Nucleotide variation in genes involved in wood formation in two pine species

David Pot[1,2], Lisa McMillan[3], Craig Echt[3,4], Grégoire Le Provost[1], Pauline Garnier-Géré[1], Sheree Cato[3] and Christophe Plomion[1]

[1]UMR 1202 BIOGECO, INRA, 69 route d'Arcachon, 33612 Cestas Cédex, France; [2]UMR 1096 PIA, CIRAD, Avenue Agropolis, 34398 Montpellier Cédex 5, France; [3]Cell Wall Biotechnology Centre (CBC), Forest Research Institute, Rotorua, New Zealand; [4]Southern Institute of Forest Genetics, USDA Forest Service, 23332 Hwy 67, Saucier, MS, USA

Author for correspondence:
Christophe Plomion
Tel: +33 5 57 12 28 30
Fax: +33 5 57 12 28 81
Email: plomion@pierroton.inra.fr

Received: *23 November 2004*
Accepted: *15 February 2005*

Summary

- Nucleotide diversity in eight genes related to wood formation was investigated in two pine species, *Pinus pinaster* and *P. radiata*.
- The nucleotide diversity patterns observed and their properties were compared between the two species according to the specific characteristics of the samples analysed.
- A lower diversity was observed in *P. radiata* compared with *P. pinaster*. In particular, for two genes (*Pp1*, a glycin-rich protein homolog and *CesA3*, a cellulose synthase) the magnitude of the reduction of diversity potentially indicates the action of nonneutral factors. For both, particular patterns of nucleotide diversity were observed in *P. pinaster* (high genetic differentiation for *Pp1* and close to zero differentiation associated with positive Tajima's *D*-value for CesA3). In addition, *KORRIGAN*, a gene involved in cellulose–hemicellulose assembly, demonstrated a negative Tajima's *D*-value in *P. radiata* accompanied by a high genetic differentiation in *P. pinaster*.
- The consistency of the results obtained at the nucleotide level, together with the physiological roles of the genes analysed, indicate their potential susceptibility to artificial and/or natural selection.

Key words: candidate gene, nucleotide variation, *Pinus*, selection, wood formation.

New Phytologist (2005) **167**: 101–112

© *New Phytologist* (2005) **doi**: 10.1111/j.1469-8137.2005.01417.x

Introduction

Identification of genes controlling quantitative trait variation is one of the great challenges of the post-genomic era. This knowledge is important not only for biomedicine but also for agriculture. In this latter field, such information would provide a way to manage and use the genetic variability in breeding and gene conservation programmes. The availability of markers linked to economically and ecologically relevant traits would be of particular interest in long-lived forest trees species such as conifers. Such tools would enhance the efficiency of artificial selection by reducing the duration of breeding cycles and increasing the genetic gain in each cycle. They would also provide criteria to manage functional genetic diversity, which is key to preserving adaptability of forest trees to their changing environment. Traditional forest tree breeding programmes have provided the forest industry with improved genotypes for wood production (e.g. 30% realized genetic gain for the volume of the bole in *Pinus pinaster*; Alazard & Raffin, 2003). The introduction of criteria towards wood quality selection is now considered as an important objective to ensure the sustainability of the wood market through the availability of raw material well suited to end-use products (Pot *et al.*, 2002).

Wood property quantitative trait loci (QTLs) have been identified in many forest tree species, attesting the existence of major gene effects controlling part of the variation of wood and its end-use properties (Bradshaw & Stettler, 1995; Grattapaglia *et al.*, 1996; Kumar *et al.*, 2000; Lerceteau *et al.*, 2000; Arcade *et al.*, 2002; Moran *et al.*, 2002; Neale *et al.*, 2002; Brown *et al.*, 2003; Markussen *et al.*, 2003). Co-localizations of QTLs and candidate genes (Moran *et al.*, 2002;

Brown *et al.*, 2003; Chagné *et al.*, 2003) have also been reported. However, given the large confidence intervals generally associated with QTLs (Mangin *et al.*, 1994), these findings did not permit their validation. Complex trait dissection allowing the identification of individual genes is currently underway through association studies in humans and model animals (e.g. *Drosophila*). Recently, Thornsberry *et al.* (2001) have successfully transferred this approach to plants. In theory, association studies should be performed at a whole-genome level (also known as a genome scan); however, due to the specific features of conifers (short distance linkage disequilibrium (Brown *et al.*, 2004; Neale & Savolainen, 2004) and extremely large genome size (Wakamiya *et al.*, 1993)), a candidate-gene approach is the only possible way to understand the molecular basis underlying quantitative variation in these species.

Wood formation includes four major steps: cell division, cell expansion, secondary cell wall formation and cell death. These steps involve expression of a number of structural genes, coordinated by transcription factors, mainly involved in the biosynthesis of polysaccharides (cellulose: 40–50% of dry wood; hemi-cellulose: 25%; and pectins), lignins (25–35%), and cell wall proteins. A number of genes that determine cell wall composition and cell shape have been identified by classical biochemical analysis (e.g. lignification genes, reviewed in Whetten *et al.*, 1998), and more recently by the application of the genomic tools such as gene or protein expression profiling (Plomion *et al.*, 2000; Hertzberg *et al.*, 2001; Le Provost *et al.*, 2003; Gion *et al.*, 2005) and the screening of large collections of *Arabidopsis thaliana* mutants (Fagard *et al.*, 2000; Mouille *et al.*, 2003).

Several studies have shown that wood structure and composition are influenced by environmental changes (Liphschitz & Waisel, 1970; Barber *et al.*, 2000). The extent of these modifications has also been shown to be genetically regulated (Rozenberg *et al.*, 2002), suggesting the potential functional role of xylogeneic genes in forest trees adaptation (Costa *et al.*, 1998; Riccardi *et al.*, 1998). In this context, it is possible that nucleotide diversity of these genes and their homologs in pine are involved in genetic variation of wood properties and, as such, may be subject to natural selection pressures in pine species.

For this study, eight candidate genes were selected based on their likely involvement in the determination of wood properties. Three were homologous to *Arabidopsis thaliana* cell wall mutant genes specifically involved in the cellulose and hemicellulose biosynthesis (a membrane-bound endo-1,4-beta-glucanase, *KORRIGAN*, and two cellulose synthases, *CESA3* and *CESA4*). Five expressional candidate genes were also analysed. These genes have been identified through differential expression studies between different types of wood characterized by distinct chemical composition and structure (reviewed in Plomion *et al.*, 2001). *Pp2* (MYB-like transcriptional factor), *Pp4* (ACC oxidase) and *Pp6* (25S ribosomal gene) have been identified as being up-regulated in early wood, whereas *Pp1* (glycine-rich protein homolog) was found to be up-regulated in late wood-forming tissue (Le Provost *et al.*, 2003). *Pr1* (unknown function protein) was isolated from wood forming tissue in *P. radiata* (S. Cato, unpublished data).

