Molecular Ecology (2005) **14**, 3191–3201 doi: 10.1111/j.1365-294X.2005.02646.x

Genomic organization of molecular differentiation in Norway spruce (*Picea abies*)

V. ACHERÉ,* J. M. FAVRE,* G. BESNARD*† and S. JEANDROZ*
*UMR INRA/UHP 1136: Tree-microbe Interactions, Faculté des Sciences, Université Henri Poincaré — Nancy 1 BP 239, F-54506 Vandoeuvre-lès-Nancy, France, †UNIL, Department of Ecology and Evolution, Biology Building, CH-1015 Lausanne, Switzerland

Abstract

Diversity and differentiation among three populations representing the geographical domains commonly recognized within the natural distribution area of *Picea abies* were analysed by using a set of 292 AFLP (amplified fragment length polymorphism), SSR (single sequence repeat) and ESTP (expressed sequence tags polymorphism) markers. As usually observed in forest trees, results showed high within-population diversity (H_S reaching 0.79) and low among-population differentiation ($G_{ST} \approx 2\%$). The genomic organization of differentiation was then investigated on the basis of a subsample of 150 AFLP, SSR and ESTP mapped markers. The number of the loci differentiating the Baltico-Nordic from the central European populations (25 loci) and, within the central European populations, the Alpine from the Hercyno-Carpathian populations (12 loci), were different. These 37 differentiated loci, with individual G_{ST} values ranging from 0.008 to 0.20, were evenly distributed on all linkage groups and mostly followed the neutral expectations, suggesting genome-wide effects on differentiation. Nine of them however behave as 'outlier' loci indicating possible locus-specific selective effects. Contribution of ongoing evolutionary forces and historical effects to the geographical differentiation of the species are discussed.

Keywords: intraspecific differentiation, mapped markers, natural selection, outlier loci, *Picea abies*

Received 30 January 2005; revision received 15 April 2005; accepted 13 May 2005

Introduction

In general, forest trees are highly polymorphic plant species and, in contrast with most herbaceous species, they typically display low among population differentiation (Hamrick *et al*. 1992; Müller-Starck *et al*. 1992; Hamrick 2004). These distinctive characteristics are certainly connected to specific biological traits of these species, such as long generation time, predominant outcrossing and high level of long-distance gene flow via pollen (Hamrick *et al*. 1992; Hamrick & Godt 1996).

As in other species, genetic variation in forest trees is strongly influenced by ongoing evolutionary forces such as mutation, genetic drift, gene flow and natural selection. But in such species with long generation time, historical events also play a crucial role. Indeed, the number, geographical location and size of glacial refugia, the routes of postglacial expansion and the possible occurrence of secondary contacts between populations migrated from different refuge sites, have been shown to have important consequences on the genetic structure of many forest tree species (Taberlet *et al*. 1998; Austerlitz *et al*. 2000; Petit *et al*. 2003; Jaramillo-Correa *et al*. 2004). Furthermore, the dramatic intensification of human activities during the last centuries has increasingly impacted the forest biodiversity, even though documented studies are often missing (Ledig 1992).

Picea abies (L.) Karst. is one of the most important coniferous species of Europe from both ecological and economical points of view. It covers a wide distribution range, stretching from western Europe to the Ural Mountains and from Fennoscandia to Greece. Based on palynological, geographical and morphological traits, this wide natural range is commonly divided into three domains generally considered as resulting from postglacial expansion from three putative refuges (Huntley & Birks 1983; Schmidt-Vogt 1986), i.e. (i) the Baltico-Nordic domain deriving from a refuge putatively located in the Moscow area, (ii) the Hercyno-Carpathian domain issued from a refuge localized in the Carpathian mountains and (iii) the Alpine domain colonized from a refuge situated in the Dinaric Alps.

Correspondence: Sylvain Jeandroz. Fax: (33)383903277; E-mail: sylvain.jeandroz@scbiol.uhp-nancy.fr

Genetic variation in this species has been investigated by using nuclear, mitochondrial and chloroplast DNA markers (Lagercrantz & Ryman 1990; Bucci & Vendramin 2000; Vendramin *et al*. 2000; Sperisen *et al*. 2001; Collignon *et al*. 2002; Bastien *et al*. 2003). As in most forest tree species, nuclear markers revealed high within-population diversity and low population differentiation (Lagercrantz & Ryman 1990; Collignon *et al*. 2002). Cytoplasmic markers, either paternally (chloroplast markers) or maternally (mitochondrial markers) inherited, were more informative and allowed to clearly distinguish two groups of populations, namely a northern group consisting of the populations of the Baltico-Nordic domain and a central European group including the populations of the Alpine and Hercyno-Carpathian domains (Vendramin *et al*. 2000; Sperisen *et al*. 2001). The random amplified polymorphic DNA (RAPD) analysis carried out by Collignon *et al*. (2002) corroborated this dual genetic structure: 33% of scored markers showed significant differentiation between the northern and central European provenances. These results indicate early divergence of the Baltico-Nordic populations and argue in favour of a reproductive isolation during the glacial times that was afterwards partially maintained during the postglacial expansion.

Differentiation analyses within the central European group did not lead to such straightforward conclusions (Lagercrantz & Ryman 1990; Bucci & Vendramin 2000; Collignon *et al*. 2002). For instance, polymorphism analysis at one mitochondrial minisatellite locus (*mh* 44) did not reveal a consistent relationship between the genetic structure and the geographical origin of the populations (Bastien *et al*. 2003). Two major haplotypes were detected, but they were found in populations from both the Alpine and Hercyno-Carpathian domains, although a preferential haplotype-domain association could be observed. The UPGMA clustering based on RAPD data did not better separate the populations from the Alpine and Hercyno-Carpathian domains than mtDNA analysis. The dendrogram showed an intricate pattern, even though small distinct clusters consistent with geographical origin could be discerned (Collignon *et al*. 2002). Higher differentiation however was revealed when cluster analysis on quantitative traits was performed, suggesting that adaptation to local environment had occurred (Collignon *et al*. 2002).

It can thus be suspected that the differentiation patterns observed between the northern and central European groups of populations and, within the central European group, between the populations of the Hercyno-Carpathian and Alpine domains, result from different combinations of evolutionary forces and historical effects.

In molecular studies of genetic diversity, population parameters are usually estimated from the polymorphism observed at putatively neutral marker loci. The data generally reveal a great variation in the level of differentiation among loci. If we focus on coniferous species, Yang *et al*. (1996) and Tani *et al*. (2003) reported G_{ST} values at isozyme loci ranging from 0.0001 to 0.06 and from 0.006 to 0.12 in *Pinus contorta* and *Pinus parviflora*, respectively. The corresponding values of mean differentiation were 0.019 and 0.044. In *Picea mariana*, Isabel *et al*. (1995), recorded individual G_{ST} values at RAPD loci comprised between 0.0 and 0.13, for a mean differentiation value of 0.03. These examples clearly indicate that some loci can behave very differently from the rest of the genome.

