assumed. The Baltico–Nordic domain includes southern Sweden, northern Sweden, and Russia and the Alpine domain includes Switzerland, Germany, and Italy.

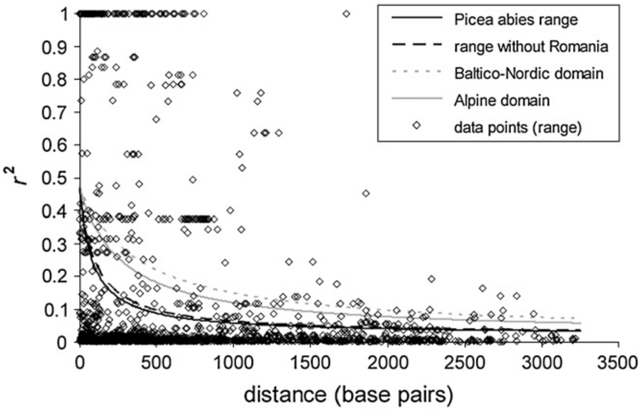

FIGURE 2.—Plot of the squared correlation of allele frequencies (r^2) *vs.* distance in base pairs between polymorphic sites across 22 loci for different subsets of populations.

correction among 1411 pairwise comparisons between informative SNPs. LD decayed fast within genes, with r^2 dropping below 0.2 within ~100 bp (Figure 2).

Statistical tests of neutrality and evaluation of alternative models: Mean values of both Tajima's *D*- and Fay and Wu's *H*-statistics were negative, with values of −0.88 and −0.74, respectively (excluding *ebs* for which no outgroup was available; Table 4). Coalescent simulations using both statistics led to the rejection of the standard neutral model and population growth, but simulations that assumed a severe and rather ancient bottleneck followed by moderate population growth were consistent with the data. Table 4 gives the values of mean Tajima's *D* and mean Fay and Wu's *H* for a subset of the models that were tested. Tajima's *D* obtained from simulations under the standard neutral model is significantly larger than the observed value, whereas under the growth model Tajima's *D* no longer differs from the observed value but Fay and Wu's *H* is now significantly larger than the observed value. Various growth models were tested (supplemental Table 3 at http://www.genetics.org/supplemental/): none led to negative values for both mean *D* and mean *H*. Unless it was extremely severe a recent bottleneck was inconsistent with the data as it led to an excess of common variants and a positive Tajima's *D* or required an extremely large ancestral effective population size (θ as large as 16; Figure 3). On the other hand, as long as the bottleneck is ancient enough, the data can be explained by different combinations of time at which the bottleneck ended and severity without requiring unrealistically large θ-values (Figure 3). The same analysis was also carried out within the Baltico–Nordic and Alpine domains and led to similar conclusions, the acceptance regions being somewhat larger in the Baltico–Nordic domain than in the Alpine domain (supplemental Figures 1 and 2 at http://www.genetics.org/supplemental/).

Only five genes showed significant Tajima's *D*-values, namely *col1*, *ebs*, *phynrII*, *vip3*, and *se121* when Tajima's *D* and Fay and Wu's *H* were calculated for individual loci (Table 2). To assess whether demography alone could explain those departures or whether the frequency spectrum at those loci would still depart from the rest of the genome when demography is taken into account we tested them against a bottleneck model that could not be rejected globally. *Ebs* was not considered as we did not have an outgroup and *phynrII* was discarded because it had only two segregating sites. The bottleneck model was accepted for *col1* and *se121* but was rejected for *vip3* (Table 5). Additional factors, such as selection or a more complex demographic model, might therefore need to be invoked to account for the polymorphism at *vip3*.

