ORIGINAL PAPER

S. C. González-Martínez · J. J. Robledo-Arnuncio ·
C. Collada · A. Díaz · C. G. Williams · R. Alía ·
M. T. Cervera

Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines

Abstract Microsatellite transfer across coniferous species is a valued methodology because de novo development for each species is costly and there are many species with only a limited commodity value. Cross-species amplification of orthologous microsatellite regions provides valuable information on mutational and evolutionary processes affecting these loci. We tested 19 nuclear microsatellite markers from *Pinus taeda* L. (subsection *Australes*) and three from *P. sylvestris* L. (subsection *Pinus*) on seven Eurasian hard pine species (*P. uncinata* Ram., *P. sylvestris* L., *P. nigra* Arn., *P. pinaster* Ait., *P. halepensis* Mill., *P. pinea* L. and *P. canariensis* Sm.). Transfer rates to species in subsection *Pinus* (36–59%) were slightly higher than those to subsections *Pineae* and *Pinaster* (32–45%). Half of the trans-specific microsatellites were found to be polymorphic over evolutionary times of approximately 100 million years (ten million generations). Sequencing of three trans-specific microsatellites showed conserved repeat and flanking regions. Both a decrease in the number of perfect repeats in the non-focal species and a polarity for mutation, the latter defined as a higher substitution rate in the flanking sequence regions close to the repeat motifs, were observed in the trans-specific microsatellites. The transfer of microsatellites among hard pine species proved to be useful for obtaining highly polymorphic markers in a wide range of species, thereby providing new tools for population and quantitative genetic studies.

S.C. González-Martínez and J.J. Robledo-Arnuncio have contributed equally to this work and are to be considered as joint first authors

S. C. González-Martínez · J. J. Robledo-Arnuncio · R. Alía ·
M. T. Cervera (✉)
Unidad de Genética Forestal, CIFOR-INIA,
P.O. Box 8111, 28080 Madrid, Spain
e-mail: cervera@inia.es
Tel.: +34-91-3476857
Fax: +34-91-3572293

C. Collada · A. Díaz
Departamento de Biotecnología, ETSIM,
Ciudad Universitaria s/n,
28040 Madrid, Spain

C. G. Williams
Graduate Genetics Program, TAMU 2135,
Texas A & M University,
College Station, TX 77843-2125, USA

Introduction

Trans-specific microsatellites are a valuable tool in pines because de novo development is difficult in the large, highly duplicated conifer genome (Kostia et al. 1995; Soranzo et al. 1998; Joyner et al. 2001; Mariette et al. 2001). Although a number of methods have been developed to circumvent problems associated with repetitive DNA (Zhou et al. 2002), the transfer of microsatellites between related conifer species is one of the more promising (Echt et al. 1999; Kutil and Williams 2001; Shepherd et al. 2002). Shepherd et al. (2002) reported successful microsatellite transfer in hard pines (subgenus *Pinus*) for species that diverged a few million generations ago. Polymorphism was observed to be high within these narrow phylogenetic limits (Shepherd et al. 2002). Microsatellite transfer across pine subgenera (about 14 million generations since divergence) has also been shown to be feasible, but the extent of polymorphism is unknown at this level (Kutil and Williams 2001). Trans-specific microsatellites can provide a test of evolutionary models for microsatellite DNA regions (Karhu et al. 2000; Primmer and Ellegren 1998) because they are often phylogenetically informative (see Zhu et al. 2000). Variation relevant to infer phylogenetic relationships can be found both in flanking sequences and in the repeat motif. Trans-specific microsatellites are also useful for anchoring genetic maps in comparative mapping, for analysing the conservation of gene order across orthologous linkage groups and for verifying quantitative trait loci (QTLs) (Devey et al. 1999; Meksem et al. 2001).

A common observation in microsatellite transfer studies is that allele lengths are longer in focal species than in non-focal species. These results may possibly be explained by the presence of an inherent bias in protocols that select for longer-than-average microsatellites in the focal species (i.e., "ascertainment bias" as defined by Ellegren et al. 1995). However, evidence of ascertainment bias remains contradictory. It has not been found in intensive studies of wasps and cattle (Crawford et al. 1998; Zhu et al. 2000). In other cases, such as between humans and chimpanzees, inter-specific length differences in orthologous microsatellites cannot entirely be explained by ascertainment bias (Cooper et al. 1998). In fact, Cooper et al. (1998) showed that longer microsatellites in humans could be explained by a mutational bias in favour of microsatellite expansions and a higher average genome-wide microsatellite mutation rate in the human lineage.

Pines are woody perennial species with a long lifespan. They are characterized by highly outcrossing mating systems and approximately 10 years per generation as a general rule (Kutil and Williams 2001). Mediterranean pines are a heterogeneous assembly of hard (diploxylon) pine species that have been traditionally considered to be relic taxa from the Tertiary period, belonging to different evolutionary lineages (Klaus 1989). The complex phylogenetic relationships among the Mediterranean pine species themselves (subsections *Pinaster* and *Pineae*, based on Frankis 1993 and Liston et al. 1999) and among these and other Eurasian pines (subsection *Pinus*) have been the subject of active investigation using morphological and molecular data, and the conclusions drawn from these studies are still under debate (see Price et al. 1998 for a review). Pine species occur over a broad area in the Mediterranean basin, showing a remarkable ecological plasticity (Barbéro et al. 1998) and playing an important biological and commercial role. Some of the singular ecological adaptations of Mediterranean tree species have taken on a special significance in recent years as more precise predictions on climate change become available. The genetic variability of those Eurasian forest tree species, such as *P. halepensis* or *P. pinaster*, or of marginal populations from southern locations of widespread species such as *P. sylvestris*, which are adapted to warm and dry conditions, are of great interest to breeders and genetic conservation programmes.