The distribution across the genome of these 'outlier' loci (Luikart *et al*. 2003) can be of much interest to distinguish between locus-specific and genome-wide effects. Indeed, under a standard neutral model, genetic drift and migration should affect all loci, while mutation and natural selection are expected to differ among loci. Outlier loci can thus be considered as genomic signatures of natural selection. The relevance of this approach has been convincingly illustrated in human genetics (Akey *et al*. 2002; Kayser *et al*. 2003). Akey *et al*. (2002) in particular, calculated individual F_{ST} values and examined the distribution at genome, chromosome and individual gene level of more than 26 000 single nucleotide polymorphism (SNP) loci. From this analysis, they were able to identify 174 candidate genes subject to selection that provided a frame for a first natural selection map of the human genome. This approach could help investigate the mechanisms underlying genetic variation.

In forest trees, such an approach has been used to study the interspecific differentiation between two sympatric and interfertile European white oaks (*Quercus robur* and *Quercus petraea*). Twenty outlier loci distributed over nine linkage groups (the haploid number of chromosomes is 12 in both species) were identified, indicating genome-wide selective effects in species divergence (Scotti-Saintagne *et al*. 2004).

In this study, we applied a similar approach to analyse the geographical differentiation observed within the natural distribution area of *Picea abies*. We first carried out a multilocus analysis of the genetic variation using AFLP (amplified fragment length polymorphism), SSR (single sequence repeat = microsatellite) and ESTP (expressed sequence tags polymorphism) mapped and unmapped markers. Then, in order to investigate the genomic organization of molecular variation, we considered the differentiation at each mapped locus individually. Tests of selective neutrality were performed and the distribution across the genome of the loci significantly involved in the genetic variation was examined using the saturated genetic map recently published by Acheré *et al*. (2004).

Materials and methods

Plant material and DNA isolation

Analyses were performed on a sample of 92 trees originating from the three geographical domains usually recognized

within the natural distribution range of *Picea abies*, with 30 trees for the Baltico-Nordic domain (D1), 33 for the Hercyno-Carpathian domain (D2) and 29 for the Alpine domain (D3). These trees were collected in the international IUFRO provenance test 1964/1968 (Krutzsch 1974) located in the Amance forest (northeastern France). Each tree represented a distinct provenance. The same sample was previously used for RAPD (Collignon *et al*. 2002) and mtDNA (Bastien *et al*. 2003) studies of genetic variation in this species. Total DNA was isolated from frozen needles, using the DNeasy Plant Mini Kit (QIAGEN), according to the manufacturer's recommendations.

Choice of mapped AFLP, microsatellite and ESTP markers

To analyse genetic diversity, we selected AFLP, SSR and ESTP markers evenly distributed over the 12 linkage groups of the consensus map of *P. abies* published by Acheré *et al*. (2004). Distribution of markers was analysed by comparing the observed number of markers per 5-cM interval with the expected Poisson distribution frequencies using a chi-squared test as described in Acheré *et al*. (2004).

Regarding AFLP markers, we chose 15 highly polymorphic primer-enzyme combinations (PECs) which generated a total of 345 segregating markers in the mapping progeny of Acheré *et al*. (2004). These PECs were named a1, a2, a3, a4, a7, a8, a9, a10, a13, a22, a24, a41, a43, a44 and a45, as in the original publication by Acheré *et al*. (2004). To identify homologous fragments generated by a given PEC, amplification products of the parents and six individuals representing the mapping progeny were loaded on each polyacrylamide gel together with those of the D1, D2 and D3 tree samples. It has been well established that most comigrating AFLP fragments from individuals of the same species are homologous (Rouppe van der Voort *et al*. 1997; Waugh *et al*. 1997; Virk *et al*. 2000). The AFLP markers identified were then designated by the name of the PEC followed by the size of fragment in base pairs, e.g. a1-186.

For microsatellites we selected 25 primer pairs developed by Pfeiffer *et al*. (1997), Hodgetts *et al*. (2001), Scotti *et al*. (2000, 2002a, b) and Besnard *et al*. (2003) from different *Picea* species, and for ESTPs two primer pairs designed by Schubert *et al*. (2001) from *P. abies* cDNA sequences. We checked that all these SSR and ESTP markers were single locus, codominant and multiallelic. Experimental protocols were as described in Acheré *et al*. (2004).

Data analysis

Sampled trees from each of the D1, D2 and D3 domains were considered as populations. Data obtained with the dominant (AFLP) and codominant (SSR and ESTP) markers were analysed separately.

The polymorphic AFLP fragments were scored independently as present (1) or absent (0) and a binary matrix constructed. This matrix was first used to carry out a phenotypic analysis of differentiation. The genetic structure within the 92 sampled trees was depicted using a correspondence analysis (CA) based on the chi-squared distances between paired individuals (NTSYS-PC 2.02; Rohlf 2000) and the differentiation among populations (Φ_{ST}) estimated by an analysis of molecular variance (AMOVA) carried out on a matrix of squared standard Euclidean distances between all pairs of AFLP phenotypes (Excoffier *et al*. 1992). Significance was obtained by using 500 permutations.

A genotypic analysis was then performed using those AFLP markers with frequency lower than $1 - (3/N)$, where N is the sample size, as recommended by Lynch & Milligan (1994) to reduce bias in estimating the population genetics parameters (Isabel *et al*. 1995, 1999; Szmidt *et al*. 1996; Krutovskii *et al*. 1999). Allelic frequencies were calculated assuming Mendelian segregation and Hardy–Weinberg equilibrium. The mean diversity within population (H_S), the total diversity (H_T) and the differentiation at each locus (G_{ST}: Nei 1973) were computed with the POPGENE version 1.31 software (Yeh *et al*. 1997). The significance of differentiation was estimated by using the SPAGEDI version 1.1 software (Hardy & Vekemans 2002). The procedure is based on the comparison of observed values of differentiation (F_{ST}) with the corresponding frequency distributions obtained by means of 10 000 random permutations of individuals among populations (95% confidence interval).

