Theor Appl Genet (2004) 109: 1204–1214
DOI 10.1007/s00122-004-1683-z

ORIGINAL PAPER

D. Chagné · P. Chaumeil · A. Ramboer · C. Collada · A. Guevara · M. T. Cervera · G. G. Vendramin · V. Garcia · J-M Frigerio · C. Echt · T. Richardson · C. Plomion

Cross-species transferability and mapping of genomic and cDNA SSRs in pines

Received: 3 December 2004 / Accepted: 30 March 2004 / Published online: 22 September 2004
© Springer-Verlag 2004

Abstract Two unigene datasets of *Pinus taeda* and *Pinus pinaster* were screened to detect di-, tri- and tetranucleotide repeated motifs using the *SSRIT* script. A total of 419 simple sequence repeats (SSRs) were identified, from which only 12.8% overlapped between the two sets. The position of the SSRs within their coding sequences were predicted using *FrameD*. Trinucleotides appeared to be the most abundant repeated motif (63 and 51% in *P. taeda* and *P. pinaster*, respectively) and tended to be found within translated regions (76% in both species), whereas dinucleotide repeats were preferentially found within the 5′- and 3′-untranslated regions (75 and 65%, respectively). Fifty-three primer pairs amplifying a single PCR fragment in the source species (mainly *P. taeda*), were tested for amplification in six other pine species. The amplification rate with other pine species was high and corresponded with the phylogenetic distance between species, varying from 64.6% in *P. canariensis* to 94.2% in *P. radiata*. Genomic SSRs were found to be less transferable; 58 of the 107 primer pairs (i.e., 54%) derived from *P. radiata* amplified a single fragment in *P. pinaster*. Nine cDNA-SSRs were located to their chromosomes in two *P. pinaster* linkage maps. The level of polymorphism of these cDNA-SSRs was compared to that of previously and newly developed genomic-SSRs. Overall, genomic SSRs tend to perform better in terms of heterozygosity and number of alleles. This study suggests that useful SSR markers can be developed from pine ESTs.

Communicated by D.B. Neale

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-004-1683-z

D. Chagné · P. Chaumeil · A. Ramboer · J.-M. Frigerio · C. Plomion (✉)
UMR 1202 BIOGECO-INRA,
Equipe de génétique, 69 route d'Arcachon,
33610 Cestas Cédex, France
e-mail: plomion@pierroton.inra.fr
Tel.: +33-557-122838
Fax: +33-557-122881

C. Collada
Departamento de Biotecnología, Escuela Técnica Superior de Ingenieros de Montes de Madrid, UPM, Ciudad Universitaria sn,
28040 Madrid, Spain

A. Guevara · M. T. Cervera
Departamento de Genética Forestal, CIFOR-INIA,
Carretera de la Coruna Km 7,
28040 Madrid, Spain

G. G. Vendramin
Istituto di Genetica Vegetale, Sezione di Firenze, Consiglio Nazionale delle Ricerche,
Via Madonna del Piano,
50019, Sesto Fiorentino,
Firenze, Italy

V. Garcia
UMR Physiologie et Biotechnologie Végétale, INRA Bordeaux,
71 avenue E. Bourleaux,
33883 Villenave d'Ornon Cédex, France

C. Echt · T. Richardson
New Zealand Forest Research Institute Ltd,
Private Bag 3020 Rotorua, New Zealand

Introduction

In contrast to other plant species, few polymorphic single-copy nuclear microsatellite markers or simple sequence repeats (SSR) have been reported in the Pinaceae (reviewed in Table1). The genome structure of these species, characterised by a large physical size (22 pg/C, Leitch et al. 2001) with a large amount of repeated sequence (Kriebel 1985; Kamm et al. 1996; Kossack and Kinlaw 1999; Elsik and Williams 2000) has been the main obstacle to the development of useful markers. In addition, the ancient divergence time between coniferous species (Price et al. 1998) and the complexity of their genomes means that transferability of single-copy SSRs among genera and even within *Pinus* (the most studied genus) is generally poor, resulting in a large proportion of amplification failure, non-specific amplification, multi-banding

patterns or lack of polymorphism (Echt et al. 1999; Mariette et al. 2001). Given the high cost of developing useful SSR markers, cross-species transferability is a valuable attribute.

In an attempt to circumvent these genome-related problems, Elsik and Williams (2001) removed most of the repetitive portion of the genome using a DNA reassociation kinetics-based method, and Zhou et al. (2002) targeted the low-copy portion of the genome using an undermethylated region enrichment method. Both approaches yielded remarkable enrichment for useful SSR markers in *Pinus taeda*. Scotti et al. (2002a, b) used an alternative strategy based on the pre-screening of single-copy microsatellite containing clones, using dot blot hybridisation analysis, and also obtained a high number of single-copy polymorphic SSR markers in *Picea abies*. *Pinus taeda* SSRs developed by Elsik and Williams (2001) and Zhou et al. (2002) transferred quite well between American hard pines (Shepherd et al. 2002), but were shown to be less transferable in the phylogenetically divergent Mediterranean hard pines (Gonzalez-Martinez et al. 2004). Interestingly, perfect trinucleotide SSRs transferred from American to Mediterranean pines better than other motifs (Kutil and Williams 2001).

Simple sequence repeats have been found in all genomic regions, including coding regions (Toth et al. 2000). By developing a cDNA library enriched in SSRs, Scotti et al. (2000) showed the presence of microsatellites within the coding regions of Norway spruce (*Picea abies*), a species belonging to the Pinaceae. The availability of expressed sequence tags (ESTs) resulting from large sequencing projects is potentially a valuable source of SSRs that can be evaluated with less intensive laboratory development. Recently, cDNA-SSRs were obtained from EST databases developed in several plant species such as grape (Scott et al. 2000), cereals (Temnykh et al. 2000, 2001; Cho et al. 2000; Cordeiro et al. 2001; Kantety et al. 2002; Eujayl et al. 2002; Varshney et al. 2002; Gao et al. 2003) and *Arabidopsis* (Cardle et al. 2000; Morgante et al. 2002). These EST-derived markers showed good transferability between phylogenetically related species (Eujayl et al. 2003; Gupta et al. 2003).

The objectives of this study were threefold: (1) to investigate the relative occurrence and types of SSRs present in the coding regions of two pine genomes, (2) compare polymorphism levels of SSRs derived from cDNA and genomic sources, and (3) compare the transferability of cDNA-SSRs and genomic SSR markers across several pine species.

Materials and methods

In silico SSR detection in pine ESTs

Public EST database were independently assembled for *Pinus pinaster* and *P. taeda* using StackPack (Christoffels et al. 2001). A total of 18,498 *P. pinaster* ESTs provided 2,893 contigs and 5,001 singletons (http://cbi.labri.fr/outils/SAM/COMPLETE/index.php). For *P. taeda*, 8,070 contigs and 12,307 singletons resulted from 75,047 ESTs (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/).