In the present study, nucleotide variation of these eight genes was analysed within and between two pine species: *Pinus pinaster* Ait. and *Pinus radiata* D.don, both of which are economically and ecologically important. Both species are currently the target for conservation efforts, and the accurate determination of their genetic structure at the functional level would help refine conservation strategies.

P. pinaster has a highly fragmented distribution over 4 Mha in the Mediterranean basin. This natural range includes highly variable climatic conditions, from more than 1000 mm rainfall in Tova (Corsica) to less than 100 mm in Oria (Spain), and soil structure that varies from sandy dunes to shallow rocky soils. The genetic structure of the species has been described using several sets of markers (reviewed in Burban & Petit, 2003) and reveals 18 geographically structured races belonging to three major groups: an Atlantic group, comprising populations from western France and the greater part of Spain and Portugal; a Mediterranean group, consisting of all eastern European populations, and including eastern Spanish populations up to Andalucía and the small stand of Punta Cires in Morocco; and a North African group comprising all the other African populations. Because of the fragmentation of its natural range, maritime pine exhibits a relatively high genetic differentiation among populations at nuclear markers in comparison to other conifer species. A high level of genetic differentiation was also observed for survival, adaptation to different climatic conditions, growth and phenology, resistance to insects and drought tolerance (reviewed in González-Martínez *et al.*, 2002).

P. radiata grows naturally in five locations: Año Nuevo, Monterey and Cambria on the Californian mainland coast and Guadalupe and Cedros islands off the coast of Baja California. These five locations differ substantially from each other with respect to soil, elevation, temperature, rainfall and ecosystem associates. At the genetic level, significant differentiation was observed between the different populations (ranging between 0.119 and 0.26, depending on the type of markers and the populations considered; Moran *et al.*, 1988; Wu *et al.*, 1999; Karhu, 2001). Although the natural range of *P. radiata* is extremely small, it is the world's most widely planted fast-growing softwood species. It is cultivated on a commercial scale in Australia, Chile, South Africa and New Zealand.

The objectives of this study were twofold. The first was to study the patterns of nucleotide diversity of the eight chosen candidate genes in *P. radiata* and *P. pinaster*. More explicitly, we described for the first time in these two species, the type (SNP vs INDEL), nature (silent vs nonsynonymous) and genomic location (coding vs noncoding) of nucleotide polymorphisms. The second goal was to investigate whether nucleotide diversity patterns were compatible with neutral models or not.

Table 1 List of *Pinus pinaster* and *Pinus radiata* populations

Species	Country	Population	Latitude	Longitude	Altitude (m)	Sample size[a]	Group[b]
P. pinaster	Tunisia	Tabarka	36°57′ N	8°46′ E	200	2 (8)	Mediterranean
	France	Corsica Porto Vecchio	41°28′ N	9°12′ E	150	2 (7)	Mediterranean
		Corsica Vivario	41°20′ N	9°09′ E	600	2 (5)	Mediterranean
		Corsica Zonza	41°45′ N	9°11′ E	760	2 (6)	Mediterranean
		Aquitaine Castets	43°52′ N	1°08′ W	60	2 (3)	Atlantic
		Aquitaine Mimizan	44°08′ N	1°18′ W	35	2 (9)	Atlantic
		Aquitaine Souston	43°41′ N	1°25′ W	35	2 (4)	Atlantic
		Aquitaine Hourtin	45°10′ N	1°08′ W	40	4 (10)	Atlantic
		Aquitaine Medoc	45°34′ N	1°13′ W	40	2 (18)	Atlantic
	Portugal	Leiria Mata	40°00′ N	8°45′ W	50	1 (5)	Atlantic
		Leiria Velha	40°00′ N	8°45′ W	50	1 (4)	Atlantic
	Morocco	Punta Cires	35°55′ N	5°28′ W	20	1 (6)	Atlantic
		Tamjout	33°52′ N	4°02′ W	1600	1 (4)	North African
P. radiata	New Zealand	NZ breeding population (land race)				23	Año Nuevo and Monterey

[a]Number of megagametophytes analysed per gene. For *CesA3*, a wider sample was studied. The sample size analysed for each population for this gene is indicated in parentheses.
[b]*P. pinaster* groups based on Burban & Petit (2003); *P. radiata* groups based on Burdon *et al.* (1997a,b).

Materials and methods

Plant material and DNA extraction

Haploid megagametophytes, a maternal tissue surrounding the diploid embryo in conifer seeds, were harvested from germinated seedlings just before the seed coat was cast off. Genomic DNA was extracted as described by Plomion *et al.* (1995). *P. pinaster* nucleotide diversity was assessed using megagametophytes collected from natural stands across the species natural range (Table 1). Twenty-four gametes from 13 provenances belonging to the three main groups identified by Baradat and Marpeau (1988) were included in this exploratory analysis. In a second step, for one of the genes (*CesA3*) the sample size was extended to 91 megagametophytes (Table 1). *P. radiata* nucleotide diversity was estimated using 23 megagametophytes collected from individual trees of the New Zealand breeding population (Forest Research, Rotorua, New Zealand). Previous studies, based on monoterpene analysis (Burdon *et al.*, 1997a) and morphological traits (Burdon *et al.*, 1997b) have shown that the local race was introduced from the USA during the 19th century and mostly derived from the Año Nuevo population, with some admixture from the Monterey population.

Primer design, PCR amplification and DNA sequencing

For each gene, a BLAST search (Altschul *et al.*, 1997) was first run to identify homologs in pine expressed sequence tag (EST) databases available at http://cbi.labri.fr/outils/SAM/COMPLETE/index.php for *P. pinaster* and http://fungen.org/Projects/Pine/Pine.htm for *Pinus taeda* (Table 2). From the multiple alignments of the retrieved sequences, a consensus sequence was then derived for each candidate using SEQUENCHER v4.1.4 (Genecodes, Inc, Ann Arbor. Michigan USA). Primer pairs (Table 3) were designed from the consensus sequence using PRIMER 3 (Rozen & Skaletsky, 2000).

PCR products were sequenced using the Big Dye terminator kit (Amersham Bioscience, Uppsala, Sweden) and an ABI 3100 automatic sequencer (Applied Biosystem, Foster City, CA, USA) according to the manufacturers' specifications. A single sequence was obtained per megagametophyte for each candidate gene. Singleton polymorphisms were verified through re-sequencing of the affected megagametophyte sample.

Landscape of nucleotide diversity

Sequence alignment and nucleotide polymorphism detection were performed with SEQUENCHER v4.1.4. Each polymorphic site was visually checked on the chromatograms in order to distinguish true polymorphisms from scoring errors. The use of haploid tissues greatly facilitated the sequence analysis, allowing the direct definition of the haplotypes (multilocus combinations of polymorphisms) without cloning or using an expectation maximization (EM) algorithm (Long *et al.*, 1995).