Regarding the codominant SSR and ESTP markers, fragment size was estimated with the GENEIMAGIR software. The allele distribution among populations was compared for each polymorphic locus with Fisher exact test, using GENEPOP version 3.3 software (Raymond & Rousset 1995). Diversity (H_S, H_T) and differentiation (G_{ST}) estimates were calculated from allele frequencies by using the FSTAT version 2.9.3 5 software (Goudet 2002). Significance of differentiation at each locus was tested by a log-likelihood (G) based exact test (Goudet *et al*. 1996) computed with GENEPOP.

Identification of AFLP, SSR and ESTP outlier loci was carried out following the Beaumont and Nichols' approach (1996) that uses computer simulation to model neutral loci with the FDIST2 software. An infinite alleles mutational model was used. The simulated data was used to derive F_{ST} using Cockerham & Weir's formula (1987). The distribution of F_{ST} as a function of heterozygosity was characterized by estimating the 0.05, 0.50 and 0.95 quantiles of the distribution. In a first step, the neutral expectation was based on the overall mean value of F_{ST} calculated from all markers. Markers with F_{ST} values outside the 0.95 limits corresponding to the null hypothesis were then removed and a new analysis was performed with a recalculated mean value of F_{ST}. Markers with F_{ST} values that fell outside the 0.95 limits after this second analysis were considered as outlier loci.

Comparison of H_S, H_T and G_{ST} obtained with dominant and codominant markers was performed with the non-parametric Mann–Whitney rank sum test applied to single-locus estimates of these parameters. The G_{ST} distributions for the dominant and codominant markers were compared with Fisher exact test after grouping of G_{ST} values into 10 classes with an interval of 0.01.

Results

Genetic diversity

AFLP markers. A total of 265 polymorphic were scored from the 15 selected PECs. The number of fragments polymorphic in at least one population ranged from 4 to 39 per PEC (17.7 per PEC on average). Of these 265 AFLP markers, 143 were homologous to mapped loci of the saturated genetic map of Acheré *et al*. (2004) based on similarity of fragment size. The genome position of the remaining 122 was unknown (unmapped markers).

When the Lynch and Milligan's restrictions (1994) for dominant markers were applied, the number of polymorphic AFLP markers was reduced to 203. The mean diversity within populations (H_S) and the total diversity (H_T) amounted to 0.378 and 0.389, respectively. The genetic diversity estimates (H_S or H_T) did not differ significantly whether the mapped (117) or unmapped (86) markers were considered (Mann–Whitney rank sum test; $P > 0.05$).

SSR and ESTP markers. All the analysed SSRs and ESTPs were found to be polymorphic in the D1, D2 and D3 populations. The 25 microsatellites revealed a high number of alleles, ranging from 5 to 55 per locus (25 on average). For both ESTPs, the number of alleles was similar to that reported by Schubert *et al*. (2001), e.g. 5 alleles of 212–230 bp and 4 alleles of 368–420 bp for PA0034 and PA0043, respectively.

Microsatellite markers revealed high genetic diversity, with H_S and H_T values of 0.789 and 0.796, respectively (Table 1). With the ESTP markers, genetic diversity was lower ($H_S = 0.493$ and $H_T = 0.505$; Table 1). Altogether these values were significantly higher than those obtained with the dominant AFLP markers (Mann–Whitney rank sum test; $P < 0.001$).

Genetic differentiation

Results of the correspondence analysis (CA) carried out at the whole range level (92 trees from the three geographical domains) using the frequencies of the 265 AFLP markers are shown in Fig. 1. The two first axes accounted for 7.3% of the total variation. A clear separation of the Baltico-Nordic trees (D1) was observed along axis 1. Trees of the central European populations (D2, D3) were mixed together, forming a distinct composite group.

Table 1 Genetic diversity and differentiation between the D1, D2 and D3 populations at 25 SSR and 2 ESTP loci

Marker	Loci	References	H_S	H_T	G_{ST}	Lg
SSR	UAPgAG150(A)	[5]	0.799	0.807	0.011	1
	EAC6E09	[3]	0.920	0.928	0.008	1
	SpAGC2	[1]	0.897	0.917	0.022*	2
	PAAC3	[2]	0.949	0.953	0.004	2
	EAC6B01	[3]	0.940	0.947	0.007	3
	UAPsTG25	[5]	0.544	0.576	0.055*	3
	EAC7H7	[3]	0.934	0.938	0.004	5
	SpAC1F7	[1]	0.573	0.579	0.010	5
	SpAC1H8	[1]	0.935	0.945	0.010	5
	SpAGC1	[1]	0.682	0.684	0.002	5
	EAC1D10	[3]	0.980	0.982	0.002	5
	SpAC1B8	[1]	0.984	0.983	–0.001	6
	SpAG2	[1]	0.913	0.918	0.005	6
	EATC2C01	[4]	0.943	0.945	0.002	7
	pgGB5	[6]	0.755	0.759	0.005	7
	EAC1G5	[3]	0.977	0.981	0.004	8
	EATC1D02A	[4]	0.795	0.804	0.011*	9
	paGB8	[6]	0.960	0.965	0.006	10
	pgGB7	[6]	0.569	0.584	0.027*	10
	UAPgCA91	[5]	0.406	0.407	0.003	11
	EATC1B2	[4]	0.303	0.303	–0.001	11
	EATC1E3	[4]	0.317	0.327	0.029*	11
	UAPgCA24	[5]	0.965	0.972	0.008	11
	paGB3	[6]	0.766	0.767	0.001	12
	PAAC19	[2]	0.925	0.929	0.004	12
	Overall		**0.789**	**0.796**	**0.009***	
ESTP	PA0043	[7]	0.634	0.657	0.036*	5
	PA0034	[7]	0.352	0.353	0.005	9
	Overall		**0.493**	**0.505**	**0.025***	

[1], Pfeiffer *et al*. 1997; [2], Scotti *et al*. 2000; [3], Scotti *et al*. 2002a; [4], Scotti *et al*. 2002b; [5], Hodgetts *et al*. 2001; [6], Besnard *et al*. 2003; [7], Schubert *et al*. 2001; H_S, mean within-population diversity; H_T, total gene diversity; G_{ST}, genetic differentiation; *significant differentiation at 5% level; Lg, linkage group number (Acheré *et al*. 2004).

The analysis of molecular variance (AMOVA) based on the whole 265 AFLP markers, showed that 95.6% of the molecular variation was found within populations and 4.3% among populations. The two-level variance partitioning indicated a higher differentiation ($\Phi_{ST} = 5.2\%$; $P < 0.002$) between the Baltico-Nordic (D1) and central European (D2 + D3) groups than between the D1, D2, D3 ($\Phi_{ST} = 4.3\%$; $P < 0.002$) or the D2, D3 ($\Phi_{ST} = 1.9\%$, $P < 0.002$) populations.