Eurasian hard pines show distinct patterns of genetic variation and population structure. Highly heterozygous species, such as *P. sylvestris* (Prus-Glowacki and Stephan 1994), are sympatric with species showing a remarkable lack of polymorphism, such as *P. pinea* (Fallour et al. 1997). *Pinus halepensis* has a widespread and continuous range, yet *P. pinaster* has a widespread and scattered range and *P. canariensis* is an island endemic. Trans-specific microsatellites are needed to address evolutionary questions related to gene flow and adaptation at a multispecific (ecosystem) level. To date, there has not been a microsatellite-transfer study for pines in the Mediterranean region, and few de novo microsatellites are available— namely, *P. sylvestris* (Kostia et al. 1995; Soranzo et al. 1998; Karhu et al. 2000), *P. halepensis* (Keys et al. 2000) and *P. pinaster* (Mariette et al. 2001).

In the investigation reported here we tested microsatellite transfer from *P. taeda* and *P. sylvestris* to seven Eurasian hard pines widely distributed along the Mediterranean basin (*P. pinaster*, *P. pinea*, *P. halepensis*, *P. canariensis*, *P. n*igra, *P. sylvestris* and *P. uncinata*). Transferred markers (i.e., the clear amplification pattern of a single locus) were screened for gene diversity and allelic richness in order to demonstrate their practical usefulness in genetic studies. Because the set of species included in this study showed a wide range of phylogenetic distances, we were also able to address the extent of polymorphism retention from a few million years ago to over 100 million years ago (estimated divergence time between New World and Eurasian hard pines), thereby extending previous results by Shepherd et al. (2002). Furthermore, trans-specific microsatellites were used to (1) study sequence variation in the flanking and repeat regions across different Eurasian pine species and (2) evaluate the usefulness of the transferred markers for a population-level distinction between *P. sylvestris* and *P. uncinata*, two closely related cross-mating species.

Materials and methods

Species classification and DNA sources

We followed the taxonomic classification of Price et al. (1998), with the exception of recognizing a subsection *Pinaster* (Frankis 1993; Liston et al. 1999), which includes the species placed by Price et al. (1998) in subsections *Halepenses* and *Canarienses* together with two species from subsection *Pinus*—*P. pinaster* and *P. merkusii*. Two hard pine (subgenus *Pinus*) species were used as focal taxa of microsatellite loci: *P. taeda* and *P. sylvestris*. *Pinus taeda* belongs to the subsection *Australes* (New World hard pines), while *P. sylvestris* belongs to subsection *Pinus* (Eurasian hard pines). Seven native pine taxa occurring in the Mediterranean basin (including Canary Islands pine, a former Mediterranean-distributed species) were included as non-focal species for transfer, namely *P. pinea* (subsection *Pineae*); *P. halepensis*, *P. canariensis* and *P. pinaster* (subsection *Pinaster*); *P. nigra*, *P. sylvestris* and *P. uncinata* (subsection *Pinus*). Note that *P. sylvestris* was used as a both focal and non-focal species.

Control DNA samples of *P. taeda* were randomly selected from a three-generation outbred pedigree (Devey et al. 1991). Twenty individuals of each of the other seven species were randomly selected from three to four different native populations, with one exception. This exception was *P. uncinata*, which has a very restricted distribution range; samples were collected in just one population in the Pyrenees (northeastern Spain). DNA was extracted from needle tissue following the protocol described by Dellaporta et al. (1983). Of the 20 DNA samples collected, two were used to evaluate amplification and optimize PCR conditions. All 20 samples were used to estimate the polymorphism of the transferred loci or to confirm monomorphism. In addition, two samples from *P. taeda* and *P. sylvestris* were used as focal-species controls.

Table 1 Results of transfer and PCR conditions used to amplify 22 microsatellite loci in seven Eurasian hard pine species