Basic parameters including the number of single nucleotide polymorphisms (SNPs), insertion–deletions (INDELs), synonymous (S) and nonsynonymous (NS) mutations were calculated using the SITE software (Hey & Wakeley, 1997). Nucleotide diversity was estimated as θ_w (based on the number of segregating sites; Watterson, 1975) and π (based on the average number of nucleotide differences per site between sequences; Nei, 1987). These parameters were computed with SITE, without considering INDELs, at three different levels: (i) the whole sequenced region; (ii) noncoding regions (including introns, 3′ and 5′ untranslated regions (UTRs)); and (iii) coding regions, subdivided in two components – S and NS.

Table 2 Summary of the studied genes

Gene ID	Function	Accession[a]	Base pairs screened			Number of homologs with pine EST (E-value < 1^{-10})		
			Total	Exon	Intron	3′ UTR	*P. pinaster*[b]	*P. taeda*[c]
KORRIGAN	membrane-bound endo-(1-4)-β-glucanase (EC:2.4.1.12)	BV079723	937	566	371	0	8	47
CesA3	cellulose synthase (EC:2.4.1.12)	BV079715 + BV079717	1048	810	238	0	8	49
CesA4	cellulose synthase (EC:2.4.1.12)	BV079716	489	396	0	93	7	22
Pp1	glycine-rich protein homolog	BV079718	493	240	0	253	22	133
Pp2	MYB-like transcriptional factor MBF1	BV079719	494	494	0	0	1	3
Pp4	ACC oxidase	BV079720	270	270	0	0	1	42
Pp6	25S rRNA gene	BV079721 + BV079722	902	902	0	0	33	176
Pr1	unknown protein	BQ701569	113	113	0	0	0	15

[a]Accession number deposited in dbSTS
[b]Search performed at http://cbi.labri.fr/outils/SAM/COMPLETE/index.php
[c]Search performed at http://fungen.org/Projects/Pine/Pine.htm

Table 3 List of primer pairs and amplification conditions

Gene ID	Primer pairs		Amplification conditions	
	Forward primer	Reverse primer	T_a (°C)	Mg^{2+} (mM)
KORRIGAN	GCAGGACTATGGTGTTTTAAGC	TATTCCCCCAGTATCACCCC	59–50	3
CesA4	AGATCTTGCTCAATGCCTCG	CCAAACTTCACTGTCACATCG	59–50	3
CesA3a[a]	GCTTTGAGAAGTCGTTTGGC	GTATGCCAGTCTTTCCAGCC	64–55	2
CesA3b[a]	CATTGGTTCGAGTCTCTGCC	TAACACACCAAGAGGCCACC	59–50	3
Pp1	GAGTTCTCAAGGGATGTCGG	TAACACACCAAGAGGCACC	59–50	3
Pp2	AACAGATCATCCATCTCGGG	ACAGATGGTCATTGATCGCC	59–50	3
Pp4	GAACATCTACCCTGCTTGCC	TGAAATTCCTAACATGCTCCC	59–50	3
Pp6a[a]	TTTTGATCCTTCGATGTCGG	GAATCTCAGTGGATCGTGGC	59–50	3
Pp6b[a]	AAATTCAACCAAGCGCGG	CTTTTAACAGATGTGCCGCC	59–50	2
Pr1	ATCGCATGGGAGTTGCAG	CATGTCAGCCTCGGTTTGG	64–55	2

[a]For *Pp6* and *CesA3*, two primer pairs were designed.

The number of haplotypes and the haplotype diversity were calculated using the DNASP software (Rozas & Rozas, 1999).

Tests for selection were performed to estimate whether the considered genes followed the model of neutral evolution (Kimura, 1983) or not. Tajima's *D*-test, based on the allelic distribution (Tajima, 1989), was carried out using ARLEQUIN 2.000 software (Shneider *et al.*, 2000). As implemented in this software, significance of this test was tested by generating random samples under the hypothesis of neutrality and population demographic equilibrium. This test was performed assuming the absence of recombination, making it conservative.

Levels of differentiation (F_{ST}) between Corsican and Aquitaine populations were estimated for all the studied genes with the analysis of molecular variance (AMOVA) (Excoffier *et al.*, 1992) as implemented in the ARLEQUIN 2.000 software (Schneider *et al.*, 2001). In addition, differentiation among all the studied populations was also estimated for *CesA3*.

Considering first the small number of sequences analysed for most of the genes, and second the existence of significant differentiation for some of them, linkage disequilibrium (LD) was only computed for *CesA3* for which a larger sample size was available. Given the absence of significant differentiation for this gene, LD between polymorphic sites was estimated using the whole set of sequence with DNASP. Fisher's exact tests and Bonferroni correction for multiples tests were computed to determine whether the detected associations were significant or not.

Total divergence between *P. pinaster* and *P. radiata*, estimated as the average number of nucleotide substitutions per site, was finally calculated using DNASP.

Results

Nucleotide variation at the intraspecific level

Sequence data for almost the complete set of gametes were obtained for six out of the eight genes analysed. For *Pp1* and *Pp2*, only 12 and 14 high-quality sequences, respectively, were

obtained in *P. radiata*, probably as a result of the coamplification of other family members. Sequences were deposited in dbSTS (http://www.ncbi.nlm.nih.gov/dbSTS/) and SNPs in dbSNP (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp).

The regions analysed covered a total of 4.7 kb, corresponding to 3.8 kb of coding sequence and 0.9 kb of-noncoding regions (intron and 3′ UTR) (Table 2). A total of 32 (29 SNPs and three INDELs) and 13 (exclusively SNP) intraspecific polymorphisms were detected in *P. pinaster* and *P. radiata*, respectively. All the INDELs were single-based and located in noncoding regions. A total of 10 singletons were identified (seven in *P. pinaster* and three in *P. radiata*). All the nonsynonymous polymorphisms were conservative or moderately conservative according to the classification of Grantham (1974).

The average nucleotide diversity was slightly higher for *P. pinaster* (0.00241) than for *P. radiata* (0.00186). This difference mainly relied on *Pp1* and *CesA3*, for which 11 and nine polymorphic sites were detected in *P. pinaster*, whereas only one and two polymorphic sites, respectively, were detected in *P. radiata*. Although the numbers of sequences analysed were smaller for *P. radiata* for these two genes, π will not be better estimated with a sample of sequences above 10 as its variance levels off very quickly (Tajima, 1983), thus the divergence of the estimates probably does not result from these unequal sample sizes. Apart from these two genes, close correspondence between the nucleotide diversity estimates in the two species was observed.