The differentiation between the three populations, as measured by the SSR and ESTP codominant markers or by the 203 AFLP dominant markers in conformity with the Lynch and Milligan's restrictions (1994), was significant, with overall G_{ST} values of 0.009, 0.025 (Table 1) and 0.029, respectively. The frequency distribution of individual G_{ST} values of the AFLP markers (Fig. 2) differed significantly from that of the microsatellites (Fisher test; $P = 0.0008$).

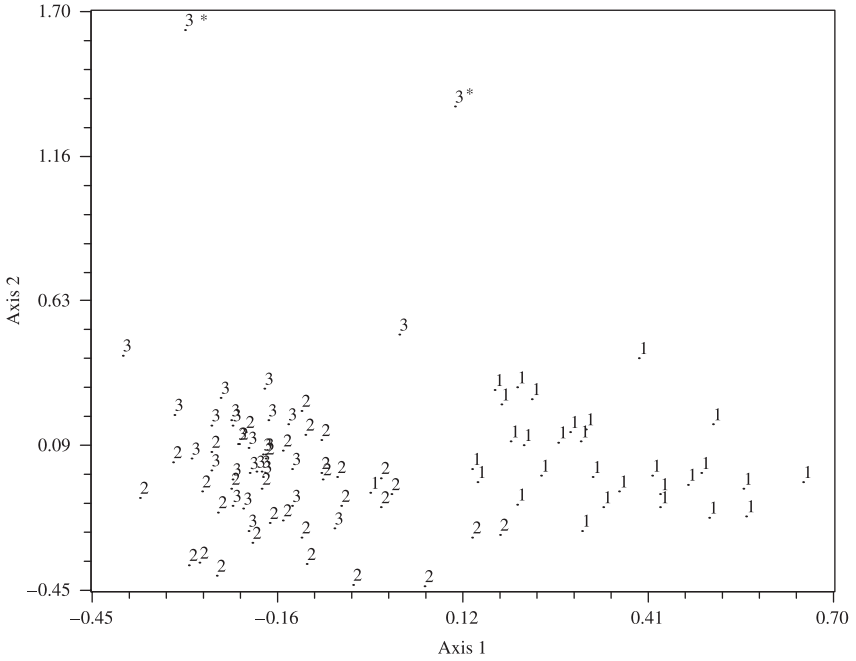

Fig. 1 The first two axes of the correspondence analysis (CA) computed from the AFLP data: distribution of the 92 trees according to their original domain (1, Baltico-Nordic domain; 2, Hercyno-Carpathian domain; 3, Alpine domain). Trees labelled by one asterisk showed several missing data. Their origin (domain 3) has been confirmed by mtDNA data (Bastien *et al*. 2003).

Two-thirds of AFLPs displayed G_{ST} values > 1%, compared to only one-third for SSRs. Five AFLPs (2.46%) showed among population differentiation exceeding 10%. Comparison of G_{ST} calculated from mapped or unmapped AFLP markers did not reveal significant difference (Mann–Whitney rank sum test, $P > 0.05$).

Genomic distribution of differentiation

Given the dual structure of genetic variation within the natural distribution range of the species confirmed by the data reported in the previous paragraph, we first examined the genomic distribution of differentiation between the D1 and D2 + D3 pooled populations. One hundred and fifty-two mapped markers (125 AFLPs, 25 SSRs and 2 ESTPs; Fig. 3), representing about 20% of the total markers positioned on the saturated linkage map of Acheré *et al*. (2004), were found to be polymorphic. Comparison of the observed and expected Poisson distributions indicated that these polymorphic markers are randomly distributed over the 12 linkage groups ($\chi^2 = 1.30$; 3 d.f.; $P < 0.05$; with 7–18 markers per linkage group) with an average spacing between two adjacent markers of 13.4 cM (Fig. 3).

Differentiation analysis revealed significant overall G_{ST} between the D1 and D2 + D3 populations ($P < 0.05$), with values of 0.018, 0.008 and 0.027 for AFLP, SSR and ESTP markers, respectively. However, only 25 out of the 152 mapped loci (16%) were significantly involved in the differentiation. These 25 differentiated loci (17 AFLPs, 7 SSRs and 1 ESTP) were evenly distributed ($\chi^2 = 0.82$; 2 d.f.; $P < 0.05$) over 11 of the 12 linkage groups of the map (1–4 per linkage group; Fig. 3). No significantly differentiated loci were detected on linkage group 4. On linkage group 2, four differentiated loci (3 AFLPs and 1 SSR) occurred in a 32-cM interval, with in-between distances of 4.6, 10.9 and 16.7 cM (Fig. 3). However the comparison with the expected Poisson distribution indicated that this linkage group did not contain any cluster of markers ($\chi^2 = 0.23$; 2 d.f.; $P < 0.05$).

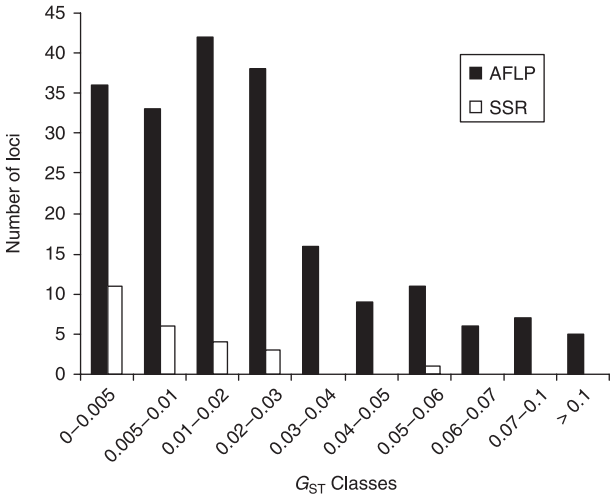

Fig. 2 Distribution of G_{ST} values of the 203 AFLP and 25 SSR analysed markers.