Locus	Focal taxon[a]	Repeat motif	Expected size (bp)	T_m[b] (°C)	PCR[c] protocol	Results[d]						
						PU	PS	PN	PP	PPA	PH	PC
PtTX3032	PT	$(GAT)_{35}(GAC)_3GAT(GAC)_8...(GAC)_6AAT(GAT)_6$	335	59↓49	a[e]	1	1	4	3	4	4	4
PtTX3107	PT	$(CAT)_{14}$	182	60↓50	a	1	1	1	1	2	1	2
PtTX3116	PT	$(TTG)_7...(TTG)_5$	146	55↓45	a[e]	1	1	3	1	3	1	1[f]
PtTX4001	PT	$(CA)_{15}$	224	60↓50	a	1	1	1	2	4	2	1[f]
PtTX4011	PT	$(CA)_{20}$	305	60↓50	a	1	1	1[f]	2	2	4	2
PtTX2094	PT	$(TTTG)_4...(TTTG)_9...(TTTG)_9T_{11}$	324	65↓55	a[e]	2	2	2	2	2	2	2
PtTX2128	PT	$(GAC)_8$	245	60↓50	a	2	2	2	2	2	2	2
PtTX3018	PT	$(GAT)_{13}$	155	55↓45	a[e]	2	2	2	2	2	2	2
PtTX3019	PT	$(CAA)_{10}$	223	55↓45	a[e]	2	2	4	4	2	4	4
PtTX3020	PT	$A_{16}(CAA)_9$	211	60↓50	a[e]	3	3	3	3	3	3	3
PtTX3063	PT	$(CAA)_7(CAT)_2(CAA)_6...(CAA)_{15}$	268	60↓50	a[e]	3	3	3	3	3	3	3
PtTX3098	PT	$(GTT)_8$	187	60↓50	a	2	2	2	2	2	2	2
PtTX2080	PT	$(CAT)_5...(CGT)_7CAT(CGT)_2(CGT)_7$	161	61↓58	a[e]	4	4	4	4	4	4	4
PtTX2082	PT	$(GT)_{14}(GAGT)_7(GA)_{13}$	253	52↓49	a[e]	4	4	4	4	4	4	4
PtTX2164	PT	$(CGT)_5...(CCG)_4...(CGT)_7(CAT)_3(CGT)_4CAT(CGT)_5TGTCGTCAT(CGT)_5(CAT)_{10}$	252	60↓50	a	4	4	4	4	4	4	4
PtTX4003	PT	$(GT)_{13}TT(AT)_4(GT)_2$	159	51↓48	a[e]	4	4	4	4	4	4	4
PtTX3055	PT	$(GAT)_5...(GAT)_8...(GAT)_6$	402	56↓53	a[e]	4	4	4	4	4	4	4
PtTX4055	PT	$(TG)_3CG(TG)_3CGTA(TG)_{12}...(GA)_{18}...(GA)_{18}$	187	51↓48	a[e]	4	4	4	4	4	4	4
PtTX4093	PT	$(CTT)_{16}...(CTT)_6...(CTT)_3...(CTT)_4$	337	60↓50	a[e]	3	3	3	4	3	3	4
SPAC11.6	PS	$(CA)_{29}(TA)_7$	165	55↓50	b	1	-	4	4	4	4	1
SPAC12.5	PS	$(GT)_{20}(GA)_{10}$	155	55↓50	b	1	-	4	4	4	4	4
SPAG7.14	PS	$(TG)_{17}(AG)_{21}$	209	55↓50	b	1	-	1	4	4	4	1

[a] PT, *Pinus taeda* (primers from Auckland et al. 2002); PS, *P. sylvestris* (primers from Soranzo et al. 1998)

[b] Starting and final temperatures of the touchdown PCR profile

[c] See Materials and methods for details

[d] Result of transfer to each of the taxa: PU, *P. uncinata*; PS, *P. sylvestris*; PN, *P. nigra*; PP, *P. pinaster*; PPA, *P. pinea*; PH, *P. halepensis*; PC, *P. canariensis*. The transfer classes are: 1, polymorphic; 2, monomorphic; 3, non-specific amplification or complex pattern; 4, no amplification

[e] Amplification tested using 5% DMSO

[f] Microsatellite loci initially monomorphic but polymorphic when tested in a larger sample size (20 individuals)

Microsatellite transfer

Transference is defined here as the positive amplification of a PCR band of the expected size. Twenty-two microsatellite markers (Table 1) were assessed for transfer. Markers were selected at random from already available sources. Nineteen markers were originally developed in *P. taeda* from different DNA libraries (16 from low-copy and under-methylated libraries; Auckland et al. 2002) and the other three markers were originally developed in *P. sylvestris* from total genomic libraries (Soranzo et al. 1998). The amplification of *P. taeda* microsatellite loci followed the protocol described by Auckland et al. (2002) (protocol a in Table 1). We attempted to optimize the PCR analyses by making the following modifications: 5% DMSO was used for amplification, and new annealing temperatures of the touchdown profile were tested (Table 1). The total number of cycles and concentration of PCR mixture products were not changed. Microsatellite loci described originally in *P. sylvestris* were tested for transfer using the protocol described in Soranzo et al. (1998) (protocol b Table 1). A Perkin-Elmer GenAmp 9700 thermal cycler (Perkin Elmer, Foster City, Calif.) was used to carry out all reactions.

PCR amplification was initially assessed on 1.7% agarose gels. Products of positive amplifications were then resolved on 6% acrylamide/bisacrylamide (19:1), 7 *M* urea and 1× TBE denaturing gels. The gels were run at 45 W constant power for 1–2 h using a Li-Cor 4200 Series automated DNA sequencer (Li-Cor, Lincoln, Neb.). Amplified fragments were sized by GENE IMAGIR ver. 3.56 software (Scanalytics), using external standards. A positive result of transfer to non-focal species was reported when a clear amplification pattern of the expected size and polymorphism were observed (class 1). Other observed results were classified as follows: class 2, monomorphic amplification product of the expected size; class 3, non-specific amplification, complex pattern or poor amplification; class 4, no amplification.