The average number of haplotypes (3.375 in *P. pinaster* vs 2.375 in *P. radiata*) and the average haplotype diversity (0.425 in *P. pinaster* vs 0.376 and *P. radiata*) were, like the total nucleotide diversity, slightly higher in *P. pinaster*. Large variations in haplotype number and haplotype diversity were observed among the genes in both species. The number of haplotypes varied from one to six. With the exception of *CesA3* and *Pp1*, the numbers of haplotypes were consistent among species.

Neutrality tests

Tajima's *D*-tests were performed exclusively for the genes presenting at least five polymorphic sites (Table 4). Significant departure from the null hypotheses of neutrality and demographic equilibrium at $P < 0.05$ was observed only for *KORRIGAN* in *P. radiata*. For all the genes, these tests were performed on the whole set of sequences available.

For *CesA3*, Tajima's *D*-test was first performed in the Aquitaine provenance exclusively, and then, according to the nonsignificant level of differentiation observed for this gene (see the next section, 'Populations differentiation in *P. pinaster*'), performed also considering all the sequences available. Both calculations yielded the same result, that is a positive but nonsignificant Tajima's *D*-value ($D = 1.12147$, $P = 0.117$ for the whole area of distribution and $D = 1.03263$, $P = 0.131$ for the Aquitaine provenance). Estimation of the local recombination parameter *R* (Hudson, 1987) for this gene, and its subsequent integration in the calculation of the Tajima's *D*-value expected distribution using coalescence simulation in DNASP, did not change its significance.

Populations differentiation in *P. pinaster*

F_{ST} estimated using all the polymorphic sites revealed a significant differentiation between Corsican and Aquitaine provenances ($F_{ST} = 0.22$). However, this high level of differentiation relied exclusively on two genes (*KORRIGAN* and *Pp1*) for which highly significant differentiation were observed (0.45 and 0.23, respectively, in Table 5). In comparison, Mariette *et al.* (2001) observed a significant G_{ST} value of between 0.049 and 0.092 for amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR), respectively. Accounting for the difference in estimation methods which show that F_{ST} values are equivalent to twice the G_{ST} values (Nei, 1987), the differentiation between both groups of populations for both *Pp1* and *KORRIGAN* is more than twice as high as that obtained for AFLP markers (0.098). Compared with SSR (0.184), although the differentiation values observed for both genes remain higher, the *Pp1* value of differentiation is just slightly higher, whereas that of *KORRIGAN* is still twice as high.

In addition, a wider sampling for CesA3 allowed us to test the differentiation between the 13 populations. No significant differentiation was observed and the estimated value is very close to zero. This result deviates from the significant differentiation observed with neutral markers at the level of the whole geographic distribution of maritime pine (Petit *et al.*, 1995).

If F_{ST} estimates are probably dependent on the very few polymorphic sites (from one to four) detected for *CesA4*, *Pp2*, *Pp4* and *Pr1*, the differentiation estimated for *KORRIGAN*, *CesA3* and *Pp1* for which at least five polymorphic sites were analysed are certainly more representative of the gene values.

Linkage disequilibrium

Linkage disequilibrium was only calculated for *CesA3*, as the other genes presented either strong population differentiation combined with only small population size analysed or low level of polymorphism. Out of the 36 tests performed (nine polymorphic sites), 11 were significant, after Bonferroni's correction for multiple testing.

Nucleotide variation at the interspecific level

Sixty-three polymorphisms including 59 SNPs and four INDELs distinguished *P. pinaster* from *P. radiata* (Table 6). All INDELs were located in the noncoding region. The total number of interspecific fixed differences varied from 0 for *Pr1* to 21 for *CesA3* (Table 6). NS fixed differences were found for five genes (*KORRIGAN*, *CesA3*, *Pp1*, *Pp2* and *Pp4*). Three of

Table 4 Pattern of nucleotide variation

Pinus pinaster Gene ID	*KORRIGAN*	*CesA3*	*CesA4*	*Pp1*	*Pp2*	*Pp4*	*Pp6*	*Pr1*	Total
Number of sequences	24	91	23	24	22	24	24	24	**256**
INDEL	1	0	0	2	0	0	0	0	**3**
SNP									
Total									
S[a]	5	9	1	9	2	1	0	2	**29**
Singleton	0	1	1	3	1	0	0	1	**7**
π	0.00176	0.00260	0.00019	0.00696	0.00121	0.00083	0	0.00573	**0.00241**
θ_w	0.00173	0.00177	0.00058	0.00515	0.00116	0.00107	0	0.0056	**0.00213**
Tajima's *D*	0.63089	1.12147	–	1.20137	–	–	–	–	
Noncoding									
S	3	4	–	7	–	1	–	–	**15**
π	0.00132	0.00102	–	0.0066	–	0.00083	–	–	**0.00244**
θ_w	0.00104	0.00082	–	0.004	–	0.00107	–	–	**0.00173**
Coding									
Total									
S	2	5	1	2	2	0	0	–	**12**
π	0.00045	0.00084	0.00019	0.00036	0.00121	0	0	–	**0.00043**
θ_w	0.00069	0.00102	0.00058	0.00114	0.0116	0	0	–	**0.00214**
Synonymous									
S	2	3	1	2	1	0	0	–	**9**
π	0.00045	0.00078	0.00019	0.00036	0.00018	0	0	–	**0.00028**
θ_w	0.00035	0.00061	0.00058	0.00114	0.00058	0	0	–	**0.00046**
Nonsynonymous									
S	0	2	0	0	1	0	0	–	**3**
π	0	0.00006	0	0	0.00103	0	0	–	**0.00015**
θ_w	0	0.00041	0	0	0.00058	0	0	–	**0.00014**
Number of haplotypes	5	6	2	6	3	2	1	3	**3.375**
Haplotype diversity (SE)	0.809 (0.057)	0.607 (0.059)	0.091 (0.081)	0.656 (0.079)	0.511 (0.091)	0.268 (0.113)	0	0.537 (0.052)	**0.425**

Pinus radiata Gene ID	*KORRIGAN*	*CesA3*	*CesA4*	*Pp1*	*Pp2*	*Pp4*	*Pp6*	*Pr1*	Total
Number of sequences	18	21	20	12	14	23	23	21	**153**
INDEL	0	0	0	0	0	0	0	0	**0**
SNP									
Total									
S[a]	5	2	1	1	2	0	0	1	**13**
Singleton	2	0	0	0	1	0	0	0	**3**
π	0.00198	0.00048	0.00101	0.00114	0.00374	0	0	0.00645	**0.00186**
θ_w	0.00291	0.00100	0.00064	0.00089	0.00359	0	0	0.00588	**0.00191**
Tajima's *D*	–1.97[b]	–	–	–	–	–	–	–	
Noncoding									
S	3	1	–	1	–	0	–	–	**5**
π	0.00139	0.00026	–	0.00114	–	0	–	–	**0.00069**
θ_w	0.00175	0.0005	–	0.00089	–	0	–	–	**0.00078**
Coding									
Total									
S	2	1	1	0	2	0	0	–	**6**
π	0.00058	0.00022	0.00101	0	0.00384	0	0	–	**0.00081**
θ_w	0.00117	0.0005	0.00064	0	0.04	0	0	–	**0.00604**
Synonymous									
S	1	0	1	0	2	0	0	–	**4**
π	0.00043	0	0.00101	0	0.00384	0	0	–	**0.00075**
θ_w	0.00058	0	0.00064	0	0.04	0	0	–	**0.00588**
Nonsynonymous									
S	1	1	0	0	0	0	0	–	**2**
π	0.00016	0.00022	0	0	0	0	0	–	**0.00005**
θ_w	0.00058	0.0005	0	0	0	0	0	–	**0.00015**
Number of haplotypes	4	2	2	2	3	1	1	4	**2.375**
Haplotype diversity (SE)	0.673	0.327 (0.123)	0.456 (0.153)	0.556 (0.085)	0.604 (0.09)	0 (0.076)	0	0.339 (0.138)	**0.376**