TABLE 4

An evaluation of alternative demographic models for the total population and the Baltico–Nordic and the Alpine domains

	Total[e]			Baltico–Nordic domain[e]			Alpine domain[e]		
	Mean π[a]	Mean *D*	Mean *H*	Mean π	Mean *D*	Mean *H*	Mean π	Mean *D*	Mean *H*
Observed	1.47	−0.88 [0.38]	−0.74 [3.84]	1.53	−0.55 [1.00]	−0.27 [3.28]	1.59	−0.66 [0.45]	−0.26 [1.45]
Model									
SNM[b]	1.47	−0.02 ($<10^{-4}$)	−0.000 (0.015)	1.53	−0.03 (0.005)	−0.00 (0.180)	1.59	−0.03 (0.001)	−0.00 (0.209)
Growth[c]	1.47	−0.80 (0.337)	0.68 ($<10^{-4}$)	1.54	−0.82 (0.955)	0.73 (0.000)	1.59	−0.80 (0.815)	0.76 (0.000)
Bottleneck[d]	1.45	−0.28 (0.079)	−1.75 (0.693)	1.53	−0.31 (0.304)	−1.81 (0.901)	1.59	−0.29 (0.233)	−1.96 (0.860)

P-values for the observed means under the model simulated are given in parentheses. The numbers within brackets are the variances across loci of the parameters.

[a] Average π per locus across loci (the average number of sites surveyed is 719 bp), mean Tajima's *D* across loci, and mean Fay and Wu's *H* across loci.

[b] Standard neutral model.

[c] The growth rate was $G = 10$. $\theta = 4.78$.

[d] We assumed a population shrinking at rate 10 up to time $t_1 = 0.003 \times 4N_e$ before present (this represents population growth in the forward direction), then going through a bottleneck of severity $f = 0.0005$ until $t_2 = 0.0035 \times 4N_e$, and then having an ancestral population the same size as the current population. Assuming that $N_e = 500{,}000$ and a generation time of 25 years, $t_1 = 150{,}000$. If we assume $N_e = 10^6$, $t_1 = 300{,}000$. In the first scenario the bottleneck would last 25,000 years. $\theta = 10.03$.

[e] The analysis was based on 21 loci in the total data set and the Baltico–Nordic domain and 19 loci in the Alpine domain as *phynrII* and *phyP2* were monomorphic in the latter.