Expected heterozygosity (H_e) was calculated following Nei (1973). Allelic richness (*A*) at a locus was measured as the number of different alleles observed in a sample of 10 to 20 diploid individuals. In order to estimate allelic richness for a fixed sample size, a rarefaction method (Hurlbert 1971) was used to estimate A_{10}, the number of different alleles at a locus for a sample of ten diploid individuals. The results obtained from non-focal species were compared with analogous population genetic data from *P. taeda* (Al-Rababah and Williams 2002). Differences in average microsatellite size (population genetic data) and length (sequencing data, see below) between *P. taeda* and non-focal species were tested using a non-parametric sign test. One goal of the present study was to develop a set of highly polymorphic markers able to distinguish—at a population level—between *P. sylvestris* and *P. uncinata*, two closely related cross-mating species. The utility of trans-specific microsatellites for this purpose was tested using a Fisher exact test of population differentiation between the two species (Raymond and Rousset 1995a). Genotypes of 20 trees per species, scored at eight trans-specific microsatellites each, were considered for the Fisher exact test, which was performed using GENEPOP version 3.3. software package (Raymond and Rousset 1995b).

Sequence variation of microsatellite loci

Two amplification products from each of three trans-specific microsatellites (PtTX3107, PtTX4001 and PtTX3116; see Results) were sequenced for all non-focal species with successful cross-amplification. PCR products were precipitated with ethanol and then cloned using the pGEM-TEasy vector (Promega, Madison, Wis.). At least four clones from each transformation were sequenced to reduce *Taq* polymerase PCR artifacts. Only consensus sequences are reported. DNA sequencing was carried out using dye terminator sequencing reagents (Perkin-Elmer) and an automated ABI 377 sequencer. GenBank accession numbers are given in Table 2. Sequences were aligned using the CLUSTAL-W method included in MEGALIGN software (DNASTAR), followed by manual alignment adjustments.

To test for polarity of mutation, we computed the rate of change for a given nucleotide position from the repeat motif as the

Table 2 GenBank accession numbers for trans-specific microsatellite sequences

Species	Microsatellite sequence		
	PtTX3107	PtTX3116	PtTX4001
Pinus uncinata	AY304038	AY304033	AY304045
Pinus sylvestris	AY304039	AY304034	AY304046
Pinus nigra	AY304040	-	AY304047
Pinus pinaster	AY304041	AY304035	AY304048
Pinus pinea	AY304042	-	-
Pinus halepensis	AY304043	AY304036	AY304049
Pinus canariensis	AY304044	AY304037	AY304050

observed number of substitutions and insertions/deletions at this position over all the sequences divided by the number of sequences analysed. Generalized linear models were tested using the GLM procedure of SAS version 9.0 statistical package (SAS Institute, Cary, N.C.). In these models, the rate of change was considered to be the response variable and the distance to the repeat motif was the explanatory variable. Models were constructed combining 5′ and 3′ flanks. Confidence intervals for parameter estimates were obtained using standard methods. Base substitution at position 52 in PtTX4001 was excluded from regression analysis because it was common to all non-focal species. Finally, non-parametric unpaired Wilcoxon tests were computed to test if there were higher rates of change within the 5–10 bp immediately adjacent to the repeat region than in flanking sequences further away. This level of analysis (5–10 bp) was chosen following Brohede and Ellegren (1999).

Results

The percentage of microsatellites successfully transferred (classes 1 and 2 in Table 1) from *P. taeda* to subsection *Pinus* was moderately high (36–53%). The levels of polymorphism varied among species within this subsection from three polymorphic markers in *P. nigra* to five in *P. sylvestris* and *P. uncinata*. The transference rate was slightly lower to pines in subsections *Pineae* and *Pinaster* (36–42%). Polymorphism was found in two markers in *P. halepensis*, *P. pinaster* and *P. canariensis* (see Table 1), whereas no polymorphism was found in *P. pinea*. Out of the ten microsatellites that were successfully transferred from *P. taeda* (classes 1 and 2, Table 1), five (50%) showed polymorphism (class 1) in some of the non-focal species, which diverged from the focal species over an evolutionary period of approximately 100 million years (or ten million generations).

Trans-specific microsatellites originally developed in *P. sylvestris* were always polymorphic but also more phylogenetically limited (Table 1). Positive cross-amplification was mainly achieved within its own subsection (*P. uncinata* and *P. nigra*), although two microsatellites were successfully transferred to *P. canariensis* (SPAC11.6 and SPAG7.14). All primers with clear amplification patterns in *P. sylvestris* were also successfully amplified in *P. uncinata*.

Polymorphism level varied among markers in non-focal species (Table 3). With the allelic richness corrected for a sample of ten diploid individuals (A_{10}), it ranged between 1.0 and 14.0. Expected heterozygosity (H_e) ranged from moderate (0.29) to high (0.93) values. Microsatellites with higher number of repeats, such as

Table 3 Gene diversity and allelic richness for eight nuclear microsatellites in *Pinus taeda* (focal species) and seven Eurasian hard pine species (*nt* non-tested microsatellites; – tested but not transferred microsatellites)