[a]Number of SNPs; [b]Significant Tajima's D-value ($P < 0.05$).

Table 5 F_{ST} estimates between Corsican and Aquitaine populations

Gene	F_{st}
CesA3	−0.05482
KORRIGAN	0.45267[a]
Pp1	0.23280[a]
Pp2	−0.14549
Pp4	0.14286
Pr1	−0.05504
All (31 polymorphic sites)	0.22395[a]
G_{ST} AFLP–SSR (Mariette *et al.* 2001)	0.049–0.092

[a]Significant test.

these NS fixed differences were moderately radical regarding the amino acid modification (Grantham, 1974): two sites in *Pp2* (modifications SER to ARG and GLY to ARG) and one site in *Pp4* (VAL to SER).

Under neutral evolution, interspecific divergence is expected to be proportional to intraspecific nucleotide diversity. Comparison of divergence and nucleotide diversity revealed that only *Pr1* diverged from this pattern. However, due to the small size of the fragment analysed (113 bp), no particular hypothesis could be provided. A wider exploration of the diversity of this gene will be required before any conclusion can be drawn.

Discussion

Adequacy between the sampling strategy and SNP detection probability

The probability P of detecting the two alleles at a SNP locus depends on three parameters: (i) the number of gametes sampled, N; (ii) the frequency of the rare allele in the population, p; and (iii) the organization of gene diversity. In the absence of differentiation among populations, $P = 1 - (1 - p)^N$. In the present study, for each species, on average 21 gametes were sequenced for each DNA fragment, resulting in a detection probability of 89% for a rare allele frequency of 10%.

In respect to *P. pinaster*, the probability of detecting polymorphic loci was probably maximized considering: (i) the scattered sample used in our study covering the three main groups of diversity; (ii) the moderate level of genetic differentiation at the neutral level between geographical provenances (G_{ST} = 0.14–0.17; Petit *et al.* 1995) for isozymes, proteins and terpenes, with populations from France, Portugal, Corsica, Spain, Italy, Sardinia; and (iii) the rather low differentiation within provenances (G_{ST} = 0.04 for isozymes, cpSSR, nuclear SSR and AFLP markers, within Spain, Portugal, Aquitaine and Corsica: Mariette *et al.*, 2001; González-Martínez *et al.*, 2002; Ribeiro *et al.*, 2002). However, it is important to note that the North African group, which constitutes a singular mitochondrial lineage, with highly differentiated populations, was under-represented in this study and probably led to an underestimation of the nucleotide diversity of some of the genes.

Concerning *P. radiata*, as reported in the Material and Methods section, the sample used in this study corresponds to the first generation of the New Zealand breeding population, which derived from the Año Nuevo population with some admixture from the Monterey population. Johnson and Lipow (2002) showed that first-generation seed orchards retain most of the genetic diversity present in the natural populations from which they were derived. As a consequence, the results obtained for *P. radiata* should reflect the nucleotide diversity present in its ancestral populations. Indeed, using nuclear and chloroplast microsatellite loci, no significant changes in diversity were found between the five natural populations of *P. radiata*, and the current New Zealand breeding populations (T. Richardson, Forest Research, New Zealand, pers. comm.). It is, however, important to note that, according to the selection criteria used to select the first-generation breeding population (i.e. growth and form), some of the genes controlling these traits could have been submitted to artificial selection events leading to a reduction in their diversity.

Nucleotide diversity in wood formation related genes

Polymorphic sites were found in almost all the genes analysed, providing the basis to initiate association studies to test the

Table 6 Fixed differences between *Pinus pinaster* and *Pinus radiata* and estimates of total divergence D (x,y)

		SNP			
Gene ID	Number of INDELs	Noncoding	Coding synonymous	Coding nonsynonymous	D (x,y)
KORRIGAN	3	2	3	1	0.00686
CesA3	0	8	8	5	0.02098
CesA4	1	1	3	0	0.0083
Pp1	0	6	3	2	0.0234
Pp2	0	0	6	6	0.02444
Pp4	0	0	2	2	0.0155
Pp6	0	2	0	0	0.00228
Pr1	0	0	0	0	0

Table 7 Estimates of nucleotide diversity in different species

Species	Number of loci	Number of genotypes	Length (bp)	Coverage of the natural distribution	Total nucleotide diversity (π)	Reference
Pinus pinaster	10	22–91	4 746	yes	0.00241	this study
Pinus radiata	10	12–24	4 746	no	0.00186	this study
Pinus taeda	19	32	17 580	yes	0.00395	Brown *et al.* (2004)
Pinus taeda	28	NA	NA	NA	0.00489 (θ_w)	Neale & Savolainen (2004)
Pinus taeda	18	32	10 116	yes	0.00533	S. C. González-Martínez, CIFOR-INIA, Madrid, pers. comm.
Pinus sylvestris	2	12–15	4 136	yes	0.0007	García-Gil *et al.* (2003)
Pinus sylvestris	1	20	2 045	yes	0.0014	Dvornyk *et al.* (2002)
Cryptomeria japonica	7	48	10 158	yes	0.00252	Kado *et al.* (2003)
Pseudotsuga menziesii	12	NA	NA	NA	0.00853	Neale & Savolainen (2004)
Populus tremula	5	24	6 188	no	0.0111	Ingvarsson (2005)
Quercus petraea	7	27	3 083	yes	0.00722	J. Derory, INRA Pierroton, pers. comm.
Glycine max L. Merr.	142	25	76 000	no, restricted to ancestors of North American cultivars	0.00125	Zhu *et al.* (2003)
Arabidopsis thaliana	9	20	–	yes	0.0067	reviewed in Aguadé (2001)
Beta vulgaris	37	2	18 002	no	0.0076	Schneider *et al.* (2001)
Zea mays	18	36	6 935	no, restricted to US elite maize breeding pool	0.0063	Ching *et al.* (2002)
Zea mays	6	12–25	NA	yes	0.00871	reviewed in White & Doebley (1999)

NA, data not available.

involvement of these genes in the variability of the traits of interest. The availability of haploid tissue enabled the definition of the different haplotypes, allowing a reduction of the polymorphic sites to be genotyped. For instance in *P. pinaster*, only 21 markers (SNPs and INDELs) will have to be genotyped to define the haplotypic composition, instead of the 32 polymorphic sites discovered. This subset of SNP tags was defined exclusively based on the haplotypes observed in the studied sample. Although linkage disequilibrium analysis would allow a reduction of this SNP tag set, such analysis was not performed, given the high differentiation observed for some of the analysed genes and the small sample size of each population analysed.