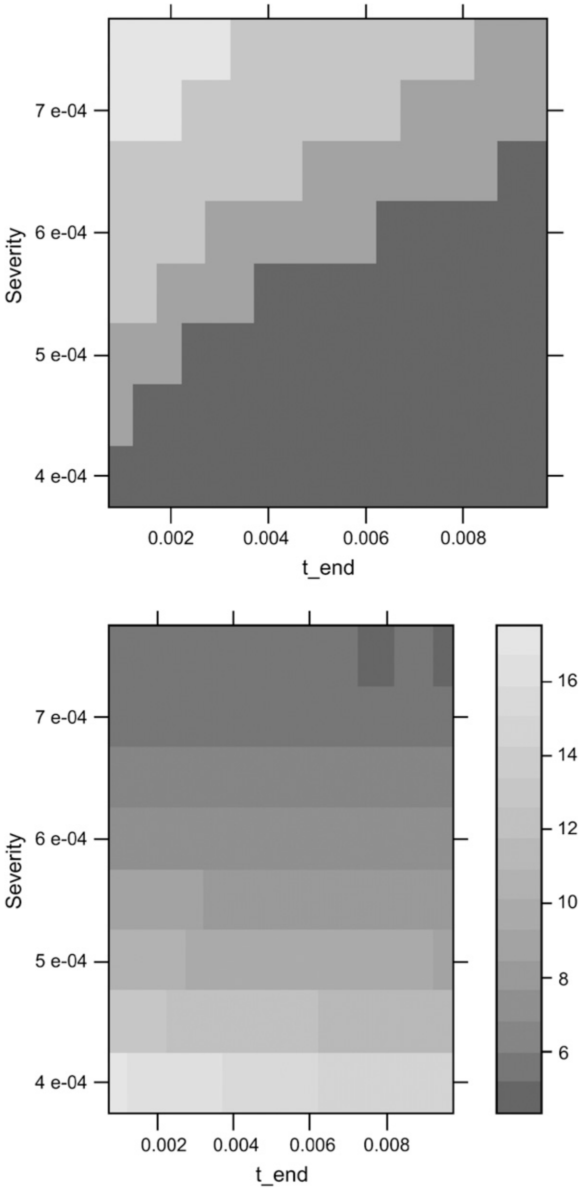

FIGURE 3.—Evaluation of different bottleneck models in the total data. (Top) Significance level of the multilocus neutrality test for different combinations of severity and time at which a bottleneck ended (*t*_end). The duration of the bottleneck was 0.0015. The *P*-value reported was in all cases that for Tajima's *D*. The *P*-value for Fay and Wu's *H* was always >0.05. From darker to lighter shading: $P > 0.05$, $0.01 < P \leq 0.05$, $0.001 < P \leq 0.01$, and $P \leq 0.001$. (Bottom) Corresponding average θ-values used in coalescent simulations. Lightest shading, $\theta > 16$; darkest shading, $\theta < 5$.

However, we note that no significant signal of selection could be detected on any SNP using BEAUMONT and BALDING's (2004) method (data not shown).

DISCUSSION

Norway spruce has a low to moderate level of nucleotide diversity ($\pi_s = 0.0039$, $\theta_{Ws} = 0.0058$), low levels of LD, which decayed by 50% within <100 bp, and a moderate level of population structure ($F_{ST} = 0.12$). Using multilocus tests based on summary statistics of the allele frequency spectrum we showed that the standard neutral model can be rejected and that a severe bottleneck predating the Last Glacial Maximum is sufficient to explain the data. This is true when all populations are considered but also within both the Baltico–Nordic and Alpine domains when those are analyzed separately. Hence, although nucleotide diversity was slightly higher in the Baltico–Nordic than in the Alpine domain the two domains seem to have experienced rather similar demographic histories.

Nucleotide diversity: The average level of silent nucleotide diversity in *P. abies*, $\pi_s = 0.0039$, confirmed earlier results from S. DEGLI IVANISSEVICH and M. MORGANTE (unpublished data), who found $\pi_s = 0.0041$ for 21 EST loci sequenced across 12 individuals (note that the 11 control loci we analyzed were selected on the basis of that study), and supports the contention that conifers are characterized by a low level of nucleotide diversity. Compared to other conifers, the level of silent polymorphism was of the same order of magnitude as that in *Cryptomeria japonica* ($\pi_s = 0.0038$ across 7 genes, KADO *et al.* 2003) and *P. sylvestris* ($\pi_s \approx 0.0041$ across 14 genes, DVORNYK *et al.* 2002; GARCÍA-GIL *et al.* 2003), but lower than that in *P. taeda* ($\pi_s = 0.0064$ across 19 wood-production candidate genes, BROWN *et al.* 2004; $\pi_s = 0.0079$ across 18 drought stress candidate genes, GONZÁLEZ-MARTÍNEZ *et al.* 2006). These estimates of silent nucleotide diversity are higher than that in soybean ($\pi_s = 0.0015$, ZHU *et al.* 2003) but twofold lower than that in *A. thaliana* ($\pi_s = 0.0083$, SCHMID *et al.* 2005) and an order of magnitude lower than estimates in aspen ($\pi_s = 0.0160$, INGVARSSON 2005) and wild relatives of maize ($\pi_s = 0.012$–0.013, TIFFIN and GAUT 2001). Hence, our results indicate that nucleotide diversity in the Norway spruce gene pool is indeed low.