Locus	*Pinus taeda* (focal species)				*P. uncinata*				*P. sylvestris*				*P. nigra*				*P. pinaster*				*P. pinea*				*P. halepensis*				*P. canariensis*			
	n[a]	A (A_{10})[b]	H_e[c]	Allele size range (average)	n	A (A_{10})	Allele size range (average)	H_e	n	A (A_{10})	Allele size range (average)	H_e	n	A (A_{10})	Allele size range (average)	H_e	n	A (A_{10})	Allele size range (average)	H_e	n	A (A_{10})	Allele size range (average)	H_e	n	A (A_{10})	Allele size range (average)	H_e	n	A (A_{10})	Allele size range (average)	H_e
PtTX3107	18	5 (4.54)	0.73	160–191 (172)	18	3 (2.78)	153–159 (156)	0.29	20	5 (4.98)	150–165 (159)	0.80	15	4 (3.86)	153–159 (158)	0.59	16	4 (3.48)	147–182 (152)	0.37	20	1 (1.00)	150	0	20	2 (2.00)	159–165 (161)	0.40	20	1 (1.00)	147	0
PtTX4001	20	7 (6.61)	0.84	211–237 (224)	19	5 (4.17)	201–215 (204)	0.44	20	7 (6.64)	201–219 (210)	0.82	16	6 (5.12)	202–217 (209)	0.73	12	1 (1.00)	207	0	–	–	–	–	20	1 (1.00)	201	0	20	7 (6.31)	210–225 (218)	0.82
PtTX4011	19	5 (4.31)	0.69	283–314 (292)	14	3 (2.98)	260–278 (262)	0.70	12	2 (2.00)	258–264 (261)	0.49	14	2 (2.00)	161–165 (163)	0.50	13	1 (1.00)	165	0	12	1 (1.00)	177	0	–	–	–	–	10	1 (1.00)	285	0
PtTX3116	19	6 (5.64)	0.77	132–167 (145)	17	8 (6.71)	119–166 (133)	0.68	10	10 (10.00)	119–226 (161)	0.79	–	–	–	–	14	4 (3.70)	123–148 (132)	0.60	–	–	–	–	19	2 (2.00)	114–120 (118)	0.48	20	4 (3.64)	146–173 (165)	0.52
PtTX3032	18	14 (11.01)	0.90	270–355 (322)	14	10 (9.11)	323–434 (352)	0.87	16	18 (13.97)	320–419 (376)	0.93	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SPAC11.6	nt	nt	nt	nt	10	9 (9.00)	147–179 (156)	0.72	15	12 (10.34)	145–169 (154)	0.88	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	18	7 (6.42)	123–141 (135)	0.81
SPAC12.5	nt	nt	nt	nt	16	6 (5.49)	149–177 (156)	0.78	17	17 (12.64)	133–197 (154)	0.92	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SPAG7.14	nt	nt	nt	nt	19	13 (9.62)	193–301 (227)	0.87	14	11 (9.81)	187–211 (201)	0.88	12	8 (7.46)	183–237 (191)	0.78	–	–	–	–	–	–	–	–	–	–	–	–	19	14 (10.21)	178–222 (197)	0.87

[a] *n*, Sample size
[b] *A*, Allelic richness; A_{10}, allelic richness after rarefaction to a sample size of 20 (ten diploid individuals)
[c] H_e, Genetic diversity as estimated by Nei (1973)

PtTX4001, PtTX3032 or SPAG7.14, generally corresponded with higher expected heterozygosity values. When all five polymorphic markers transferred from *P. taeda* were taken into consideration, in 22 out of 26 cases the average allele size was larger in the focal species; the only exceptions were those corresponding to microsatellites with a complex repeat motif pattern (PtTX3116 and PtTX3032; Tables 1, 3). This difference was highly significant, as shown by a non-parametric sign test ($P<0.001$).

Important polymorphism differences were also apparent among species for each marker, with *P. sylvestris* being consistently more polymorphic at all transferred loci (Table 3). Despite its restricted distribution, levels of diversity in *P. uncinata* remained high when compared with widespread species such as *P. sylvestris* (Table 3). Trans-specific polymorphism was common between *P. sylvestris* and *P. uncinata*, with an average of 30% of the alleles showing the same size in both species (data not shown). However, *P. uncinata* was clearly differentiated from *P. sylvestris*, as shown by a Fisher exact test of differentiation based on allele frequencies ($P<0.01$).

Cloning and sequencing of successfully transferred markers enabled flanking and repeat regions to be analysed across all seven species (Fig. 1). Basic repeat structure was conserved in two loci—PtTX3107 and PtTX4001. Only in three species—*P. halepensis* at PtTX3107, and *P. pinaster* and *P. canariensis* at PtTX4001—were slight changes in the repeat motif observed. Sequences flanking the repeat region were also highly preserved in PtTX3107. Only five point mutations were identified in four different species: *P. pinaster*, *P. pinea*, *P. halepensis* and *P. canariensis* (Fig. 1). The flanking region of PtTX4001 was more variable, with point mutations per 100 bp from 2.63 in *P. halepensis* to 4.21 in *P. canariensis* and *P. nigra* and with one base insertion in the latter. Pines belonging to subsection *Pinus* (*P. uncinata*, *P. sylvestris* and *P. nigra*) shared two point mutations that were not present in *P. pinaster*, *P. halepensis* or *P. canariensis* (Fig. 1). In contrast to PtTX3107 and PtTX4001, microsatellite locus PtTX3116 showed a complex pattern of variation among species. The main tri-nucleotide repeat region in the focal taxon (TTG) was only found in one of the non-focal species (*P. canariensis*), instead tri- and six- nucleotide compound motifs (TTA and TTATTG) or perfect repetitions of the six-nucleotide motif TTATTG (*P. pinaster*) were present (Fig. 1). The flanking sequence of the PtTX3116 microsatellite was also considerably variable, in particular in the regions closer to the repeat motif.