Pinus pinaster and *P. taeda* unigene sets were searched for tandemly repeated motifs of 2, 3 and 4 bp using the *SSRIT* SSR search tool (Temnykh et al. 2001; http://www.gramene.org/db/searches/ssrtool), with 14, 15 and 20 as the minimum repeat length, respectively. We associated the *SSRIT* Perl script with the *FrameD* gene prediction software (Schiex et al. 2003) to determine if the detected repeat motifs were located in the 5′ or 3′ untranslated regions (UTRs) or in the open reading frames (ORF). *FrameD* was developed to predict the position of the translated regions in EST sequences. Because *FrameD* uses interpolated Markov models (IMM; Salzberg et al. 1998) to build probabilistic models of coding sequences, a pine-specific IMM was constructed to enhance the prediction in *P. taeda* and *P. pinaster* sequences. We used 67 kb from 65 pine full-length coding sequences to build the *Pinus* IMM (Table S2). Finally, the sequences containing microsatellites in *P. pinaster* and *P. taeda* were compared in order to check the redundancy of the sequences containing SSRs in both species.

PCR primer design and amplification

We designed 56 PCR primer pairs (set no. 1) flanking the microsatellites identified with our in silico analysis using Primer v3.0 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) with default parameters, except that we used a range of 40–55% for the primer GC%, GC clamps of 2 bases and a maximum T_m difference of 10. We kept the expected amplified fragment length below 500 bp to avoid the risk of the presence of introns, which may induce PCR failure. Fifty-three out of 56 PCR primers were designed based on *P. taeda* sequences and three were developed from *P. pinaster* sequences. The PCR primers were chosen to represent the broadest range of SSRs possible considering the repeat type (di-, tri- or tetranucleotide), the motif (e.g., AG, AT), the length (5–26 repeats) and the position (UTR or ORF). In addition to these new SSRs, we also included a set of 16 cDNA-SSRs previously developed from *P. taeda* sequences (set no. 2, C. Echt, http://dendrome.ucdavis.edu/Gen_res.htm). This second set resulted from a SSR search using a preliminary sequence dataset of about 10,000 *P. taeda* ESTs.

A third set of 107 PCR primers (set no. 3) was developed from *P. radiata* genomic SSRs and screened for amplification success in pine species (C. Echt and T. Richardson, unpublished data). A fourth set of three SSR markers described by Mariette et al. (2001) was also used (set no. 4).

DNA was isolated using the protocol described by Doyle and Doyle (1990). PCR reactions were performed with 15 ng of genomic DNA in a total reaction volume of 10 µl, with 1× reaction buffer (Gibco BRL), 2 mM $MgCl_2$, 1 µM of each primer, 0.2 mM of dNTP and 0.5 U of *Taq*

polymerase (Gibco BRL) on a Stratagene Robocycler Gradient 96 (Stratagene, La Jolla, Calif., USA) using the following cycles: preliminary denaturing (94°C, 5 min) followed by 30 cycles of denaturing (94°C, 30 s), annealing (locus-specific temperature, 30 s), and extension (72°C, 1 min), and a final extension (72°C, 10 min). An additional touchdown was performed for some loci (10 cycles with the annealing temperature decreasing by 1°C for every cycle).

Amplification success was checked on 1.5% agarose gels. We checked that the amplification showed a single band pattern with a size corresponding to the expected length. Amplifications resulting in multiple bands were discarded from further analysis since they could result from non-specific amplification or paralogous loci. The useful loci were then run on a LICOR automated sequencer using the same conditions described by Mariette

Table 1 Di-, tri- and tetranucleotide SSR detection in *Pinus pinaster* and *P. taeda* unigenes using *SSRIT* software

Development of SSR loci					Transfer of SSR loci		
Species	SSR origin	Number of primer pairs tested	Number of polymorphic single copy SSR loci	Reference	Number of markers tested in other conifers	Number of successfully transferred markers	Reference
Pinus radiata $^{H-am}$	EGL	2	2	Smith and Devey 1994	2	2^{H} 0^{S} 0^{P}	Echt et al. 1999
					2	1^{H-me}	Karhu et al. 2000
					2	0^{H-me}	Mariette et al. 2001
					2	0^{H-am}	Shepherd et al. 2002
Pinus radiata $^{H-am}$	TGL	43*	2	Fisher et al. 1998			
	EGL		11		7	7^{H} 3^{S} 1^{P}	Fisher et al. 1998
					4	3^{H-me} 3^{H-am} 0^{S} 0^{P}	Echt et al. 1999
					2	0^{H-me}	Mariette et al. 2001
					7	4^{H-am}	Shepherd et al. 2002
					20	11^{H-am}	Devey at al. 1999
Pinus radiata $^{H-am}$	EGL	50	10	Devey et al. 2002			
Pinus taeda $^{H-am}$	ELCL	18	16	Elsik et al. 2000b	7	7^{H-am} 5^{H-me}	Kutil and Williams 2001
					25**	13^{H-am}	Shepherd et al. 2002
					19**	10^{H-me}	Gonzalez-Martinez et al. 2003
Pinus taeda $^{H-am}$	ELCL	29	15	Elsik and Williams 2001			
	EGL	37	8				
Pinus taeda $^{H-am}$	ELCL	8	8	Kutil and Williams 2001	8	8^{H-am} 2^{H-me}	Kutil and Williams 2001
Pinus taeda $^{H-am}$	EUML	36	19	Zhou et al. 2002			
Pinus contorta $^{H-am}$	EGL	5	5	Hicks et al. 1998			
Pinus sylvestris $^{H-me}$	TGL	2	0	Kostia et al. 1995			
Pinus sylvestris $^{H-me}$	EGL	37	7	Soranzo et al. 1998	3	3^{H-me}	Gonzalez-Martinez et al. 2003
Pinus halepensis $^{H-me}$	EGL	25	8	Keys et al. 2000	8	7^{H-me}	Keys et al. 2000
					8	1^{H-me}	Mariette et al. 2001
Pinus pinaster $^{H-me}$	EGL	29	2	Mariette et al. 2001			
Pinus densiflora $^{H-as}$	EGL	14	6	Lian et al. 2000	6	6^{H-as} 5^{H-am} 0^{S}	Lian et al. 2000
Pinus strobus ^{S}	EGL	77	19	Echt et al. 1996	15	12^{S} 0^{H} 0^{P}	Echt et al. 1999
					28	3^{H}	Karhu et al. 2000
					4	0^{H}	Mariette et al. 2001
					5	0^{H}	Shepherd et a. 2002
Pinus strobus ^{S}	EGL	4	0	Echt et al. 1999	4	0^{S}	Echt et al. 1999
Picea sitchensis ^{SP}	EGL	7	4	van de Ven and Mac Nicol 1996			
Picea abies ^{SP}	EGL	36	7	Pfeiffer et al. 1997			
Picea abies ^{SP}	EGL	96	34	Paglia et al. 1998			
Picea abies ^{SP}	ECDL	6	6	Scotti et al. 2000			
Picea abies ^{SP}	EGL	55	16	Scotti et al. 2002a (tri)			
Picea abies ^{SP}	EGL	53	16	Scotti et al. 2002b			
Picea glauca ^{SP}	EGL	13	13	Hodgetts et al. 2001	13	12^{SP}	Hodgetts et al. 2001
Picea glauca ^{SP}	EGL	16	6	Rajora et al. 2001	6	6^{SP}	Rajora et al. 2001
Pseudotsuga menziesii ^{P}	EGL	102	48	Amarasinghe and Carlson 2002	50	31^{P}	Amarasinghe and Carlson 2002
Tsuga heterophylla ^{P}	EGL	16	11	Amarasinghe et al. 2003			
Cryptomeria japonica ^{P}	EST	3	2	Moriguchi et al. 2003			
	EGL	67*	31				
	TGL		1				
Total	-	776	333 (43%)		213	108 (50%)	