Variation in average nucleotide diversity estimates across species can be caused by a combination of factors such as differences in individual sampling strategies, parts of the genome considered, selection, demographic history, and differences in mutation rate. The studies cited above were based on samples covering the entire species distribution ranges, or wild and cultivated genotypes of different origins, so differences in individual sampling are unlikely to account for the magnitude of the variation in estimates among species. Silent variation varied 30-fold across genes in our study and 50-fold in *P. taeda* (BROWN *et al.* 2004), so the studied genes do not appear to be biased toward a particular group. Because the EST genes were selected to be variable, our estimate might even be biased upward. Estimates of average polymorphism could also be reduced by selective sweeps that diminish variation at and around particular genes or by purifying selection. However, the limited amount of LD makes selective

TABLE 5

Evaluation of a bottleneck model at individual genes that depart from the standard neutral model

			SNM			Bottleneck[a]		
Gene	Test statistic	Observed value	0.025	0.5	0.975	0.025	0.5	0.975
col1	*D*	−1.57	*−0.99*	*−0.11*	*0.96*	−2.51	−2.10	3.68
	H	−5.47	−12.17	0.72	7.65	−67.64	0.30	0.75
Vip3	*D*	−1.68	*−1.57*	*−0.08*	*1.84*	−1.47	0.00	2.80
	H	0.41	−2.96	0.28	1.32	*−3.89*	*0.00*	*0.32*
se121	*D*	−1.76	*−1.57*	*−0.06*	*1.99*	−1.77	0.00	3.15
	H	0.23	−2.36	0.24	1.08	−7.76	−0.00	0.36

SNM: standard neutral model. The significant departures are in italics. The values are the 0.025, 0.5, and 0.975 quantiles.

[a] The bottleneck model used here is the same as the one described in Table 3.

sweeps an unlikely explanation for low nucleotide diversity and recurrent hitchhiking events would not account for the negative values of Fay and Wu's *H* (Przeworski 2002; Haddrill *et al.* 2005). Similarly, models of weak negative selection predict an excess of low-frequency-derived mutations and hence a positive value of Fay and Wu's *H*, which is not consistent with our data. In brief, while selection may partly explain the low level of nucleotide variation at individual loci it does not seem to be sufficient to explain the low level across loci.

Two possible explanations therefore remain for the relatively low level of polymorphism in conifers, namely demographic history and/or mutation rates. On the basis of sequence variation at 19 nuclear genes (amounting to a total of almost 18,000 bp) Brown *et al.* (2004) estimated the substitution rate per year to be 1.17 × 10^{-10} in *P. taeda*, a value similar to that reported for *P. sylvestris* (Dvornyk *et al.* 2002) and an order of magnitude lower than angiosperm mutation rates. This estimate was based on a divergence time between *P. pinaster* and *P. taeda* of 120 million years. There are grounds, however, to question this mutation rate estimate. First, the divergence time retained to calculate it corresponds to the early diversification of the Pinaceae in the early Cretaceous (∼120–140 MYA), not to the divergence time of two species within the genus. Wang *et al.* (2000) inferred that Pinus species diverged from one another in the early to mid-Cretaceous (∼70 MYA), which is consistent with the first appearance of Pinus in the fossil record in the early Cretaceous. Picea species appeared later on, in the middle Pliocene (∼45 MYA), and apparently diversified ∼20 MYA, if not later (Bouillé and Bousquet 2005). Hence 120 MYA is likely to be a gross overestimate of the divergence between *P. pinaster* and *P. taeda* and 1.17 × 10^{-10} an underestimate of the mutation rate. Bouillé and Bousquet (2005), considering three nuclear genes (amounting to a total of ∼2000 bp), obtained a mutation rate of 2.23 × 10^{-10} to 3.32 × 10^{-10} in Northern American Picea species. As nucleotide diversity varies a lot across loci this estimate cannot be taken at face value but is, in any case, two- to threefold higher than the one reported by Brown *et al.* (2004). Second, estimates of molecular divergence between Pinus and Picea based on an extensive EST database lead to an estimate of the mutation rate of ∼1 × 10^{-9}/year if the divergence time between pine and spruce is that of the diversification of the Pinaceae in the early Cretaceous (∼120–140 MYA) (Savolainen and Wright 2004). Finally, Willyard *et al.* (2006) used divergence at multiple nuclear and chloroplast loci, exemplar taxa, and two calibration points to show that divergence times among pine lineages have often been overestimated and, consequently, absolute mutation rates have been underestimated. They obtain a nuclear silent mutation rate in Pinus of 0.70–1.31 × 10^{-9} sites/year. Hence, the particularly low levels of nucleotide diversity in Norway spruce are probably not exclusively due to low mutation rates and we may have to turn to population demographic history for additional explanations.