Application of the Beaumont and Nichols' procedure (1996) to the 152 markers allowed us to identify three outlier loci (2 AFLPs and 1 SSR; mean F_{ST} value: 0.026). To avoid bias in the differentiation estimate, the mean F_{ST} value

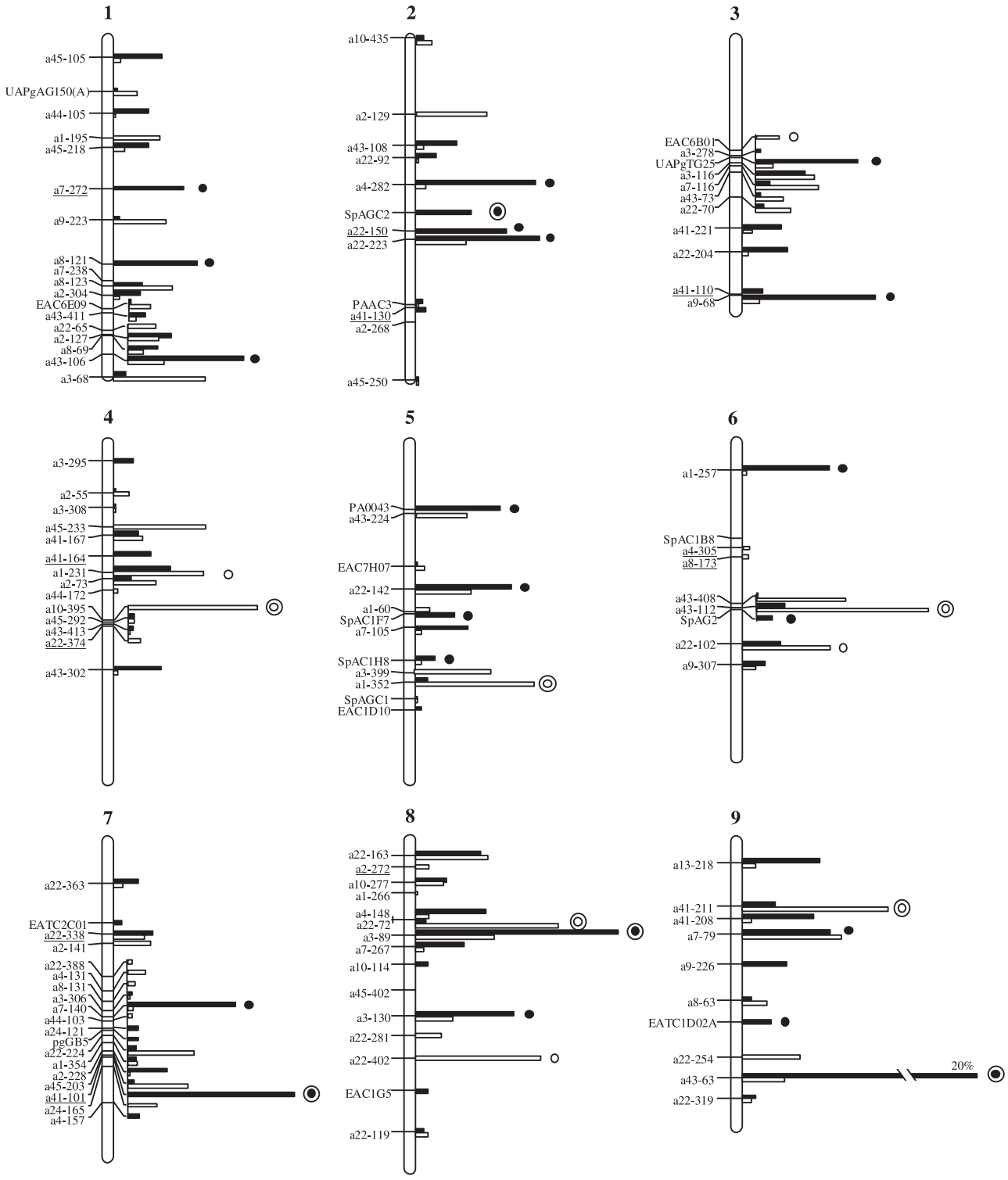

Fig. 3 Location and G_{ST} values of the analysed markers along the 12 linkage groups of the consensus map by Acheré *et al*. (2004).
Black bars correspond to the differentiation between the D1 and D2 + D3 pooled populations.
● Significant differentiation at 5% level.
◉ Outlier loci.
– White bars indicate to the differentiation between the D2 and D3 populations.
○ Significant differentiation at 5% level.
◎ Outlier loci.
Underlined markers were only used in one of the analysed differentiations.

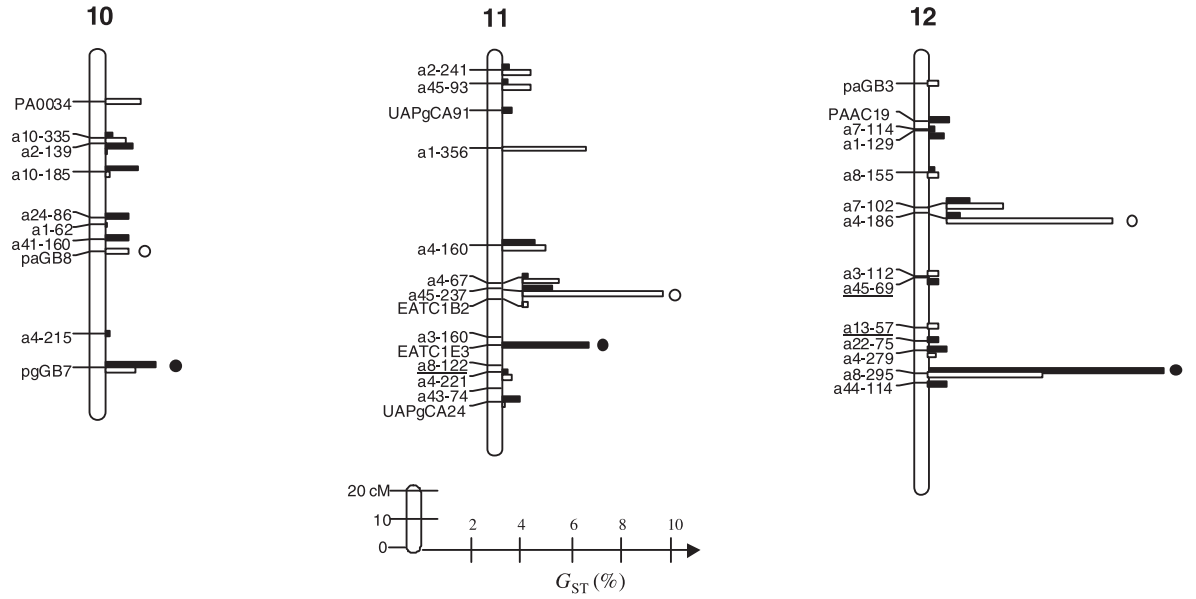

Fig. 3 *Continued*

was recalculated after removal of these three highly divergent loci (recalculated F_{ST} value = 0.020). One additional outlier AFLP locus could then be detected. In total, four loci showing F_{ST} values exceeding the 0.95 limits of the simulated distribution of differentiation under selective neutrality were found (Fig. 4a). These four outliers, putatively affected by natural selection, were all significantly involved in the differentiation between the D1 and D2 + D3 populations, representing 16% of the differentiated loci. They were scattered across four different linkage groups (Lg 2, Lg 7, Lg 8 and Lg 9; Fig. 3).