et al. (2001) to precisely determine the length of each amplification product (i.e., allele).

Sequencing

Amplified fragments in *P. pinaster* were cloned and sequenced as described by Dubos and Plomion (2003) in order to check the orthology of the same markers as based on sequence identity.

Plant material

Polymorphism and reliable co-dominant inheritance were tested in three *P. pinaster* mapping pedigrees (the INRA-F2 pedigree, Costa et al. 2000; the INRA-G2 pedigree Chagné et al. 2002; and the AFOCEL-F1 pedigree, Ritter et al. 2002) for which saturated genetic maps are available, and a fourth (INIA-F1) which is under construction (M.T. Cervera, unpublished data). Loci that were polymorphic in at least one mapping pedigree were also tested on 26 unrelated *P. pinaster* elite trees from the Aquitaine region (south-western France). These trees are first generation selections for the *P. pinaster* breeding programme and were used to estimate the level of diversity (heterozygosity and number of alleles) of the SSRs.

Samples from seven species belonging to the genus *Pinus* (subgenus *Pinus*): *P. canariensis*, *P. halepensis*, *P. pinaster*, *P. pinea*, *P. radiata*, *P. sylvestris*, and *P. taeda* were used to test the amplification rate of the cDNA-SSR markers.

Mapping

Markers segregating in the INRA-G2 and INRA-F2 mapping pedigree were visually scored and assigned two allele genotypes. We used Joinmap v3.0 (Van Oijen and Voorrips 2001) using a minimum LOD of 6.0 for genetic map construction. The Arlequin software (Schneider et al. 2000) was used to estimate genetic diversity parameters based on the genotypes of the 26 unrelated *P. pinaster* individuals.

Results

SSR detection in pine ESTs and sequence annotation

A total of 251 and 168 SSRs were found in *P. taeda* and *P. pinaster* unigene sets (Table S3). This corresponds to enrichment rates of 1.2 and 2.1%, respectively (Table 1). The most common repeat types were trinucleotides (63% in *P. taeda* and 51% in *P. pinaster*), followed by dinucleotides (36% in *P. taeda* and 45% in *P. pinaster*). Tetranucleotide repeats were almost absent (1% in *P. taeda* and 3% in *P. pinaster*). These results were obtained for a minimum repeat number of 7, 5 and 5 for di-, tri- and tetranucleotide motifs, respectively. These thresholds are comparable to those used by Cardle et al. (2000) and Scott et al. (2000), and correspond to perfect motifs only. If we used less stringent detection criteria (e.g., minimum of 5 repeats for dinucleotides, as in Morgante et al. 2002) and allowed the detection of compound motifs we have estimated that the SSR enrichment would increase by twofold.

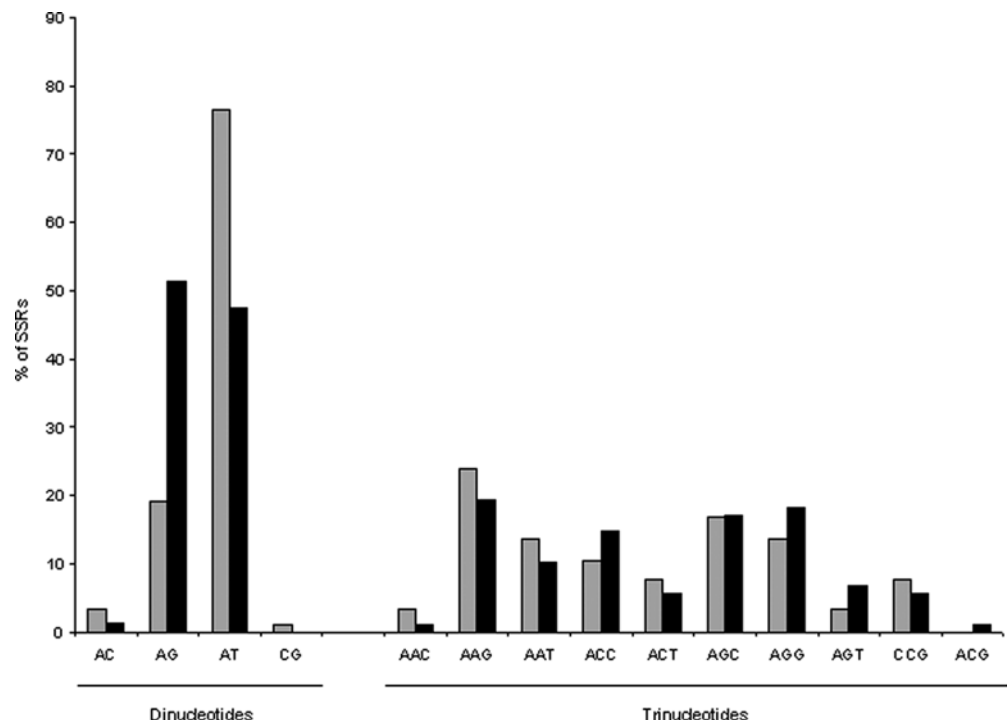

Fig. 1 Distribution of the different classes of di- and trinucleotide SSRs in *Pinus taeda* (*grey boxes*) and *P. pinaster* (*black boxes*) unigenes

Regarding the types of repeated motif (Fig. 1), the AT and AG motifs were the most represented among the dinucleotides (76 and 19% in *P. taeda*, and 47 and 51% in *P. pinaster*, respectively), whereas the AC and CG types were rare (<3% in both species). Regarding trinucleotides, the AAG motif was the most common repeat type (23.9 and 19.3% in *P. taeda* and *P. pinaster*, respectively), followed by AGC and AGG motifs.