In spite of the exploratory nature of this study, limited to a restricted set of genes, it is interesting to note that the results obtained here agree with previous nucleotide surveys in conifers (Table 7). Although comparative diversity analyses using allozymes have shown that conifers are among the most genetically diverse organisms (Hamrick & Godt, 1990), nucleotide data do not support this statement. Indeed, the nucleotide diversity of conifers is higher than in humans but lower than in *Zea mays*. Interestingly, the nucleotide diversity levels reported in broadleaved trees such as *Populus* or *Quercus* are also significantly higher than in conifers (Table 7); the reasons for this divergence remain to be found.

Lower diversity in *P. radiata*: consequences of neutral process or genes controlling traits submitted to selection

A trend towards lower nucleotide diversity was observed for *P. radiata* compared with *P. pinaster*. This result is consistent with our previous knowledge regarding the populations analysed. Although *P. pinaster* is characterized by a large geographic distribution, the natural range of *P. radiata* is extremely small. In addition, the populations analysed in this study covered different ranges of the distribution according to the considered species. For *P. pinaster*, almost the whole geographic distribution was analysed, whereas for *P. radiata*, only a subset of the total variation was analysed. As a consequence, the lower nucleotide diversity observed for *P. radiata* agrees with its lower population effective size compared with *P. pinaster*.

Although a lower diversity is expected in *P. radiata* under a neutral model of evolution, two genes (*Pp1* and *CesA3*) presented an abnormally strong reduction of diversity in this species. A plausible hypothesis would be the concomitant effects of the smaller population effective size combined to the existence of natural and/or artificial selection acting on these genes. Such a scenario would have lead to the elimination of some alleles, resulting in an unusually low diversity level. Several concomitant results in *P. pinaster* support this hypothesis.

For *Pp1* in *P. pinaster*, a higher differentiation than at the neutral level was observed. Such a differentiation pattern would be consistent with a 'diversifying selection' acting at this locus in *P. pinaster*. Evidence of selection at the molecular level for this gene would be consistent with its physiological role: *Pp1* is a glycine-rich protein (GRP) that has been shown to be differentially expressed between differentiating xylem associated with different types of wood characterized by different physical and chemical properties; in other words, early vs late wood (Le Provost *et al.*, 2003) and opposite vs compression wood (Allona *et al.*, 1998; Zhang *et al.*, 2000; Le Provost *et al.*, 2003). Cell wall GRPs are localized in vascular

tissues and are thought to provide elasticity as well as tensile strength during vascular development (Cassab, 1998). Polymorphisms inducing variation of these properties would definitely affect the adaptation of the tree to its environmental conditions and thus be preferentially fixed in certain conditions. In the case of *P. radiata*, according to the negative genetic correlations often reported between growth and wood quality in conifers (Rozenberg & Cahalan, 1997; Pot *et al.*, 2002), the reduction of diversity observed may have resulted from the artificial selection on growth applied to the New Zealand land race.

The absence of differentiation observed for *CesA3* in *P. pinaster* compared with the significant level observed for neutral markers (Petit *et al.*, 1995) provides a strong indication of balancing selection acting on this gene. The positive Tajima's *D*-values reported for this gene for the whole area of distribution and for the Aquitaine provenance tend to confirm this hypothesis. Indeed, such values would not be expected in the case of no differentiation. Furthermore, the relatively high haplotype structure observed for this gene (high haplotype diversity, low number of haplotypes compared with the number of polymorphic sites, high level of linkage disequilibrium) also indicates the same tendency toward the action of balancing selection. These hypotheses of possible deviations from neutrality for *CesA3* are consistent with its role in cellulose biosynthesis. Cellulose is one of the major components of the cell wall. In temperate zones, climatic variation during the annual course of the vascular cambium give rise to early wood formed early during the growing season, and late wood formed in late summer. This environmental pressure could strongly affect the major change in cellulose content recognized between these two types of wood.

As in the case of *Pp1*, the reduction of diversity observed for *CesA3* in *P. radiata* would be consistent with the involvement of this gene in the genetic determinism of wood quality, a trait negatively correlated to growth.

KORRIGAN, a gene involved in polysaccharides biosynthesis, as a putative target of natural selection

Several results that include a high differentiation between Corsican and Aquitaine populations in *P. pinaster* and a significant negative Tajima's *D*-value in *P. radiata* suggest *KORRIGAN* as a potential target of selection in these species.

The high differentiation observed in *P. pinaster* is consistent with the existence of diversifying selection that would have lead to the prevalence of different haplotypes, as a consequence of their role in local adaptation to the particular environmental conditions encountered.

In *P. radiata*, the significant negative Tajima's *D*-value may result from a past selection event on this gene, or may be a recent one, which would be consistent with the relatively strong haplotype structure (only four haplotypes for 18 sequences and five polymorphic sites). Thus the excess of rare frequency polymorphisms would be consistent with a hitchhiking event in the *P. radiata* population. An alternative neutral hypothesis would be the recent expansion of the New Zealand breeding population.

The role of *KORRIGAN* is consistent with deviation from neutrality. Indeed, *KORRIGAN* is involved in the biosynthesis of cellulose, the main compound of the cell wall and whose amount is genetically controlled (Zobel & Buijtenen, 1989; Pot *et al.*, 2002; Sewell *et al.*, 2002), and which provides strength and flexibility to plant tissue. It encodes a β-1,4 endoglucanase, which catalyses the cleavage of the cellodextrin from the sistosterol cellodextrin (Nicol *et al.*, 1998; Peng *et al.*, 2002) before the proper synthesis of the cellulose microfibrils by the cellulose synthase complex. Its importance in this pathway has already been underlined. It is indeed strongly differentially expressed between early and late wood, presenting an overexpression in late wood which is characterized by a higher proportion of cellulose (accession AL750476 in Le Provost, 2003).

Recent studies tend to confirm the central role of this gene in the genetic variability of cell wall composition. Indeed, significant relationships between *KORRIGAN* polymorphisms and polysaccharides content were detected (coincident with QTLs in a three-generation outbreed pedigree; Pot, 2004). Also a significant association was observed in the *P. pinaster* first-generation breeding population between one *KORRIGAN* SNP and cellulose content (P. Garnier-Géré, pers. comm.). These observations reveal the potential importance of this gene in the variability of polysaccharide content, a trait that may be subjected to natural selection pressures.