In a second step of the analysis, we examined the genomic distribution of the differentiation between the D2 and D3 central European populations. One hundred and fifty mapped markers (123 AFLPs, 25 SSRs and 2 ESTPs; Fig. 3) were found to be polymorphic. They were randomly distributed ($\chi^2 = 6.07$; 4 d.f.; $P < 0.05$) over the 12 linkage groups of the map (9–18 per linkage group), with an average spacing between two adjacent markers of 13.6 cM.

Mean differentiation as revealed by AFLPs was significant (overall G_{ST}: 0.017; $P < 0.05$), but not when SSR and ESTP markers were used (overall G_{ST}: 0.002 for both). The number of loci significantly involved in the differentiation decreased to 12 (10 AFLPs and 2 SSRs; Fig. 3), representing 8% of the polymorphic markers. They were evenly scattered ($\chi^2 = 0.195$; 2 d.f.; $P < 0.05$) across nine linkage groups (1–2 per linkage group), with a minimum distance of 19 cM (Fig. 3).

The Beaumont and Nichols' test for neutrality (1996) applied to the 150 polymorphic markers allowed the detection of three outlier AFLPs (mean $F_{ST} = 0.020$). After recalculation of the mean F_{ST} without these three divergent loci (recalculated F_{ST} value = 0.015), two additional outlier

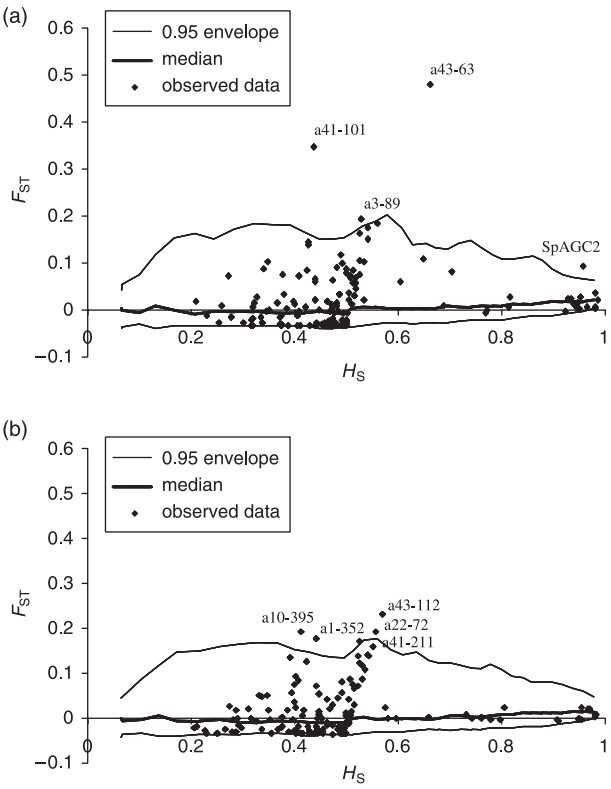

Fig. 4 Distribution of F_{ST} values calculated from mapped loci as a function of heterozygosity (H_S). The quantiles have been estimated following Beaumont & Nichols (1996). (a) Differentiation between the D1 and D2 + D3 populations. (b) Differentiation between the D2 and D3 populations. Outlier loci are referenced as in Fig. 3.

AFLPs were identified. The total number of outlier loci identified was thus of five, all AFLPs (Fig. 4b). These outlier AFLPs were significantly involved in the D2 vs. D3 differentiation, representing 42% of the differentiated loci. They were distributed on five different linkage groups (Lg 4, Lg 5, Lg 6, Lg 8 and Lg 9; Fig. 3) and differed from those involved in the D1 vs. D2 + D3 differentiation.

Discussion

Molecular differentiation

Diversity estimates obtained in this study by using AFLP, SSR and ESTP data confirmed that the organization of genetic variation in *Picea abies* is similar to that observed in most forest tree species: the major proportion of gene diversity is harboured within populations, whereas the geographical differentiation is low (Hamrick *et al*. 1992; Hamrick & Godt 1996). We found for instance that the mean within-population diversity (H_S) was 0.79, while the differentiation (G_{ST}) estimated from AFLP data was only 0.029, in good consistency with the low values previously found with allozymes (0.05; Lagercrantz & Ryman 1990) or RAPDs (0.04; Collignon *et al*. 2002).

Beyond this general observation, it is important to notice that results obtained from different types of markers showed some discrepancies. This 'marker effect' affected in particular the diversity estimates which were higher when calculated from multiallelic and codominant SSR markers than from biallelic and dominant AFLP markers. Conversely, genetic differentiation was higher when inferred from AFLP markers as compared to SSRs. The same observations were made in *Quercus petraea* and *Quercus robur* by Mariette *et al*. (2002) and interpreted as a consequence of the higher mutation rate of microsatellites compared with AFLP and other types of markers in general (Weber & Wong 1993). Homoplasy among scored alleles of microsatellites can accentuate this underestimation of differentiation (Estoup *et al*. 1995).

Consistent with the multilocus analysis, the locus-per-locus examination of G_{ST}, revealed a significantly higher variation in amplitude for AFLP dominant markers than for microsatellite codominant markers. Their distribution, however, did not follow a general L-shaped curve, as reported in the studies of Mariette *et al*. (2002) and Scotti-Saintagne *et al*. (2004) on *Q. robur* and *Q. petraea*. We found 66% of AFLP markers showing G_{ST} values > 1% compared to only 29% in Scotti-Saintagne *et al*. (2004). This difference clearly suggests that the number of genomic regions involved in the intraspecific differentiation of a wide-ranging species such as *P. abies* is greater than that of the loci contributing to the interspecific divergence between the two sympatric and interfertile oak species studied by these authors.

Genomic differentiation

Several evolutionary forces such as mutation, random genetic drift, gene flow or natural selection influence variation across genomes and populations (Gaut *et al*. 2000). Selection and mutation have locus-specific effects while genetic drift and gene flow act at genome-wide scale (Lewontin & Krakauer 1973; Luikart *et al*. 2003).