Figure 2 shows the position of the detected SSRs in the gene sequences of both species based on the results obtained with *FrameD* (Schiex et al. 2003). Significant differences between di- and trinucleotide SSRs were observed. Dinucleotides were found mostly in the UTRs (75 and 65% in *P. taeda* and *P. pinaster*, respectively), whereas trinucleotides were more frequent in the ORFs (76% in both species). For both type of repeats, SSRs were less abundant in the 5′ UTR than in the 3′ UTR.

By assembling the *P. taeda* and *P. pinaster* contigs and singletons that contained SSRs using StackPack (Christoffels et al. 2001), we found that only 22 of the 171 (12.8%) *P. pinaster* sequences matched contig sequences in the *P. taeda* unigene set, providing a catalogue of 397 non-redundant putative SSR markers for pines.

Transferability of cDNA and genomic SSRs in pines

As a representative sample, 72 primer pairs (sets no. 1 and 2) were designed from cDNA-SSR sequences. Fifty-two out of the 69 *P. taeda* and one out of the three *P. pinaster* cDNA-SSRs amplified a single band of the expected size in the source species. The multi-banding pattern observed for five loci could be attributed to non-specific amplifications or the presence of multi-gene families that are frequent in pines (Kinlaw and Neale 1997). The lack of amplification obtained for 14 loci, could be explained by the quality of the primer pairs and/or the presence of introns. Table 2 summarises the amplification success for

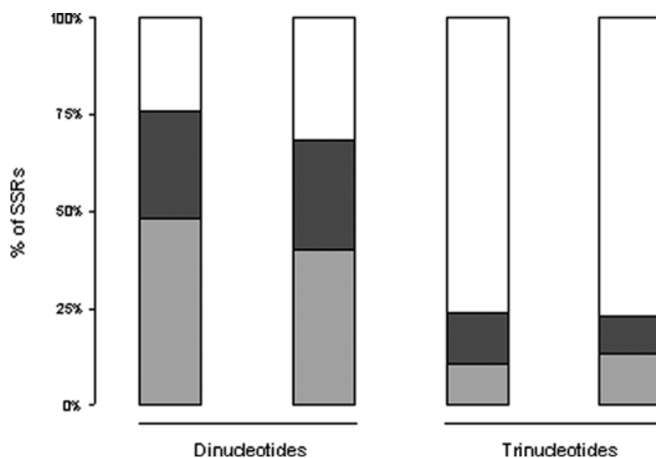

Fig. 2 Distribution of the di- and trinucleotide SSRs within the open reading frame (ORF, in *white*) or in the 5′ untranslated regions (UTR, *dark grey*) and 3′ UTR (*light grey*) in *P. taeda* and *P. pinaster* contigs. Sequences for which no ORF could be detected were not considered

these 53 cDNA-SSR markers in seven pine species. Overall, the amplification rates in non-source species ranged between 64.6% in *P. canariensis* and 94.2% in *P. radiata*. This transferability rate was comparable to the result obtained with EST-derived markers in pines (Brown et al. 2001; Chagné et al. 2003; Komulainen et al. 2003).

Fifty-eight out of 107 (54%) of the set no. 3 *P. radiata* SSR markers amplified a single band in *P. pinaster*. This transferability rate was higher than of Gonzalez-Martinez et al. (2004) in *P. pinaster* using *P. taeda*-derived SSRs (42%), and that of Shepherd et al. (2002) in *P. elliottii* and *P. caribaea* using *P. radiata*-derived SSRs (44%). Overall, the interspecific transferability of cDNA-SSR markers was higher than that of the genomic SSRs.

Polymorphism, orthology, and genetic mapping of cDNA and genomic SSRs in *Pinus pinaster*

Among the 46 single-copy cDNA and 58 genomic SSR loci that amplified in *P. pinaster*, nine (19.5%) and seven (12%) were found to be polymorphic in at least one of the four mapping pedigrees, respectively. Six out of 18 (33%) of the cDNA-SSRs located in UTRs were polymorphic, compared to three out of 30 (10%) of those located in ORFs. This result suggests that a pre-annotation of the sequences containing SSRs can be used to enrich for primer pairs that yield polymorphic cDNA-SSR markers. If we consider the repeat type and position of the cDNA-SSRs (Table 2), then it should be noted that five out of 17 dinucleotide cDNA-SSRs (29%) were polymorphic in at least one *P. pinaster* mapping pedigree whereas four out of 35 (11%) trinucleotide cDNA-SSRs were polymorphic.

We verified the orthology for the seven polymorphic SSR loci originated from *P. radiata* genomic library by sequencing PCR products obtained by amplifying *P. pinaster* DNA. The high levels of sequence identity found for six of the loci (Table 3) were comparable to the levels found between orthologous pine ESTs in previous studies (Brown et al. 2001; Chagné et al. 2003; Komulainen et al. 2003). Interestingly, one locus (NZPR1702_b) was not homologous between the species and did not contain an SSR motif. Electrophoresis on an acrylamide gel showed that this locus presented two distinct bands, 30 bp apart (i.e., two alleles corresponding to an insertion-deletion polymorphism). This locus presented the lowest genetic diversity (H=0.38), and was subsequently discarded for the comparison between genomic and cDNA SSRs (see next section).

The chromosomal assignments of 19 polymorphic SSR markers in the INRA-G2 and INRA-F2 genetic maps (Chagné et al. 2002; Costa et al. 2000) and their polymorphism state in two other *P. pinaster* pedigrees are presented in Table 4. All the loci were linked with a minimum LOD of 6.0, except for locus ssrPt_ctg275 that was not linked to any linkage group in either of the maps. The three SSR markers of set No. 4 previously developed by Mariette et al. (2001) were also mapped in both pedigrees. Overall, these SSRs made it possible to align

Table 2 Cross-specific amplification of 53 cDNA-SSR markers: locus ID and amplification in seven hard pine species. The locus nomenclature follows the recommendations of the Treegenes database (http://dendrome.ucdavis.edu/Tree_Page.htm) for pine STS, also described by Brown et al. (2001). Position in the gene: *UTR* untranslated region, *ORF* open reading frame, *NP* no protein. The annealing temperature (°C) or touchdown temperature range used for the PCR amplification are given. Amplification: *Pp Pinus pinaster* (subsection Sylvestres), *Pt Pinus taeda* (subsection Australes), *Pr Pinus radiata* (subsection Oocarpae), *Ps Pinus sylvestris* (subsection Sylvestres), *Ph Pinus halepensis* (subsection Sylvestres), *Ppi Pinus pinea* (subsection Pineae), *Pc Pinus canariensis* (subsection Canarienses), + single locus amplification, − no amplification, *NA* no data