Conclusion and perspectives

This exploratory study allowed the identification of polymorphisms in eight wood formation related genes in *P. pinaster* and *P. radiata*. This information is currently used in association studies to test their involvement in the phenotypic variability of economically important traits linked to wood structure and chemical composition in these two species.

The analysis of the patterns of nucleotide diversity obtained at the intra and interspecific levels provided some indications on adaptative evolution at the molecular level for *KORRIGAN*, *Pp1* and *CesA3*. These interpretations are consistent with the demonstrated physiological role of these genes, and with recent data obtained in QTL mapping experiments and association studies.

Acknowledgements

We gratefully acknowledge the financial support from the French Ministry of Foreign Affairs, who granted a collaborative project « Déterminisme génétique et moléculaire de la qualité du bois chez les conifères » between INRA (Cestas, France) and Forest Research (Rotorua, New Zealand), the European Union (GEMINI: QLRT-1999-00942, FEDER:2003227)

and the Aquitaine Region (2004-03-05-003FA). We thank Drs Valérie Lecorre, Philippe Rozenberg, Santiago C. González-Martínez and Antoine Kremer for their comments on earlier versions of the manuscript.

References

Aguadé M. 2001. Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis Thaliana*. *Molecular Biology and Evolution* **18**: 1–9.

Alazard P, Raffin A. 2003. La troisième génération de vergers à graines, de nouvelles variétés pour 2010. In: Groupe Pin maritime du futur, ed. Le progrès génétique en Forêt. Bordeaux, France: Groupe Pin maritime du futur, 41–48.

Allona I, Quinn M, Shoop E, Swope K, St. Cyr S, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff R, Whetten RW. 1998. Analysis of xylem formation in pine by cDNA sequencing. *Proceeding of the National Academy of Sciences, USA* **95**: 9693–9698.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Arcade A, Faivre-Rampant P, Pâques LE, Prat D. 2002. Localisation of genomic regions controlling microdensitometric parameters of wood characteristics in hybrid larches. *Annals of Forest Science* **59**: 607–615.

Baradat P, Marpeau-Bezard A. 1988. Le pin maritime. *Pinus Pinaster* Ait: biologie et génétique des terpènes pour la connaissance et l'amélioration de l'espèce. PhD Thesis. University of Bordeaux-I, France.

Barber VA, Juday GP, Finney BP. 2000. Reduced growth of Alaskan white spruce in the twentieth century from temperature-induced drought stress. *Nature* **405**: 668–6673.

Bradshaw HD, Stettler RF. 1995. Molecular genetics of growth and development in populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics* **139**: 963–973.

Brown GR, Bassoni DL, Gill GP, Fontana JR, Wheeler NC, Megraw RA, Davis MF, Sewell MM, Tuskan GA, Neale DB. 2003. Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L). III. QTL verification and candidate gene mapping. *Genetics* **164**: 1537–1546.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceeding of the National Academy of Sciences, USA* **101**: 15255–15260.

Burban C, Petit R. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* **12**: 1487–1495.

Burdon RD, Broekhuizen P, Zabkiewicz JA. 1997a. Comparison of native-population and New Zealand land-race samples of *Pinus radiata* using cortical oleoresin monoterpenes. In: Burdon RD, Moore JM, eds. *Proceedings of IUFRO '97: Genetics of Radiata Pine. NZ FRI Bulletin* **203**: 50–56.

Burdon RD, Firth A, Low CB, Miller MA. 1997b. Native populations of *Pinus radiata*. New Zealand. Performance and potential. *New Zealand Journal of Forestry Science* **41**: 32–36.

Cassab GI. 1998. *Plant cell wall proteins. Annals of Review of Plant Physiological Plant Molecular Biology* **49**: 281–309.

Chagné D, Brown G, Lalanne C, Madur D, Pot D, Neale D, Plomion C. 2003. Comparative genome and QTL mapping between maritime and loblolly pines. *Molecular Breeding* **12**: 185–195.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* **3**: 19.

Costa P, Bahrman N, Frigerio JM, Kremer A, Plomion C. 1998. Water-deficit responsive proteins in Maritime pine. *Plant Molecular Biology* **39**: 587–596.

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

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

Fagard M, Höfte H, Vernhettes S. 2000. Cell wall mutants. *Plant Physiology and Biochemistry* **38**: 15–25.

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

Gion J-M, Lalanne C, Le Provost G, Ferry-Dumazet H, Paiva J, Frigerio JM, Chaumeil P, Barré A, de Daruvar A, Brach J, Claverol S, Bonneu M, Plomion C. 2005. The proteome of maritime pine wood forming tissue. *Proteomics*. (In press.)

González-Martínez SC, Alía R, Gil L. 2002. Population genetic structure in a Mediterranean pine (*Pinus pinaster* Ait.): a comparison of allozyme markers and quantitative traits. *Heredity* **89**: 199–206.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* **85**: 862–864.

Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR. 1996. Genetic Mapping of Quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* **144**: 1205–1214.

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

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlén Mathias Teeri TT, Lundeberg J, Sundber B, Nilsson P, Sandberg G. 2001. A transcriptional roadmap to wood formation. *Proceeding of the National Academy of Sciences, USA* **98**: 14732–14737.

Hey J, Wakeley J. 1997. A coalescent estimator of the population recombination rate. *Genetics* **145**: 833–846.

Hudson RR. 1987. Estimating the recombination parameter of a finite population model without selection. *Genetical Research* **50**: 245–250.

Ingvarsson PK. 2005. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics* doi: 10.1534/genetics.104.034959.

Johnson R, Lipow S. 2002. Compatibility of breeding for increased wood production and longterm sustainability: the genetic variation of seed orchard seed and associated risks. In: Johnson, A C, Haynes, R W, Monserud, R A, eds. *Congruent management of multiple resources, Proceedings from the Wood Compatibility Initiative Workshop 18*. General Technical Report PNW-GTR-563. Portland, OR, USA: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, 169–179.

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

Karhu A. 2001. Evolution and applications of pine microsatellites. PhD Thesis. University of Oulu, Finland.

Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge, MA, USA: Cambridge University Press.

Kumar S, Spelman RJ, Garrick DJ, Richardson TE, Lausberg M, Wilcox PL. 2000. Multiple-marker mapping of wood density loci in an outbred pedigree of radiata pine. *Theoretical and Applied Genetics* **100**: 926–933.

Le Provost G. 2003. Effet de la saison et d'un stress mécanique sur la variation du transcriptome dans le xylème en formation chez le pin maritime (*Pinus Pinaster* Ait.) PhD Thesis. University of Bordeaux-I, France.