Population history: The main outline of Norway spruce population history inferred from this DNA polymorphism survey is the following. As shown previously with other molecular markers [allozymes (Lagercrantz and Ryman 1990), AFLP (Acheré *et al.* 2005), and cytoplasmic DNA (Vendramin *et al.* 2000; Sperisen *et al.* 2001)], the Norway spruce population is today genetically and geographically divided into two main domains, namely the Baltico–Nordic domain and the Alpine Central European domain, and a more limited one, the Carpathian domain, represented in the present survey by a single population, Romania. This population had a very limited polymorphism and may not be representative of the Carpathian domain. The estimate of overall population differentiation that we obtained using F_{ST} (0.117) is substantially higher than that previously reported by Lagercrantz and Ryman (1990) using isozymes (0.052) on a larger set of 70 populations covering a similar geographic range and by Acheré *et al.* (2005) using AFLPs and SSRs (0.02) on a more limited set of populations. This could, in part, be due to differences in sampling and differences in levels of variation among different types of markers (Charlesworth 1998). The choice of candidate genes putatively involved in controlling phenology-related traits for which ample variation exists among these populations ($Q_{ST} = 0.729$, R. Liesch and M. Lascoux, unpublished data) could also explain this difference, if they were under selection. However, no significant difference between candidates and control genes is visible in population differentiation levels as estimated by F_{ST} (data not shown).

The split among the two main geographic domains has been dated to a maximum of 40,000 years (Lagercrantz and Ryman 1990), coinciding with the time estimated from pollen analysis. Previous to that, our analysis suggests that the whole population went through a rather severe bottleneck. We did not attempt to date the bottleneck precisely but recent bottlenecks failed to generate negative values for both *H* and *D* and too severe ones would require unrealistically high values of θ in the ancestral population to explain the data. A more recent bottleneck may also be incompatible with the low level of linkage disequilibrium. Because of the fairly large set of bottleneck parameters that are compatible with the data it is difficult to associate the bottleneck with a particular climatic event. However, climate reconstructions extending back 400,000 years (*e.g.*, Petit *et al.* 1999) show that the average temperature fluctuated with an amplitude of ~10° and a periodicity of ~100,000 years. The bottleneck(s) suggested by the genetic data could then correspond to one of the abrupt changes in temperature that took place during the quaternary. More complex demographic models, metapopulation models, or glacial cycles models, for instance (Wakeley and Aliacar 2001; Jesus *et al.* 2006), may provide an even better fit to our data, but would be more difficult to justify and model at that stage. Finally, our inference was based on both coding and noncoding DNA. It would certainly have been better to use only noncoding DNA as was done by Haddrill *et al.* (2005). However, as there was no strong evidence of selection at loci considered individually, and given the low level of linkage disequilibrium ruling out strong hitchhiking effects, we feel that this may not have altered our conclusion. In summary, we therefore conclude that, even if demography alone is unlikely to explain the low nucleotide variation in all coniferous species, it provides a simple explanation, at least in Norway spruce.