Primer set	Locus information									Amplification						
	Locus name	Identification	Repeated motif	Number of repeat	Position in gene	Forward primer	Reverse primer	Annealing temperature	Expected length (bp)	*Pp*	*Pt*	*Pr*	*Ps*	*Ph*	*Ppi*	*Pc*
1	SsPp_cn524	Contig524[c]	AG	14	5′UTR	cgattgtttttgccttttaagc	aaatatggcggggtgtgc	50	156	+	+	+	+	+	+	−
1	SsrPt_AA739797	AA739797[b]	AT	11	3′UTR	actttgcggtgaatcagacc	aaagtaaggctgcttgcatga	51	281	+	+	+	−	−	−	−
1	SsrPt_AW010960	AW010960[b]	AT	9	ORF	atcgactaggcatcaggtgg	tcctcgtagcccagctttta	49	225	+	+	+	+	+	+	+
1	SsrPt_AW225917	AW225917[b]	AT	9	3′UTR	tgcattgaaaaatacagcgg	attatgtacgaggccccaca	49	198	+	+	+	+	+	+	+
1	SsrPt_AW981642	AW981642[b]	AAG	7	ORF	gtggcacagggttttctgat	caaaccttcggtagcctcat	60–50	245	−	+	+	NA	NA	NA	NA
1	SsrPt_AW981772	AW981772[b]	CCT	4	ORF	gatcctgttcctcctcctcc	cctggacagaaacagcaaca	49	266	+	+	+	+	+	+	+
1	SsrPt_BF049767	BF049767[b]	AG	22	ORF	ttttgggtcgtaggaacctg	taaaacgggtgtctcttcgg	51	227	+	+	+	+	+	+	−
1	SsrPt_BF778306	BF778306[b]	AG	7	NP	gaagatggagacgaagcagg	tttgcagtctgttgcctttg	60–50	172	−	+	+	NA	NA	NA	NA
1	SsrPt_ctg1376	Contig1376[a]	AT	20	NP	cgatattatggattttgcttgtga	aaatgcatgccaaacttaaatac	60–50	145	+	+	+	+	−	−	−
1	SsrPt_ctg1525	Contig1525[a]	AGG	7	ORF	ttgaaaccatataagcaatgcc	aggacctgggtaaggaggc	60–50	173	+	+	+	+	+	+	+
1	SsrPt_ctg16480	Contig16480[a]	AAAT	13	NP	ctaaaacatcggtcggaagc	atttagtccaggccatgtcg	60–50	151	+	+	NA	NA	NA	NA	NA
1	SsrPt_ctg16811	Contig16811[a]	AT	11	5′UTR	gtccatgatgttgcagattgg	tgttccccaatggtctgtc	56	199	+	+	+	−	−	−	+
1	SsrPt_ctg17601	Contig17607[a]	AAG	9	ORF	cgccattaatatgcctaccg	atctctgcgctgcttgaagt	54	225	+	+	+	+	+	+	+
1	SsrPt_ctg18103	Contig18103[a]	AT	10	NP	cctggattcatttgtggctaa	catgccaacttcttgcattg	60	184	+	+	−	+	+	+	+
1	SsrPt_ctg2300	Contig2300[a]	CCG	6	ORF	cactttgcgagagactgcac	acgctgaaggaaatcgagaa	49	173	+	+	+	+	+	+	+
1	SsrPt_ctg275	Contig275[a]	AT	16	3′UTR	acggagatatattgctggcg	aaagaataacgtgaaacaaaccc	60–50	137	+	+	+	+	−	−	−
1	SsrPt_ctg3021	Contig3021[a]	AGC	14	ORF	ctcagattcctccaaatgcg	catgcaacatatgcaaaccg	60–50	234	+	+	+	+	+	+	+
1	SsrPt_ctg3089	Contig3089[a]	AT	17	NP	ctttcttcacgttggacttctt	ttagccatggagagtgcaga	45	482	−	+	+	+	+	+	+
1	SsrPt_ctg3754	Contig3754[a]	AGC	6	5′UTR	tctttgggtttctggagtgg	gctgttgctgttgttcttgg	60–50	421	+	+	+	+	+	+	+
1	SsrPt_ctg4363	Contig4363[a]	AT	10	3′UTR	taataattcaagccaccccg	agcaggctaataacaacacgc	60–50	100	+	+	+	+	+	+	+
1	SsrPt_ctg4487a	Contig4487[a]	CCG	5	ORF	tctgctgtgtggacaaacct	ttcttggctcaaaatctcgg	60–50	155	+	+	+	+	+	−	−
1	SsrPt_ctg4487b	Contig4487[a]	CCG	10	3′UTR	atgacgcattatcaggggaa	ttgcacagaaagcaggtttg	45	254	+	+	+	+	+	−	−
1	SsrPt_ctg4698	Contig4698[a]	ATC	10	ORF	cgaaaaggtggttctgatgg	ttttccgctggatttaccac	49	246	+	+	+	+	+	+	+
1	SsrPt_ctg5167	Contig5167[a]	AAC	7	ORF	tgcagagagattcgatggg	attttggtttgtttgctggc	60–50	293	+	+	+	+	+	+	+
1	SsrPt_ctg5333	Contig5333[a]	AGC	7	ORF	gaaggagtcggcgataacag	gggaattcgacctgtgaaga	49	163	+	+	+	+	−	−	−
1	SsrPt_ctg6390	Contig6390[a]	AAG	8	5′UTR	atccacgacttgtcgacgc	atcaaccaacttaggcagcg	45	440	−	+	+	−	+	+	+
1	SsrPt_ctg64	Contig64[a]	CCG	7	ORF	ggaagctgttacaagtgcgg	atcgagaagagaggaagggc	60–50	284	+	+	+	+	+	+	+
1	SsrPt_ctg7024	Contig7024[a]	AAG	7	ORF	gggaattctgaaagacaaggg	aacttacccatcgagagcccc	60–50	277	+	+	+	+	+	+	−
1	SsrPt_ctg7081	Contig7081[a]	AAG	7	ORF	gtcatccacgttcattggc	tcacaactgaccaaactgcc	60–50	442	+	+	+	+	+	+	−
1	SsrPt_ctg7141	Contig7141[a]	CCG	8	ORF	gaatgacgcattatcagggg	tcacctttctcacctctgcc	45	381	−	+	+	+	+	+	+
1	SsrPt_ctg7170	Contig7170[a]	AGC	5	ORF	ggtttttcgatttctgaggc	aacaggtgtgcaaatagccc	60–50	385	+	+	+	+	+	+	−
1	SsrPt_ctg7425	Contig7425[a]	AAG	6	ORF	aataagaccccagaggagcc	gacgtctttcaccaaatcgc	60–50	384	+	+	+	+	−	−	−
1	SsrPt_ctg7444	Contig7444[a]	AT	10	5′UTR	tcttcaccatcggtttctcc	tggatctgtcacctcctcatc	58	285	+	+	+	+	+	+	+
1	SsrPt_ctg7731	Contig7731[a]	AT	12	5′UTR	agtggtgaagggtccatctg	gcataacacaaaagccagca	51	217	+	+	+	+	+	+	+