When all three microsatellite markers were considered together (PtTX3107, PtTX3116 and PtTX4001), there were fewer perfect repeats in the non-focal species than in the focal species, *P. taeda*, a decrease which was also associated with a total shortening of the repeat region (Fig. 1). There were 14 perfect repeats of a trinucleotide motif CAT in PtTX3107 of *P. taeda*, whereas there were between four and nine in the non-focal species. In PtTX4001, the dinucleotide motif CA was repeated 15 times in *P. taeda* but only an average of eight times in the remaining species. The difference in microsatellite length

Fig. 1 Alignment of nucleotide sequences among eight hard pine species for three microsatellite loci. The first sequence corresponds always to the focal species, *Pinus taeda*. Gaps are indicated by *dashes*, identical nucleotides are indicated by *dots*, *N*=A/C/G/T. The repeat regions are shown in *bold*. *Numbers* indicate the positions of the last nucleotide of the row when repeat regions are excluded. The name of the locus and accession number of the focal-species sequence are provided for each alignment

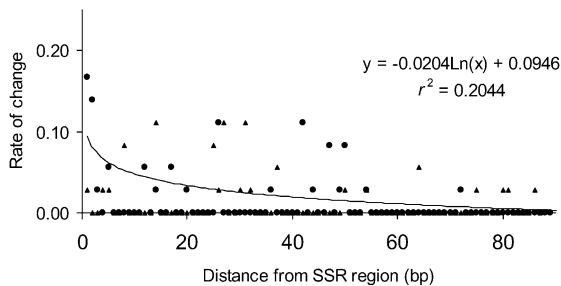Fig. 1 alignment sequences (PtTX3107, PtTX3116, PtTX4001)

Fig. 2 Rate of nucleotide change at the 5′ (*circles*) and 3′ (*triangles*) flanking sequences in the 0- to 90-bp range of distance from the repeat region. A logarithmic adjustment is shown. r^2 is the coefficient of determination. Loci used in this analysis are PtTX3107, PtTX3116 and PtTX4001

between *P. taeda* and non-focal species was highly significant as shown by a non-parametric sign test ($P<0.000$).

A polarity of mutation was also observed. The rate of change in the flanking sequence followed a logarithmic curve, moderately decreasing from the regions close to the repeat motifs to the extremes of the flanking sequences (Fig. 2). The value of the slope parameter ($b=-0.0204$), which indicates the decrease of the rate of change with distance, was significantly different from zero. The 95% confidence intervals for this parameter varied from -0.0289 to -0.0118. The non-parametric Wilcoxon test showed a significantly higher rate of change within the five nucleotides closer to the repeat

motif ($P<0.004$) than for the rest of the sequence as well as a marginal significance level when the number of nucleotides in the proximity of the repeat region was increased to ten ($P<0.083$). Polarity for mutation was only apparent in the 5′ flanking region of the repeat motifs (Fig. 2), in particular for those loci showing a higher sequence variation (see, for example, the 5′ flanking region of PtTX3116 in Fig. 1).

Discussion

Nuclear microsatellites from *P. sylvestris* (three loci tested) transferred preferentially within its own subsection (*Pinus*), while some of the 19 microsatellites from *P. taeda* (subsection *Australes*) amplified across subsections *Pinus* (36–53%), *Pineae* (36%) and *Pinaster* (36–42%). The relatively high transfer rate (53%) from *P. taeda* to *P. sylvestris* and *P. uncinata*, the latter of which belong to subsections that diverged probably more than 100 million years ago (or 10 million generations) (Geada-López et al. 2002), suggests the utility of low-copy and under-methylated microsatellites as a generic source of molecular markers in hard pines and extends the results obtained by Shepherd et al. (2002) in a New World hard pine complex.

In pines, Echt et al. (1999) successfully transferred 80% of the microsatellites tested within subgenus *Strobus* but only 29% among subgenus *Strobus* and *Pinus*, which diverged over 130 million years ago (approximately 13 million generations in pines) according to the fossil record (Miller 1977). Kutil and Williams (2001), using triplet-repeat microsatellites, found high transfer rates (100%) from *P. taeda* (subsection *Australes*) to a close relative of the same subsection, *P. palustris* but only moderate (and similar to our study) transfer success (47%) to *P. halepensis* (subsection *Pinaster*). Shepherd et al. (2002) obtained the same transfer rates between two different pine subsections (*Australes* and *Oocarpae*) of New World hard pines as within one of them (*Australes*). This result might be explained by the substantially shorter period (fewer than 20 million years or two million generations; Geada-López et al. 2002) since New World hard pine subsections split.