Le Provost G, Paiva J, Pot D, Brach J, Plomion C. 2003. Seasonal variation in transcript accumulation in wood forming tissues of maritime pine (*Pinus pinaster* Ait.) with emphasis on a Cell wall Glycine Rich Protein. *Planta* **217**: 820–830.

Lerceteau EC, Plomion C, Andersson B. 2000. AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*: a preliminary study. *Molecular Breeding* **6**: 451–458.

Liphschitz N, Waisel Y. 1970. Effects of environment on relations between extension and cambial growth of *Populus euphratica* Oliv. *New Phytologist* **69**: 1064.

Long JC, Williams RC, Urbanek M. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. *American Journal of Human Genetics* **56**: 799–810.

Mangin B, Goffinet B, Rebai A. 1994. Constructing confidence intervals for QTL location. *Genetics* **138**: 1301–1308.

Mariette S, Chagné D, Lézier C, Pastuszka P, Raffin A, Plomion C, Kremer A. 2001. Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. *Heredity* **86**: 469–479.

Markussen T, Fladung M, Achere V, Favre JM, Faivre-Rampant P, Aragones A, Da Silva Perez Harvengt L, Ritter E. 2003. Identification of QTLs controlling growth, chemical and physical wood property traits in *Pinus pinaster* (Ait). *Silvae Genetica* **52**: 8–15.

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

Moran GF, Thamarus KA, Raymond CA, Qiu D, Uren T, Southerton SG. 2002. Genomics of Eucalyptus wood traits. *Annals of Forest Science* **59**: 645–650.

Mouille G, Robin S, Lecomte M, Pagant S, Höfte H. 2003. Classification and identification of Arabidopsis cell wall mutants using Fourier Transform InfraRed (FT-IR) microspectroscopy. *Plant Journal* **35**: 393–404.

Neale DB, Savolainen O. 2004. Conifers as a model species for complex trait dissection. *Trends in Plant Science* **9**: 325–330.

Neale DB, Sewell MM, Brown G. 2002. Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. *Annals of Forest Science* **5**: 595–605.

Nei M. 1987. *Molecular Evolutionary Genetics*. New York, NY, USA: Columbia University Press.

Nicol F, His I, Jauneau A, Vernhettes S, Canut H, Höfte H. 1998. A Plasma membrane-bound putative endo-1,4-B-D-glucanase is required for normal wall assembly and cell elongation in Arabidopis. *EMBO Journal* **17**: 5563–5576.

Peng L, Kawagoe Y, Hogan P, Delmer D. 2002. Sistostreol-β-glucoside as a primer for cellulose synthesis in plants. *Science* **295**: 147–150.

Petit RJ, Bahrman N, Baradat P. 1995. Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. *Heredity* **75**: 382–389.

Plomion C, Bahrman N, Durel CE, O'Malley DM. 1995. Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. *Heredity* **74**: 661–668.

Plomion C, Leprovost G, Stokes A. 2001. Wood Formation in Trees. *Plant Physiology* **127**: 1513–1523.

Plomion C, Pionneau C, Brach J, Costa P, Baillère H. 2000. Compression wood responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). *Plant Physiology* **123**: 959–969.

Pot D. 2004. Déterminisme génétique de la qualité du bois chez le pin maritime: du phénotype aux gènes. PhD Thesis. ENSA Rennes, France.

Pot D, Chantre G, Rozenberg P, Rodrigues JC, Jones Gwynn L, Pereira H, Hannrup B, Cahalan C, Plomion C. 2002. Genetic control of pulp and timber properties in maritime pine (*Pinus pinaster* Ait). *Annals of Forest Science* **59**: 563–575.

Ribeiro MM, Mariette S, Vendramin GG, Szmidt AE, Plomion C, Kremer A. 2002. Comparison of genetic diversity estimates within and among populations of maritime pine using chloroplast simple-sequence repeat and amplified fragment length polymorphism data. *Molecular Ecology* **11**: 869–877.

Riccardi F, Gazeau P, de Vienne D, Zivy M. 1998. Protein Changes in Response to Progressive Water Deficit in Maize: Quantitative Variation and Polypeptide Identification. *Plant Physiology* **117**: 1253–1263.

Rozas J, Rozas R. 1999. Dnasp, version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.

Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawest, S, Misener, S, eds. *Bioinformatic Methods and Protocols: Methods in Molecular Biology*. Totowa, NJ, USA: Human Press, 365–386.

Rozenberg P, Cahalan C. 1997. Spruce and wood quality: genetic aspects (a review). *Silvae Genetica* **46**: 270–279.

Rozenberg P, Van Loo J, Hannrup B, Grabner M. 2002. Clonal variation of wood density record of cambium reaction to water deficit in Picea abies (L.) Karst. *Annals of Forest Science* **59**: 533–540.

Schneider K, Weisshaar B, Borchardt DC, Salamini F. 2001. SNP frequency and allelic haplotype structure of Beta vulgaris expressed genes. *Molecular Breeding* **8**: 63–74.

Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL, Neale DB. 2002. Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). II. Chemical wood properties. *Theoretical Applied Genetics* **104**: 214–222.

Shneider S, Ropessli D, Excoffier L. 2000. Arlequin ver 2000: A software for population genetics data analysis. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.

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

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ESIV. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* **28**: 286–289.

Wakamiya I, Newton RJ, Johnston JS, Price HJ. 1993. Genome size and environmental factors in the genus *Pinus*. *American Journal of Botany* **80**: 1235–1241.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256–276.

Whetten R, MacKay JJ, Sederoff RR. 1998. Recent advances in understanding lignin biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**: 585–609.

White SE, Doebley JF. 1999. The molecular evolution of terminal ear 1, a regulatory gene in the genius Zea. *Genetics* **153**: 1455–1462.

Wu J, Krutovskii KV, Strauss SH. 1999. Nuclear DNA diversity, population differentiation, and phylogenetic relationship in the Californian closed-cone pines based on RAPD and allozyme markers. *Genome* **42**: 893–908.

Zhang Y, Sederoff RR, Allona I. 2000. Differential expression of genes encoding cell wall proteins in vascular tissues from vertical and bent pine trees. *Tree Physiology* **20**: 457–466.

Zhu YL, Song QJ, Hyten DL, Van Tassel CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Gregan PB. 2003. Single nucleotide polymorphisms in soybean. *Genetics* **163**: 1123–1134.

Zobel BJ, van Buijtenen JP. 1989. Wood Variation. *Springer Series in Wood Science*. New York, NY, USA: Springer-Verlag.

About *New Phytologist*

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at **www.newphytologist.org**.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *OnlineEarly* – the 2004 average submission to decision time was just 30 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £109 in Europe/$202 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office (**newphytol@lancaster.ac.uk**; tel +44 1524 592918) or, for a local contact in North America, the US Office (**newphytol@ornl.gov**; tel +1 865 576 5261).