In this study, we investigated the genomic organization of diversity on the basis of about 150 mapped markers (123/125 AFLPs, 25 SSRs and 2 ESTPs) evenly distributed over the nuclear genome of *P. abies*. The phenotypic analysis of the AFLP markers allowed us to distinguish the Baltico-Nordic population (domain D1) and the central European populations (domains D2 and D3), whereas D2 and D3 were significantly, but more weakly differentiated. This dual structure of genetic variation was fully consistent with previous results inferred from RAPDs (Collignon *et al*. 2002), cytoplasmic markers (Vendramin *et al*. 2000; Sperisen *et al*. 2001; Bastien *et al*. 2003) or quantitative traits (Collignon *et al*. 2002). Therefore, we first considered differentiation between the D1 and D2 + D3 pooled populations and then between the D2 and D3 populations. The genomic organization of differentiation appeared to be quite different. Indeed, results clearly showed that the differentiation between the Baltico-Nordic and central European pooled populations affected a significantly higher number of loci (25 loci, 16% of markers) than the differentiation among the central European populations (12 loci, 8% of markers) (*Z*-test, 5% level, $U > 1.96$). These 25 and 12 loci were scattered across the genome and their respective position on linkage groups was different. The minimal distance between these 25 loci or between these 12 loci was 4.6 cM. In this species with a particularly high genome size (18.6 pg $\approx 18.6 \times 10^9$ bp $\approx$ 2000 cM per 1C: Siljak-Yakovlev *et al*. 2002; Acheré *et al*. 2004), this distance represents as much as 43 Mb (physical/genetic size ratio of 9.3 Mb/cM).

Among the 37 significantly differentiated loci however, important variation of the G_{ST} values could be observed (from 0.008 to 0.20; Fig. 3). Most of them fit the neutral model expectations according to Beaumont and Nichols' approach (1996). However eight AFLPs and one SSR significantly deviated from the neutral model and could be considered as outlier loci (Luikart *et al*. 2003) probably subjected to selection. Detection of such outlier loci indicates locus-specific effects in the within species differentiation. Complementary approaches such as the detection of co-localization of outlier loci with QTL (quantitative trait loci) for adaptive traits, or the analysis of outlier loci distribution in natural populations showing clinal variation for phenotypic characteristics, could confirm the involvement of natural selection at these loci.

Four of these outlier loci, three AFLPs and one SSR located on distinct linkage groups, differentiated the Baltico-Nordic

from the central European group and, five, all unlinked AFLPs, distinguished the Alpine from the Hercyno-Carpathian populations. Their respective proportion among the significantly differentiated loci did not differ significantly (*Z*-test, 5% level, $U < 1.96$). Selective effects involved in both analysis of differentiation seem therefore to have comparable impact on molecular adaptation.

Finally, the results clearly indicate that all differentiated loci are not similarly affected by evolutionary forces. Most loci, evenly distributed over the linkage groups, probably reflect genetic drift and genome-wide gene flow effects that occurred during the quaternary history of the species. A higher number of loci (21) was involved in the differentiation of the Baltico-Nordic (D1) and central European populations (D2 + D3), than in the differentiation of the Hercyno-Carpathian (D2) and Alpine (D3) populations (seven loci). This difference can be interpreted as a greater long-term genetic legacy of refugial isolation of Baltico-Nordic populations compared to the others. Much data indeed, including molecular studies of mitochondrial or chloroplast polymorphism (Vendramin *et al*. 2000; Sperisen *et al*. 2001; Bastien *et al*. 2003) argue in favour of an early divergence of the Baltico-Nordic domain, probably resulting from an ancient separation. However, secondarily, high levels of gene flow have probably partially counterbalanced the long-term consequences of refugial isolation. Such exchanger has occurred intensively among the populations of the Hercyno-Carpathian and Alpine domains (D2 and D3) which probably met during their postglacial expansion and exchanged genetic material or completely admixed (Bergmann 1991). This resulted in a reduced number of genomic regions involved in differentiation and a partial erasing of the genetic signature of Pleistocene isolation of the Carpathian and Dinaric refuges. Long-distance wind dispersal of pollen has certainly favoured such introgressions, but ancient and repeated anthropic actions have also been implicated (Ledig 1992; Jeandroz *et al*. 2004).

The detection of nine outlier loci indicates that selective effects acting differentially in the different geographical domains also occurred, probably in connection to local adaptation to specific environmental conditions within the Nordic and central European populations. Indeed, owing to the wide latitudinal and longitudinal extension of its natural distribution range, *P. abies* is subjected to varied environmental pressures that can favour directional selection for some traits (Collignon *et al*. 2002). Assuming that the sample of 150 mapped markers used in this study represents a good coverage of the genome, the proportion of loci that might be under the effect of selection reaches 2.5% to 3.3%. These percentages are of the same magnitude than those recently reported by Campbell & Bernatchez (2004), who found that in lake whitefish (*Coregonus clupeaformis*) ecotypes, 1.4% to 3.2% of scored AFLP markers could be implicated in the adaptive differentiation. But, on the other hand, they are lower than the 12.7% detected by Scotti-Saintagne *et al*. (2004) in the divergence between *Q. petraea* and *Q. robur*. This higher percentage indicates that the number of genomic regions subjected to selection is probably higher in the interspecific divergence of these two oak species than in the intraspecific variation of *P. abies*.

In the future, although AFLP represents an important source of markers in term of genome covering, an extensive mapping of coding loci (ESTP) should improve the identification of outlier loci and allow a better understanding of the genetic basis of adaptation in this species.

Acknowledgements

We want to thank L. Bernatchez, A.M. Collignon, S. Mariette, C. Scotti-Saintagne and three anonymous reviewers for their help and useful comments on the manuscript.

References

Acheré V, Faivre Rampant P, Jeandroz S *et al*. (2004) A full saturated linkage map of *Picea abies* including AFLP, SSR, ESTP, 5S rDNA and morphological markers. *Theoretical and Applied Genetics*, **108**, 1602–1613.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating at high-density SNP map for signatures of natural selection. *Genome Research*, **12**, 1805–1814.

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*, **154**, 1309–1321.

Bastien D, Favre JM, Collignon AM, Sperisen C, Jeandroz S (2003) Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce [*Picea abies* (L.) Karst.]. *Theoretical and Applied Genetics*, **107**, 574–580.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **263**, 1619–1626.

Bergmann F (1991) Causes and consequences of species-specific genetic variation patterns in European forest tree species: examples with Norway spruce and Silver fir. In: *Genetic Variation in European Population of Forest Trees* (eds Müller-Starck G, Ziehe M), pp. 192–204. Sauerländer's-Verlag, Frankfurt am Main.

Besnard G, Acheré V, Faivre Rampant P, Favre JM, Jeandroz S (2003) A set of cross-species amplifying microsatellite markers developed from DNA-sequence databanks in *Picea* (Pinaceae). *Molecular Ecology Notes*, **3**, 380–383.

Bucci G, Vendramin GG (2000) Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular Ecology*, **9**, 923–934.