Linkage disequilbrium: Linkage disequilibrium was limited within genic regions; LD decayed by half within <100 bp, confirming earlier results of Rafalski and Morgante (2004). However, the present analysis of LD is biased by the pattern in *col1*, the only gene for which a fragment >3000 bp was sequenced, and indeed estimates of LD at two other long fragments (our unpublished data) were somewhat higher. The pattern of LD was only weakly influenced by population structure, since similar results were obtained when sequences from Romania, the most divergent population, were not included. The rapid decay of LD, consistent with the prevailing outcrossing mating system and the high level of heterozygosity of this species, was similar to the one observed in another outcrossing plant, maize (Tenaillon *et al.* 2001). The very low level of LD could also provide an explanation for the differences in variability at allozyme and nucleotide levels: a limited number of segregating sites per locus recombining freely can lead to a high haplotype diversity.

Conclusions: The level of population structure detected in this study and the overall departure from the standard neutral model of spruce populations imply that these factors will have to be taken into account when carrying out association-mapping studies (Marchini *et al.* 2004; Campbell *et al.* 2005; Helgason *et al.* 2005) and when interrogating SNP databases for signatures of natural selection (Akey *et al.* 2002), respectively. The rapid decay of LD in spruce will allow high-resolution mapping in association studies, given that the right candidate genes are chosen, but will also require a high-density marker screening due to the limited predictive power of single SNPs over neighbor sequence diversity. Finally, if confirmed by more extensive studies, the rapid decay of LD also implies that hitchhiking is likely to have played a limited role in the species evolution.

We thank Skogforsk, Randolph Schirmer, Felix Gugerli, Magdalena Palada, Vladimir Semerikov, and Giovanni Vendramin for providing *Picea abies* seeds. M.L. thanks Peter Andolfatto for kindly providing the code used to carry out multilocus tests of neutrality and answering questions about it and Erik Lagercrantz for writing a Ruby script that greatly facilitated these analyses. We are grateful to two anonymous referees for constructive comments on the manuscript. This research was funded by the European Commission, project TREESNIPS (QLRT -2001-01973), the Carl Tryggers Foundation, the Philip-Sörensen Foundation, and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning. M.H. acknowledges a "Mobility of Researchers" grant from the National Research Fund of Luxembourg and is currently a postdoctoral researcher at the National Fund for Scientific Research of Belgium.

LITERATURE CITED

Acheré, V., J. M. Favre, G. Besnard and S. Jeandroz, 2005 Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). Mol. Ecol. **14:** 3191–3201.

Akey, J. M., G. Zhang, K. Zhang, L. Jin and M. D. Shriver, 2002 Interrogating a high-density SNP map for signatures of natural selection. Genome Res. **12:** 1805–1814.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Beaumont, M. A., and D. J. Balding, 2004 Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. **13:** 969–980.

Bennett, K. D., 1997 *Evolution and Ecology. The Pace of Life.* Cambridge University Press, Cambridge, UK.

Bolle, C., C. Koncz and N. H. Chua, 2000 PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. Genes Dev. **14:** 1269–1278.

Bouillé, M., and J. Bousquet, 2005 Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. Am. J. Bot. **92:** 63–73.

Brown, G. R., G. P. Gill, R. J. Kuntz, C. H. Langley and D. B. Neale, 2004 Nucleotide variation and linkage disequilibrium in loblolly pine. Proc. Natl. Acad. Sci. USA **101:** 15255–15260.

Bucci, G., and G. G. Vendramin, 2000 Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. Mol. Ecol. **9:** 923–934.

Campbell, C. D., E. L. Ogburn, K. L. Lunetta, H. N. Lyon, M. L. Freedman *et al.*, 2005 Demonstrating stratification in a European American population. Nat. Genet. **37:** 868–872.

Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. **15:** 538–543.

Doyle, J., and J. Doyle, 1990 Isolation of plant DNA from fresh tissue. BRL Focus **12:** 13–15.

Dvornyk, V., A. Sirviö, M. Mikkonen and O. Savolainen, 2002 Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris.* Mol. Biol. Evol. **19:** 179–188.

Ekberg, I., G. Eriksson and I. Dormling, 1979 Photoperiodic reactions in conifer species. Holarct. Ecol. **2:** 255–263.