Table 2 (continued)

Primer set	Locus information									Amplification						
	Locus name	Identification	Repeated motif	Number of repeat	Position in gene	Forward primer	Reverse primer	Annealing temperature	Expected length (bp)	*Pp*	*Pt*	*Pr*	*Ps*	*Ph*	*Ppi*	*Pc*
1	SsrPt_ctg7824	Contig7824[a]	AT	12	3′UTR	tgacctgtcttgtgagacgc	ttttgaaacagattgcagcc	60–50	501	+	+	+	+	+	−	−
1	SsrPt_ctg7867	Contig7867[a]	CCG	6	5′UTR	ggtcgtggaggaggtaggg	actgataacagctgccccc	45	154	+	+	+	+	+	+	+
1	SsrPt_ctg8064	Contig8064[a]	ACC	6	ORF	gaacgtggttatggcggtag	tcgtggcaactatctcctcc	50	147	+	+	+	+	+	+	+
1	SsrPt_ctg865	Contig865[a]	AT	15	3′UTR	tttcagaagctcccgatttg	cttgtggacatggttaatgaag	45	232	+	+	+	+	+	+	+
1	SsrPt_ctg8767	Contig8767[a]	AGC	8	ORF	tggggaaaaatggcatacat	ggagcagacacccatggact	55	180	+	+	+	−	−	−	−
1	SsrPt_ctg9249	Contig9249[a]	AAG	7	5′UTR	ctgctccctcagctcttcc	agacgtcactgccattaccc	55	156	+	+	−	+	+	+	+
1	SsrPt_ctg946	Contig946[a]	AGG	9	3′UTR	tatcaggtataggcctccgc	aaataggagcccttctggga	53	287	+	+	+	−	−	−	−
1	SsrPt_ctg988	Contig988[a]	AT	7	3′UTR	taataattcaagccaccccg	aacattttgcacgatagccc	51	319	+	+	+	−	−	−	−
2	RPtest1	Contig4518[a]	AAT	7	5′UTR	gatcgttattcctcctgcca	ttcgatatcctccctgcttg	50	125	+	+	+	+	+	+	+
2	RPtest5	Contig6309[a]	AAC	6	ORF	acaacaataataacgggggc	acgctttagatcctcctgca	55	197	+	+	+	+	+	+	+
2	RPtest6	Contig3845[a]	TGC	5	ORF	aggattccaacagcatcacc	ctgaacatgaagcgcagtgt	55	147	+	+	+	+	+	+	+
2	RPtest8	Contig8048[a]	CCG	6	ORF	ggtgcgagattgaaattcgt	tttgcagtctgttgcctttg	60–50	196	−	+	+	NA	NA	NA	NA
2	RPtest9	Contig1667[a]	AGC	10	ORF	ccagacaacccaaatgaagg	gcctgctatcgaatccagaa	51	289	+	+	+	+	+	+	+
2	RPtest11	Contig3631[a]	ATC	7	3′UTR	aggatgcctatgatatgcgc	aaccataacaaaagcggtcg	56	213	+	+	+	+	−	+	+
2	RPt11est13	AA739656[b]	CTG	5	ORF	gatttttcaggaagaccccc	tgtaaggcacaagccctctt	51	277	+	+	+	+	−	−	+
2	RPtest15	Contig8064[a]	ACC	6	ORF	gaacgtggttatggcggtag	ccagggacagttaccagcat	56	246	+	+	+	+	+	+	+
2	RPtest16	AA739818[b]	AGT	5	ORF	cagaaatggcgtccaaattc	accccacttatatccccagc	56	132	+	+	+	+	−	+	+
2	RPtest20	Contig6393[a]	AGC	5	ORF	gttcccactcaagggttgaa	acatcatttgttgccgcata	56	259	+	+	+	−	−	−	−
2	RPtgbLP5	AF013805[b]	AAT	6	5′UTR	agaggttccaaacgagagt	tcgacttctgatttctttacatga	60–50	176	−	+	−	NA	NA	NA	NA
Amplification rate (%)										86.8	100	94.2	85.4	72.9	70.8	64.6

[a] *Pinus taeda* unigene contig numbering (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/)
[b] GenBank accession
[c] *Pinus pinaster* unigene contig numbering (http://cbi.labri.fr/outils/SAM/COMPLETE/index.php)

Table 3 *Pinus radiata* genomic SSR markers that were mapped in *P. pinaster* and marker sequence homologies between *P. pinaster* and *P. radiata*

Primer set	Locus name	Repeated motif	Forward primer	Reverse primer	Annealing temperature (°C)	Expected length (bp)	Sequence homology (%)
3	NZPR1078	AC_{10}	tggtgatcaagcctttttcc	gttgatgagtgatggcatgg	53	342	91.5
3	NZPR114	CA_{15}... CA_{13} TA_{22}	aagatgacccacatgaagtttgg	ggagctttataacatatctcgatgc	56	193	88.2
3	NZPR1702_b	AC_{15} CA_{13}...AT_5	tatgattggaccattggggt	ccaaaccctcctccacatatc	53	187	No homology
3	NZPR413	TG_{23} GT_6	tgaacctcgatggaatagcc	cccgccttgcatcaatta	53	253	89.1
3	NZPR472	AC13	gagaaaattcaaccaccgga	ggttgtagggcagtgaatcc	53	309	89.4
3	NZPR544	CA_5AC_{12} TA_5	gcgatgtgcaacccttgata	tgctattccgtcaaaaaccc	56	286	86.1
3	NZPR823_a	AC_{57}	tatcgggagcaagttatgcc	tgcactctttttcgtctcca	53	296	92.5

eight of the 12 linkage groups between the two maps. Linkage group homology was also confirmed using a set of ESTPs mapped in the INRA-G2 (Chagné et al. 2003) and INRA-F2 pedigrees (D. Chagné and P. Semat, unpublished data).

Level of diversity of cDNA and genomic SSRs in *Pinus pinaster*

The nine polymorphic cDNA-SSR loci and 10 polymorphic genomic SSR loci were genotyped in 26 unrelated *P. pinaster* trees. Their expected heterozygosities (*H*) and number of alleles (A) are shown in Table 4. Within the cDNA-SSRs, there was no significant difference between the heterozygosity values obtained in the ORF and the UTRs, or between tri- and dinucleotide SSRs (F test with a *P* value of 0.46). Within the genomic SSRs, a significant difference (F test with a *P* value of 0.11) of the diversity parameters was found between the loci transferred from *P. radiata* and those were developed from *P. pinaster* and *P. halepensis* by Mariette et al. (2001). This difference suggests that genomic SSRs tend to be less polymorphic when transferred from phylogenetically distant species; *P. radiata* belongs to the Oocarpeae subsection, whereas *P. pinaster* and *P. halepensis* belongs to the Sylvestres subsection of the pine genus (Mirov 1967). Finally, the level of diversity was not different between the transferred *P. radiata* genomic SSRs and the cDNA-SSRs (*F* test with a *P* value of 0.27).