Trans-specific microsatellites turned out to be polymorphic in 25 out of 62 cases (40%), showing moderate to high levels of heterozygosity (0.29–0.93), and therefore providing a valuable source of molecular tools for population genetic studies. This is particularly true in species for which no other microsatellite markers are available (*P. nigra*, *P. uncinata* and *P. canariensis*) or are scarce (*P. pinaster*). The set of markers developed was also useful for distinguishing between *P. uncinata* and *P. sylvestris*, thereby providing a potential tool for the characterization of forest reproductive material and the study of hybridization between these two closely related species. *Pinus sylvestris* was consistently more polymorphic at all of the transferred loci, which agrees with the high levels of variation found in this species using such various molecular markers as allozymes (Prus-Glowacki and Stephan 1994), mtDNA and RFLPs (Soranzo et al. 2000). In contrast, no trans-specific microsatellite was polymorphic in *P. pinea* and, consequently, de novo development of microsatellites seems mandatory for this species. This lack of polymorphism in *P. pinea* was a not surprising result, as this is a conifer species that has also been shown to have a remarkable low heterozygosity with other markers—allozymes (Fallour et al. 1997) and cpSSRs (Gómez et al. 2002, G.G. Vendramin personal communication) that is only comparable to the North American *P. resinosa* (Echt et al. 1998). However, the apparently monomorphic PtTX3107 in *P. pinea* presented a well-structured repeat region (see Fig. 1), which did turn out to be polymorphic when we tested an additional sample of individuals from other populations (a new 159-bp variant, with eight repetitions of the CAT motif, was found).

High sequence similarity among primer binding sites, flanking regions, and repeat motifs is treated as orthology for trans-specific microsatellites (Kutil and Williams 2001). Conserved flanking regions and consistent repeat motifs in PtTX3107 and PtTX4001 supported the cross-amplification of orthologous sequences. In contrast, repeat motifs were not conserved across species in PtTX3116, which could be an evidence of paralogy. Karhu et al. (2000) observed locus duplication at microsatellite *RPS 105* (not included in this study) in the soft pine *P. strobus*. Sequences for *RPS 105a* and *RPS 105b* had a 94% similarity in the flanking region but a totally different repeat structure. Elsik and Williams (2001) showed that pine microsatellites occur in families with the different members of the same family having different repeat motifs but the same flanking regions. Paralogy in PtTX3116 could explain the observed high similarity of flanking regions and primer-binding sites in cross-amplified sequences showing different repeat motifs.

The point mutations and indels found in the flanking sequences of microsatellite repeat motifs have been considered to be phylogenetically informative in several species, including crabs (Orti et al. 1997), wasps (Zhu et al. 2000) and pines (Shepherd et al. 2002). If we consider the variation in flanking sequences (PtTX3116 excluded) that we observed, our results are consistent with the inclusion of *P. pinaster* in the assembly of Mediterranean hard pine species (see, for instance, base substitutions at positions 66, 107 and 109 in PtTX4001) and with the differentiation of Eurasian from New World hard pines, as suggested previously by Frankis (1993), Krupkin et al. (1996) and Geada-López et al. (2002).

The rate of change in the flanking sequence of trans-specific microsatellites was higher in the regions close to the repeat motifs. This result can be interpreted as either higher sequence instability nearby the microsatellite region or as base substitution events causing imperfections in the extremes of the repeat region. Point mutations causing imperfections arise more frequently in the extremes than in the centre of repeat arrays ("polarity of substitutions", as defined by Brohede and Ellegren 1999;

Ellegren 2000). Brohede and Ellegren (1999) found a relative sequence instability within a 5- to 10-bp region at the border of microsatellite repeat regions in cattle and subsequently advanced three different models to account for this variation: (1) differences in the probability of replication slippage in nearly completely replicated repeat tracks and in tracks where replication has just started; (2) loop formation in connection with homologous recombination or non-reciprocal gene conversion; (3) a higher tendency to accept mismatches in repeat ends than in the middle of the repeat regions during homologous recombination following DNA damage. In pines, variable evolutionary rates for different parts of a microsatellite locus have been recently suggested (Karhu 2001), but differences in the rate of change in the flanking sequences, as shown by this study, have not been previously reported. However, our results are based on only three microsatellite markers with highly conserved priming sites and might not stand up to a broader microsatellite sampling.

Microsatellite markers transferred from *P. taeda* showed, in general, a smaller average length (sequencing data) or size (diversity screening data) in the non-focal species. Ascertainment bias could explain these results. However, this study was limited in its ability to estimate ascertainment bias because it did not include reciprocal transference studies (Cooper et al. 1998; Crawford et al. 1998). An alternative hypothesis to ascertainment bias is related to the evolution of microsatellite loci. Young microsatellites, evolving from a shorter allele length, have a bias toward allele length expansion (Taylor et al. 1999). Imperfections (interruptions of the repeat motif) tend to accumulate over the life cycle of a microsatellite region, which lowers the mutation rate and eventually leads to the degradation and loss of the repeat motif (Zhu et al. 2000). Mutational bias in favour of microsatellite expansions in *P. taeda* (microsatellites in young stages) or a higher average genome-wide microsatellite mutation rate in *P. taeda* than in non-focal species could also explain our results. Additional support for the evolutionary hypothesis comes from the observation that trans-specific markers that diverged over relatively short evolutionary times did not show differences in average allele size or length. This was seen in the data for *P. uncinata*, *P. sylvestris* and *P. nigra*, all of which are from subsection *Pinus*, and is supported by data for New World hard pines (Shepherd et al. 2002).

In conclusion, given the difficulty of de novo development of microsatellite markers in species with limited commercial value, cross-amplification of loci among phylogenetically close species would seem to be an adequate strategy, in particular when the objective is only to develop a limited number of markers to accomplish population or conservation genetic studies. In hard pines, successful microsatellite transfer has proved to be feasible for species that diverged as long as 100 million years ago (ten million generations). A limited sequencing effort can produce relevant information about the evolution of this widespread kind of markers, allowing the selection of those that more closely adjust to theoretical microsatellite mutation models, as the stepwise mutation model, thereby underlying common gene diversity or population structure parameter estimates.