Ewing, B., and P. Green, 1998 Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:** 186–194.

Ewing, B., L. Hillier, M. Wendl and P. Green, 1998 Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:** 175–185.

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.

Falush, D., M. Stephens and J. K. Pritchard, 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164:** 1567–1587.

Fay, J. C., and C.-I Wu, 1999 A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol. Biol. Evol. **16:** 1003–1005.

Fay, J. C., and C.-I Wu, 2000 Hitchhicking under positive Darwinian selection. Genetics **155:** 1405–1413.

Fowler, S., K. Lee, H. Onouchi, A. Samach, K. Richardson *et al.*, 1999 GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. EMBO J. **18:** 4679–4688.

García-Gil, M. R., M. Mikkonen and O. Savolainen, 2003 Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris.* Mol. Ecol. **12:** 1195–1206.

González-Martínez, S. C., E. Ersoz, G. R. Brown, N. C. Wheeler and D. B. Neale, 2006 DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. Genetics **172:** 1915–1926.

Gordon, D., C. Abajian and P. Green, 1998 Consed: a graphical tool for sequence finishing. Genome Res. **8:** 195–202.

Haddrill, P. R., K. R. Thornton, B. Charlesworth and P. Andolfatto, 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Res. **15:** 790–799.

Hamrick, J. L., and M. J. W. Godt, 1996 Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **351:** 1291–1298.

Hayama, R., and G. Coupland, 2004 The molecular basis of diversity in the photoperiodic flowering responses of Arabidopsis and rice. Plant Physiol. **135:** 677–684.

Helgason, A., B. Yngvadóttir, B. Hrafnkelsson, J. Gulcher and K. Stefánsson, 2005 An Icelandic example of the impact of population structure on association studies. Nat. Genet. **37:** 90–95.

Hill, W. G., and B. S. Weir, 1988 Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. **33:** 54–78.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159:** 1805–1817.

Ingvarsson, P. K., 2005 Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). Genetics **169:** 945–953.

Jesus, F. F., J. F. Wilkins, V. N. Solferini and J. Wakeley, 2006 Expected coalescence times and segregating sites in a model of glacial cycles. Genet. Mol. Res. **5:** 466–474.

Kado, T., H. Yoshimaru, Y. Tsumura and H. Tachida, 2003 DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). Genetics **164:** 1547–1559.

Lagercrantz, U., and N. Ryman, 1990 Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. Evolution **44:** 38–53.

Lin, C., M. Ahmad and A. R. Cashmore, 1996 Arabidopsis cryptochrome 1 is a soluble protein mediating blue light-dependent regulation of plant growth and development. Plant J. **10:** 893–902.

Marchini, J., L. R. Cardon, M. S. Phillips and P. Donnelly, 2004 The effects of human population structure on large genetic association studies. Nat. Genet. **36:** 512–517.

McVean, G., P. Awadalla and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:** 1231–1241.

Myers, S., L. Bottolo, C. Freeman, G. McVean and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science **310:** 321–324.

Neale, D. B., and O. Savolainen, 2004 Association genetics of complex traits in conifers. Trends Plant Sci. **9:** 325–330.

Petit, J. R., J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola *et al.*, 1999 Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature **399:** 429–436.

Piñero, M., C. Gómez-Mena, R. Schaffer, J. M. Martínez-Zapater and G. Coupland, 2003 EARLY BOLTING IN SHORT is related to chromatin remodelling factors and regulates flowering in Arabidopsis by repressing FT. Plant Cell **15:** 1552–1562.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics **155:** 945–959.

Przeworski, M., 2002 The signature of positive selection at randomly chosen loci. Genetics **160:** 1179–1189.

Putterill, J., F. Robson, K. Lee, R. Simon and G. Coupland, 1997 The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell **80:** 847–857.

Rafalski, A., and M. Morgante, 2004 Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet. **20:** 103–111.