Table 4 Chromosomal assignment and genetic diversity parameters of the three classes of microsatellites genotyped on 26 unrelated *P. pinaster* trees. The mapping location in the INRA-G2 (following linkage group numbering of Chagné et al. 2002) and INRA-F2 maps (following linkage group numbering of Costa et al. 2000) are indicated. *M* Monomorphic, *P* polymorphic, *UL* unlinked, *H* heterozygosity, *A* number of alleles

Marker type	Primer set	Locus ID	Mapping pedigree				Genetic diversity	
			INRA-G2	INRA-F2	AFOCEL-F1	INIA-F1	*H*	A
cDNA-SSR	1	RPtEST11	5	2	P	M	0.74	4
	1	RPtEST13	10	M	M	M	0.66	3
	2	SsrPp_cn524	6	1	P	M	0.81	5
	2	SsrPt_ctg275	P/UL	P/UL	P	P	0.74	8
	2	SsrPt_ctg4363	M	12	P	M	0.68	4
	2	SsrPt_ctg7824	10	M	M	M	0.35	2
	2	SsrPt_ctg988	11	M	P	M	0.55	3
	2	SsrPt_ctg1525	M	11	M	M	0.16	2
	2	SsrPt_ctg64	3	3	M	P	0.68	4
P. radiata genomic SSR	3	NZPR1078	2	7	P	M	0.68	4
	3	NZPR114	M	5	M	P	0.68	5
	3	NZPR1702_b	11	6	P	M	0.38[a]	2[a]
	3	NZPR413	4	8	P	P	0.58	4
	3	NZPR472	1	M	P	P	0.67	4
	3	NZPR544	M	3	M	P	0.41	4
	3	NZPR823_a	5	M	P	P	0.67	3
P. pinaster and *P. halepensis* genomic SSR	4	FRPp91	1	9	P	P	0.85	9
	4	FRPp94	10	5	P	P	0.80	8
	4	ITPh4516	3	3	P	P	0.84	8

[a]These values were not taken into account for the comparison of diversity parameters between cDNA and genomic SSRs

Discussion

Composition and distribution of SSRs in the expressed genome of pine

The SSR composition of the coding region of the pine genome was first compared to the results published in other plant species. In dicotyledonous species where cDNA-SSR evaluations have been reported: i.e., *Vitis vinifera* (Scott et al. 2000) and *Arabidopsis thaliana* (Cardle et al. 2000; Morgante et al. 2002), the most represented repeat types, i.e., AG, AT, AAG, AGG and AGC, were also found to be the most frequent in pines (Fig. 1). Conversely, the most common repeated motif in monocotyledonous species (Varshney et al. 2002), CCG, was quite rare in pines (5.2 and 7.2% in *P. pinaster* and *P. taeda*, respectively). This result suggests that the SSR composition of gymnosperms genes is more similar to that of dicots than monocots. However, given the few number of species analysed, this interpretation remains to be confirmed.

The presence of a majority of trinucleotides in the ORFs (Fig. 2) was also in agreement with that whichhas been described in other plants. Morgante et al. (2002) showed a strong positive selection for trinucleotides in the translated regions of *A. thaliana*. Metzgar et al. (2000) explained the excess of triplet repeat microsatellites in the coding regions by the effect of important mutation pressures. Indeed, a mutation in a mono-, di-, tetra- or pentanucleotide SSR in the ORFs would result in a frameshift that could change the translated protein structure and function.

Morgante et al. (2002) detected much higher levels of SSRs in the 5′ UTRs, especially AG/CT repeats. The rather small number of SSRs detected in the 5′ UTRs of pine genes (17.4%, Table S3) contrasted with their results and could reflect a true feature of pine genes or it could simply be that the low coverage of the 5′-end in the pine ESTs has provided a bias. Some support for the latter view comes from ESTs obtained from the sequencing of the 5′ ends of 3′ anchored cDNAs (Frigerio et al. 2004; Kirst et al. 2003). Therefore, the 5′ UTRs were probably under-represented in the two pine EST collections analysed.

Transferability of cDNA and genomic SSRs in pines

From 64.6 to 94.2% of the pine cDNA-SSRs transferred to one or more of the seven pine species tested (Table 2). It has been clearly shown that the transferability of molecular markers (including SSRs) depends on the phylogenetic distance between species. Most of the markers developed in this study originated from *P. taeda*, an American pine which belongs to the *Pinus* section of the subgenus *Pinus* (Mirov 1967). It is not surprising, therefore, that the highest transfer rate was observed for *P. radiata* markers (94.2%), another American pine belonging to the same section. Similarly, the transfer rate decreased for SSR markers of Mediterranean pines of the same section (*P. pinaster*, 86.8%; *P. sylvestris*, 85.4%; *P. halepensis*, 72.9%), and was even lower with Mediterranean pine markers of the more distant section *Pinea* (*P. pinea*, 70.8%; *P. canariensis*, 64.6%). We also anticipate a lower transferability of cDNA-SSR markers in the subgenus *Strobus*, or even within other genera of the Pinaceae family. However, the transferability rates in these more distant species should be higher for cDNA-SSR markers compared to genomic SSRs (Echt et al. 1999).

Similar rates of cross-species transferability were reported using EST-derived SSR markers in the genus *Medicago* (Eujayl et al. 2003, 89%) and within the Poaceae (Gupta et al. 2003, 55%). Comparatively, genomic SSR markers have shown to be less transferable in pine (54% between *P. radiata* and *P. pinaster*, this study; 29% between *P. strobus* and *P. radiata*, Echt et al. 1999; and 42% between *P. taeda* and *P. pinaster*, Gonzalez-Martinez et al. 2004). This rate is low compared to other plant genera (e.g., up to 85% between *Glycine* spp., Peakall et al. 1998). These results suggest that the data mining of pine cDNA libraries is valuable approach to develop transferable SSR markers. Furthermore, it should be noted that the cDNA-SSR markers were obtained without library screening. Clearly the development of pine sequence databases and the in silico approach described here provides a cost-effective approach to SSR marker development.

In rice and wheat, EST-derived SSR markers have been reported to have lower rate of polymorphisms compared to SSR markers derived from genomic libraries (Cho et al. 2000; Eujayl et al. 2002). However, such differences were not found in *Medicago* (Eujayl et al. 2003) and *Picea* (Scotti et al. 2000) two highly polymorphic genera compared to the highly domesticated cereal crops. Our findings in *P. pinaster* revealed that non-source species genomic SSRs and cDNA-SSRs have similar levels of diversity and thus cDNA-SSRs are not less polymorphic.