References

Al-Rabbah M, Williams CG (2002) Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. For Ecol Manage 163:263–271

Auckland LD, Bui T, Zhou Y, Shepherd M, Williams CG (2002) Conifer microsatellite handbook. Corporate Press, Raleigh, N.C.

Barbéro M, Loisel R, Quézel P, Richardson DM, Romane F (1998) Pines of the Mediterranean basin. In: Richardson DM (ed) Ecology and biogeography of *Pinus*. Cambridge University Press, Cambridge, pp 153–170

Brohede J, Ellegren H (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. Proc R Soc London Ser B 266:825–833

Cooper G, Rubinsztein DC, Amos W (1998) Ascertainment bias cannot entirely account for human microsatellites being longer than their chimpanzee homologues. Hum Mol Genet 7:1425–1429

Crawford AM, Kappes SM, Paterson KA, deGotari MJ, Dodds KG, Freking BA, Stone RT, Beattie CW (1998) Microsatellite evolution: testing the ascertainment bias hypothesis. J Mol Evol 46:256–260

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: Version II. Plant Mol Biol Rep 1:19–21

Devey ME, Jermstad KD, Tauer CG, Neale DB (1991) Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor Appl Genet 83:238–242

Devey ME, Sewell MM, Uren TL, Neale DB (1999) Comparative mapping in loblolly and radiata pine using RFLP and microsatellite markers. Theor Appl Genet 99:656–662

Echt CS, Deverno LL, Anzidei M, Vendramin GG (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol Ecol 7:307–316

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

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet 16:551–558

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite evolution: directionality or bias in locus selection. Nat Genet 11:360–362

Elsik CG, Williams CG (2001) Families of clustered microsatellites in a conifer genome. Mol Genet Genomics 265:535–542

Fallour D, Fady B, Lefevre F (1997) Study on isozyme variation in *Pinus pinea* L.: Evidence for low polymorphism. Silvae Genet 46:201–207

Frankis MP (1993) Morphology and affinities of *Pinus brutia*. In: Ministry of Forestry (ed) Int Symp *Pinus brutia* Ten. Ministry of Forestry, Ankara, pp 11–18

Geada-López G, Kamiya K, Harada K (2002) Phylogenetic relationships of *Diploxylon* pines (subgenus *Pinus*) based on plastid sequence data. Int J Plant Sci 163:737–747

Gómez A, Aguiriano E, Alía R, Bueno MA (2002) Análisis de los recursos genéticos de *Pinus pinea* L. en España mediante microsatélites del cloroplasto. Invest Agrar Sist Recur For 11:145–154

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586

Joyner KL, Wang X-R, Johnston JS, Price HJ, Williams CG (2001) DNA content for Asian pines parallels New World relatives. Can J Bot 79:179–191

Karhu A (2001) Evolution and applications of pine microsatellites. University of Oulu, Oulu, Finland

Karhu A, Dieterich JH, Savolainen O (2000) Rapid expansion of microsatellite sequences in pines. Mol Biol Evol 17:259–265

Keys RN, Autino A, Edwards KJ, Fady B, Pichot C, Vendramin GG (2000) Characterization of nuclear microsatellites in *Pinus halepensis* Mill. and their inheritance in *Pinus halepensis* and *Pinus brutia* Ten. Mol Ecol 9:2157–2159

Klaus W (1989) Mediterranean pines and their history. Plant Syst Evol 162:133–163

Kostia S, Varvio SL, Vakkari P, Pulkkinen P (1995) Microsatellite sequences in a conifer, *Pinus sylvestris*. Genome 38:1244–1248

Krupkin AB, Liston A, Strauss SH (1996) Phylogenetic analysis of the hard pines (subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. Am J Bot 83:489–498

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

Liston A, Robinson WA, Piñero D, Álvarez-Buylla ER (1999) Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phylogenet Evol 11:95–109

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

Meksem K, Njiti VN, Banz WJ, Iqbal MJ, Kassem MyM, Hyten DL, Yuang J, Winters TA, Lightfoot DA (2001) Genomic regions that underlie soybean seed isoflavone content. J Biomed Biotechnol 1:38–44

Miller C (1977) Mesozoic conifers. Bot Rev 43:217–280

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. Proc Natl Acad Sci USA 94:10745–10749

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

Primmer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. Mol Biol Evol 15:997–1008

Prus-Glowacki W, Stephan BR (1994) Genetic variation of *Pinus sylvestris* from Spain in relation to other European populations. Silvae Genet 43:7–14

Raymond M, Rousset F (1995a) An exact test for population differentiation. Evolution 49:1280–1283.

Raymond M, Rousset F (1995b) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

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

Soranzo N, Provan J, Powell W (1998) Characterization of microsatellite loci in *Pinus sylvestris* L. Mol Ecol 7:1260–1261

Soranzo N, Alía R, Provan J, Powell W (2000) Patterns of variation at mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. Mol Ecol 9:1205–1211

Taylor JS, Durkin JM, Breden F (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. Mol Biol Evol 16:567–572

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

Zhu Y, Queller DC, Strassmann JE (2000) A phylogenetic perspective of sequence evolution in microsatellite loci. J Mol Evol 50:324–338