R Development Core Team, 2005 *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

Remington, D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA **98:** 11479–11484.

Rozas, J., J. C. Sánchez-DelBarria, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **29:** 2496–2497.

Rozen, S., and H. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 365–386 in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, edited by S. Krawetz and S. Misener. Humana Press, Totowa, NJ.

Savolainen, O., and M. Wright, 2004 Estimating divergence rates of conifers based on EST sequences conifer EST sequences, p. 7 in Population, Evolutionary and Ecological Genomics of Forest Trees. IUFRO Sections Population Genetics and Genomics, Pacific Grove, CA, September 13–17, 2004.

Schaffner, S. F., C. Foo, S. Gabriel, D. Reich, M. J. Daly *et al.*, 2005 Calibrating a coalescent simulation of human genome sequence variation. Genome Res. **15:** 1576–1583.

Schmid, K. J., S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar and T. Mitchell-Olds, 2005 A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. Genetics **169:** 1601–1615.

Schmidt, M., and H. A. Schneider-Poetsch, 2002 The evolution of gymnosperms redrawn by phytochrome genes: the Gnetatae appear at the base of the gymnosperms. J. Mol. Evol. **54:** 715–724.

Schneider, S., D. Roessli and L. Excoffier, 2000 *Arlequin Ver. 2.000: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva.

Setakis, E., H. Stirnadel and D. J. Balding, 2006 Logistic regression protects against population structure in genetic association studies. Genome Res. **16:** 290–296.

Sharrock, R. A., and P. H. Quail, 1989 Novel phytochrome sequences in Arabidopsis thaliana: structure, evolution, and differential expression of a plant regulatory photoreceptor family. Genes Dev. **3:** 1745–1757.

Simpson, G. G., and C. Dean, 2002 Arabidopsis, the Rosetta Stone of flowering time? Science **296:** 285–289.

Sperisen, C., U. Büchler, F. Gugerli, G. Mátyás, T. Geburek *et al.*, 2001 Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. Mol. Ecol. **10:** 257–263.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA **98:** 9161–9166.

Thornton, K., and P. Andolfatto, 2005 Approximate Bayesian inference reveals evidence for a recent, severe, bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics **172:** 1607–1619.

Tiffin, P., and B. S. Gaut, 2001 Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. Genetics **158:** 401–412.

Vendramin, G. G., M. Anzidei, A. Madaghiele, C. Sperisen and G. Bucci, 2000 Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce. Genome **43:** 68–78.

Voight, B. F., A. M. Adams, L. A. Frisse, Y. Qian, R. R. Hudson *et al.*, 2005 Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc. Natl. Acad. Sci. USA **102:** 18508–18513.

Wakeley, J., and N. Aliacar, 2001 Gene genealogies in a metapopulation. Genetics **159:** 893–905.

Wang, X.-Q., D. C. Tank and T. Sang, 2000 Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol. Biol. Evol. **17:** 773–781.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

Willyard, A., J. Syring, D. S. Gernandt, A. Liston and R. Cronn, 2006 Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiation for *Pinus*. Mol. Biol. Evol. (in press).

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Wright, S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley *et al.*, 2005 The effects of artificial selection of the maize genome. Science **308:** 1310–1314.

Yanovsky, M. J., and S. A. Kay, 2003 Living by the calendar: how plants know when to flower. Nat. Rev. Mol. Cell Biol. **4:** 265–275.

Zhang, H., C. Ransom, P. Ludwig and S. van Nocker, 2003 Genetic analysis of early flowering mutants in Arabidopsis defines a class of pleiotropic developmental regulator required for expression of the flowering time-switch Flowering Locus C. Genetics **164:** 347–358.

Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. van Tasseli, L. K. Matukumalli *et al.*, 2003 Single-nucleotide polymorphisms in soybean. Genetics **163:** 1123–1134.

Communicating editor: M. Nordborg