At the intraspecific level, these markers have been mapped within the different genetic maps of *P. pinaster*, which will make it possible to construct a consensus map of this species. Nevertheless, more markers will be needed to reach the saturation levels desired. The markers developed in this study were also mapped in the *P. pinaster* genetic map that was aligned with the loblolly pine map using comparative genome mapping (Chagné et al. 2003) and so can be used as orthologous markers in other conifer species.

Conclusion

We have shown in this study that database-sourced cDNA-SSRs can be efficiently developed for, and transferred across, pine species. Pine SSR markers developed in this way are less expensive to produce and are as informative as SSR markers derived from other (genomic-based) methods. However, since these markers correspond to transcribed regions, further study is necessary to determine if they behave as neutral markers or not, if they are to be

used in genetic diversity analysis and in association studies

Acknowledgements D.C. was funded by the French Ministry of Research. This research was supported by grants from France (Ministère de l'Agriculture et de la Pêche-DERF No. 61.45.80.15/02) and the European Union (TREESNIPS project: QLK3-CT-2002-01973). The maritime pine ESTs were produced with the support of the Aquitaine Région (n°2002 0307002A) and INRA (Lignome) as well as the European Union (GEMINI: QLK5-CT-1999-00942). The work at New Zealand Forest Research was funded by New Zealand's Foundation for Research, Science and Technology (CO4X005).

References

Brown GR, Kadel EE III, Bassoni DL, Kiehne KL, Temesgen B, Van Buijtenen JP, Sewell MM, Marshall KA, Neale DB (2001) Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics 159:799–809

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chagné D, Lalanne C, Madur D, Kumar S, Frigerio JM, Krier C, Decroocq S, Savoure A, Bou-Dagher KM, Bertocchi E, Brach J, Plomion C (2002) A high density genetic map of maritime pine based on AFLPs. Ann For Sci 59:627–636

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. Mol Breed 12:185–195

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCough SR, Park WD, Ayer N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oriza sativa*). Theor Appl Genet 100:713–722

Christoffels A, van Gelder A, Greyling G, Miller R, Hide T, Hide W (2001) STACK: sequence tag alignment and consensus knowledgebase. Nucleic Acids Res 29:238–238

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross-transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Costa P, Pot D, Dubos C, Frigerio J-M, Pionneau C, Bodénès C, Bertocchi E, Cervera MT, Remington DL, Plomion C (2000) A genetic map of maritime pine based on AFLP, RAPD and protein markers. Theor Appl Genet 100:39–48

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dubos C, Plomion C (2003) Identification of water-deficit responsive genes in maritime pine (*Pinus pinaster* Ait.) roots. Plant Mol Biol 51:249–262

Echt CS, Vendramin GG, Nelson CD, May-Marquardt P (1999) Microsatellite DNA as shared genetic markers among conifer species. Can J For Res 29:365–371

Elsik CG, Williams CG (2000) Retroelements contribute to the excess low-copy number DNA in pine. Mol Gen Genet 264:47–55

Elsik CG, Williams CG (2001) Low-copy microsatellite recovery from a conifer genome. Theor Appl Genet 103:1189–1195

Elsik CG, Minihan VT, Hall SE, Scarpa AM, Williams CG (2000) Low-copy microsatellite markers for *Pinus taeda* L. Genome 43:550–555

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2003) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Frigerio JM, Dubos C, Chaumeil P, Salin F, Garcia V, Barré A, Plomion C (2004) Using transcriptome analysis to identify osmotic stress candidate genes in maritime pine (*Pinus pinaster* Ait.). In: Sustainable forestry, wood products and biotechnology, BIOFOR Proceedings (in press)

Gao L, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12:245–261

Gonzalez-Martinez SC, Robledo-Arnuncio JJ, Collada C, Diaz A, Williams CG, Alia R, Cervera MT (2004) Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. Theor Appl Genet 109:103–111

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genom 270:315–323

Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of *Ty*1- *copia*-like sequences as component of large genomes in *Pinus elliottii* var. *elliottii* and other gymnosperms. Proc Natl Acad Sci USA 93:2708–2713

Kantety RV, LaRota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum wheat. Plant Mol Biol 48:501–510

Kinlaw CG, Neale DB (1997) Complex gene families in pine genomes. Trends Plant Sci 2:356–359

Kirst M, Johnston AF, Baucom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff R (2003) Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. Proc Natl Acad Sci USA 100:7383–7388

Komulainen P, Brown GR, Mikkonen M, Karhu A, Garcia-Gil MR, O'Malley D, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. Theor Appl Genet 107:667–678

Kossack DS, Kinlaw CS (1999) *IFG*, a gypsy-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. Plant Mol Biol 39:417–426

Kriebel HB (1985) DNA sequences components in the *Pinus strobus* nuclear genome. Can J For Res 15:1–4

Kutil BL, Williams CG (2001) Triplet-repeat microsatellites shared among hard and soft pines. J Hered 92:327–332

Leitch IJ, Hanson L, Winfield M, Parker J, Bennett MD (2001) Nuclear DNA C-values complete familial representation in gymnosperms. Ann Bot 88:843–849

Mariette S, Chagné D, Decroocq S, Vendramin GG, Lalanne C, Madur D, Plomion C (2001) Microsatellite markers for *Pinus pinaster* Ait. Ann For Sci 58:203–206

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Mirov NT (1967) The genus *Pinus*. Ronald, New York

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Price RA, Liston A, Strauss SH (1998) Phylogeny and systematics of *Pinus*. In: Richardson DM (ed) Ecology and biogeography of *Pinus*. Cambridge University Press, Cambridge, pp 49–68

Ritter E, Aragones A, Markussen T, Achere V, Espinel S, Fladung M, Wrobel S, Faivre-Rampant P, Jeandroz S, Favre J-M (2002) Construction and exploitation of a multifunctional and saturated genetic map for coniferous species. Ann For Sci 59:637–643

Salzberg SL, Delcher AL, Kasif F, White O (1998) Microbial gene identification using Markov interpolated models. Nucleic Acids Res 26:544–548

Schiex T, Gouzy J, Moisan A, de Oliveira Y (2003) FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. Nucleic Acids Res 31:3738–3741

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN v.2000, Genetics and Biometry Laboratory, University of Geneva, Switzerland

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Scotti I, Magni F, Fink R, Powell W, Binelli G, Hedley PE (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. Theor Appl Genet 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. Theor Appl Genet 104:1035–1041

Shepherd M, Cross M, Maguire TL, Dieters MJ, Williams CG, Henry RJ (2002) Transpecific microsatellites for hard pines. Theor Appl Genet 104:819–827

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCough SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genet Res 11:1441–1452

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genet Res 10:967–981

Van Ooijen JW, Voorrips RE (2001) Joinmap 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen. Website: http://www.joinmap.nl

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol 7:537–546

Zhou Y, Bui T, Auckland LD, Williams CG (2002) Undermethylated DNA as a source of microsatellite from a conifer genome. Genome 45:91–99