Campbell D, Bernatchez L (2004) Genomic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Cockerham CC, Weir BS (1987) Correlations, descent measures: drift with migration and mutation. *Proceedings of the National Academy of Sciences, USA*, **84**, 8512–8514.

Collignon AM, Van de Sype H, Favre JM (2002) Geographical variation in random amplified polymorphic DNA and

quantitative traits in Norway spruce. *Canadian Journal of Forest Research*, **32**, 266–282.

Estoup A, Tailliez C, Cornuet JM, Solignac M (1995) Size homoplasy and mutational process of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Molecular Biology and Evolution*, **12**, 1074–1084.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Gaut B, Le Thierry d'Ennequin M, Peek AS, Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. *Proceedings of the National Academy of Sciences, USA*, **97**, 7008–7015.

Goudet J (2002) FSTAT: *A program to estimate and test gene diversities and fixation indices*, (Version 2.9.3.2). University de Lausanne, Switzerland.

Goudet J, Raymond M, de-Meeus TFR (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Hamrick JL (2004) Response of forest trees to global environmental changes. *Forest Ecology and Management*, **197**, 323–335.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. In: *Plant Life Histories* (eds Silvertown J, Franco M, Harper JL), pp. 102–118. Cambridge University Press, Cambridge.

Hamrick JL, Godt MJW, Sherman-Broyles S (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hodgetts RB, Aleksiuk MA, Brown A *et al.* (2001) Development of microsatellite markers for white spruce (*Picea glauca*) and related species. *Theoretical and Applied Genetics*, **102**, 1252–1258.

Huntley B, Birks HJB (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0–13 000 Years Ago*. Cambridge University Press, Cambridge.

Isabel N, Beaulieu J, Bousquet J (1995) Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. *Proceedings of the National Academy of Sciences, USA*, **92**, 6369–6373.

Isabel N, Beaulieu J, Thériault P, Bousquet J (1999) Direct evidence for biased gene diversity estimates from dominant random amplified polymorphic DNA (RAPD) fingerprints. *Molecular Ecology*, **8**, 477–483.

Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004) Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology*, **13**, 2735–2747.

Jeandroz S, Collignon AM, Favre JM (2004) RAPD and mtDNA variation among autochthonous and planted populations of *Picea abies* from the Vosges Mountains (France) in reference to other French populations. *Forest Ecology and Management*, **197**, 225–229.

Kayser M, Brauer S, Stoneking M (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. *Molecular Biology and Evolution*, **20**, 893–900.

Krutovskii KV, Erofeeva SY, Aagaard JE, Strauss SH (1999) Simulation of effects of dominance on estimates of population genetic diversity and differentiation. *Journal of Heredity*, **60**, 499–502.

Krutzsch P (1974) The IUFRO 1964/68 provenance test with Norway spruce (*Picea abies* (L.) Karst.). *Silvae Genetica*, **23**, 1–3.

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

Ledig T (1992) Human impacts on genetic diversity in forest ecosystems. *Oikos*, **63**, 87–108.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Mariette S, Cottrell J, Csaikl UM *et al.* (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genetica*, **51**, 72–79.

Müller-Starck G, Baradat P, Bergmann F (1992) Genetic variation within European tree species. *New Forests*, **6**, 23–47.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **70**, 3321–3332.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Pfeiffer A, Olivieri AM, Morgante M (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome*, **40**, 411–419.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rohlf FJ (2000) *NTSYS-Pc: Numerical Taxonomy and Multivariate Analysis System*, Version 2.1. Exeter Publications, New York.

Rouppe van der Voort JNAM, van Zandvoort P, van Eck HJ *et al.* (1997) Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. *Molecular and General Genetics*, **255**, 438–447.

Schmidt-Vogt H (1986) Die Fichte, *Band II/1*. Parey Ed, Berlin.

Schubert R, Mueller-Starck G, Riegel R (2001) Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea abies* (L.) Karst. *Theoretical and Applied Genetics*, **103**, 1223–1231.

Scotti I, Magni F, Fink R *et al.* (2000) Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. *Genome*, **43**, 41–46.

Scotti I, Magni F, Paglia GP, Morgante M (2002a) Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers. *Theoretical and Applied Genetics*, **106**, 40–50.

Scotti I, Paglia GP, Magni F, Morgante M (2002b) Efficient development of dinucleotide microsatellite markers in Norway spruce (*Picea abies* Karst.) through dot-blot selection. *Theoretical and Applied Genetics*, **104**, 1035–1041.

Scotti-Saintagne C, Mariette S, Porth I *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics*, **168**, 1615–1626.

Siljak-Yakovlev S, Cerbah M, Couland J *et al.* (2002) Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea omorika* and *Picea abies*. *Theoretical and Applied Genetics*, **104**, 505–512.

Sperisen C, Büchler U, Gugerli F *et al.* (2001) Tandem repeats in plant mitochondrial genomes: application to the analysis

of population differentiation in the conifer Norway spruce. *Molecular Ecology*, **10**, 257–263.

Szmidt AE, Wang XR, Lu MZ (1996) Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* (L.) using haploid tissue analysis. *Heredity*, **76**, 412–420.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Tani N, Maruyama K, Tomaru N *et al.* (2003) Genetic diversity of nuclear and mitochondrial genomes in *Pinus parviflora* Sieb. & Zucc. (Pinaceae) populations. *Heredity*, **91**, 510–518.

Vendramin GG, Anziedei M, Madaghiele A, Sperisen C, Bucci G (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). *Genome*, **43**, 68–78.

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV (2000) Are mapped markers more useful for assessing genetic diversity? *Theoretical and Applied Genetics*, **100**, 607–613.

Waugh R, Bonar N, Baird E *et al.* (1997) Homology of AFLP products in three mapping populations of barley. *Molecular and General Genetics*, **255**, 311–321.

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

Yang RC, Yeh FC, Yanchuk AD (1996) A comparison of isozyme and quantitative genetic variation in *Pinus contorta* ssp. *latifolia* by F_{ST}. *Genetics*, **142**, 1045–1052.

Yeh FC, Yang RC, Boyle T, Ye BJ, Mao J (1997) *POPGENE, the user-friendly shareware for population genetic analysis*. Agric-Food and Forestry Molecular Biology Centre, University of Alberta, Canada.

This research forms part of the V. Acheré PhD thesis about genome evolution and population genetics of *Picea abies*. JM Favre is an emeritus professor at the Nancy I University, involved in the study of forest tree genetics. G. Besnard has now joined the Department of Ecology and Evolution (Lausanne University) and focuses on development of molecular markers to study plant re-colonization and the evolution of adaptive traits. S. Jeandroz is assistant-professor at the Nancy I University. His studies include molecular systematics and phylogeny of forest trees.
