Theor Appl Genet (2004) 109:361–369
DOI 10.1007/s00122-004-1635-7

ORIGINAL PAPER

Cherdsak Liewlaksaneeyanawin · Carol E. Ritland · Yousry A. El-Kassaby · Kermit Ritland

Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs

Received: 14 July 2003 / Accepted: 9 February 2004 / Published online: 26 March 2004

Abstract Microsatellites, or simple sequence repeats (SSRs), are usually regarded as the "markers of choice" in population genetics research because they exhibit high variability. The development cost of these markers is usually high. In addition, microsatellite primers developed for one species often do not cross-amplify in related species, requiring separate development for each species. However, microsatellites found in expressed sequence tags (ESTs) might better cross-amplify as they reside in or near conserved coding DNA. In this study, we identified 14 *Pinus taeda* (loblolly pine) EST-SSRs from public EST databases and tested for their cross-species transferability to *P. contorta* ssp. *latifolia*, *P. ponderosa*, and *P. sylvestris.* As part of our development of a *P. contorta* microsatellite set, we also compared their transferability to that of 99 traditional microsatellite markers developed in *P. taeda* and tested on *P. contorta* ssp. *latifolia.* Compared to traditional microsatellites, EST-SSRs had higher transfer rates across pine species; however, the level of polymorphism of microsatellites derived from ESTs was lower. Sequence analyses revealed that the frequencies of insertions/deletions and base substitutions were lower in EST-SSRs than in other types of microsatellites, confirming that EST-SSRs are more conserved than traditional SSRs. Our results also provide a battery of 23 polymorphic, robust microsatellite primer pairs for lodgepole pine.

Communicated by O. Savolainen

C. Liewlaksaneeyanawin (✉) · C. E. Ritland · Y. A. El-Kassaby · K. Ritland
Department of Forest Sciences,
University of British Columbia,
2424 Main Mall, Vancouver, BC V6T 1Z4, Canada
e-mail: cliewlak@interchange.ubc.ca
Tel.: +1-604-8225609
Fax: +1-604-8229102

Introduction

In the past several years, the advent of molecular markers such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), and microsatellites [or simple sequence repeats (SSRs)] have helped greatly in population genetic studies in areas such as gene diversity, mating systems, and gene mapping. Microsatellites are often regarded as the "marker of choice", being codominant and showing high variability but requiring significant investment to develop, as primer pairs specific to the microsatellite locus must be designed. Microsatellite markers can be developed in several ways: via genomic libraries, enriched genomic libraries, BAC/YAC libraries, and cDNA libraries (Scott 2001). However, reliable microsatellites are difficult to develop for conifer species due to their large genome size and the extensive repetitive nature of their DNA (Kinlaw and Neale 1997). Often, complex banding patterns (multiple loci) are obtained because of duplications, and null alleles are more frequent due to variation at primer binding sites.

Several approaches have been applied to eliminate highly repetitive DNA in conifer libraries. Elsik and Williams (2001) developed microsatellites in loblolly pine using a low-copy-enrichment method. In their study, they suggested that low-copy microsatellites provided more polymorphic and informative markers than total genomic microsatellites. Zhou et al. (2002) also reported an alternative method of microsatellite development in loblolly pine. Using a methylation-sensitive restriction enzyme (*Mcr*BC), they developed microsatellites from a library of undermethylated (UM) DNA. Although this method only eliminated some of the highly repetitive DNA and provided a single-locus inheritance microsatellite, the level of polymorphisms for UM microsatellites was lower than that of low-copy microsatellites.

Scotti et al. (2000) developed six microsatellite loci by screening a cDNA library in *Picea abies* for repeats. They reported that four of the six expressed sequence tag microsatellites (EST-SSRs) that they found provided clear banding patterns and a high level of polymorphism. This

suggests that EST sequences obtained from EST databanks might be a good source for microsatellites. As they reside in or near coding DNA, they should be more conserved than genomic sequences, allowing cross-species transferability and a lower frequency of null alleles, and they should more likely appear as a single copy in the genome, alleviating the multiple-band problem. Indeed, microsatellites from EST sequences have been recently identified in some plant species, including rice (Cho et al. 2000), grapes (Scott et al. 2000), sugarcane (Cordeiro et al. 2001), and rye (Hackauf and Wehling 2002).

Lodgepole pine (*Pinus contorta* Dougl. ex. Loud. ssp. *latifolia* Engelm.) is the most important commercial pine species in British Columbia and is widely distributed throughout the Rocky Mountain and Pacific Coast regions. Besides being an important commercial species, *P. contorta* is important in watershed management, wildlife habitat, and provides recreational and scenic value in many national parks and wilderness areas (Koch 1996). Understanding genetic history and variation in natural and domesticated populations is important for developing rational conservation and tree-improvement strategies. The use of molecular markers as genetic tools can help in providing information needed to achieve these purposes. Until now there are only five microsatellite loci developed in *P. contorta* (Hicks et al. 1998).

In this study, we report on the development of loblolly pine (*Pinus taeda*) EST-SSR markers and their cross-species transferability to *P. contorta* ssp. *latifolia* and other pines. We also investigated the sequence variation and possible evolution at EST-SSR and traditional microsatellite loci for the focal and nonfocal species. The sequence analyses tested whether (1) microsatellites derived from EST libraries are more conserved than those derived from other libraries, and (2) the polymorphism detected in *Pinus* spp. is the result of expansion or contraction of the microsatellite motif itself or is due to an accumulation of sequence differences coupled with insertion/deletion events in the flanking regions outside the SSR domains. Our results also provide a battery of 23 polymorphic, robust microsatellite primer pairs for *P. contorta*.

Materials and methods

Source of microsatellites

P. taeda EST databases

A total of 55,000 *P. taeda* EST sequences were obtained from Dr. David Neale (University of California Davis, http://dendrome.ucdavis.edu/Gen_res.htm). Only dinucleotides of eight or more repeats and trinucleotides of six or more repeats were selected for this study. Sequences with sufficient repeat lengths and flanking regions on both sides were selected for primer design. A total of 14 primer pairs (LOP1–LOP14) were designed (nine dinucleotide repeats and five trinucleotide repeats) using Oligo 6.0 software (LifeScience Software Resource, Minn.) (Table 1).

P. taeda microsatellites

Ninety-nine polymorphic markers were selected from three different sources: seven genomic SSR loci, 56 low-copy SSR loci, and 36 undermethylated SSR loci. One *P. taeda* (PtTX) marker (PtTX2146) derived from a total genomic library amplified the same marker as EST-SSR locus RPTest9 (Elsik et al. 2000). From this point on, we will treat this locus as an EST-SSR. The lists of PtTX microsatellite series can be found in the Conifer Microsatellite Handbook (Auckland et al. 2002).

Cross-species transferability of microsatellites

Eight *P. taeda* EST-SSRs (Table 2) and 99 PtTX SSRs were used to examine the cross-species transferability of microsatellites on four to eight individuals of *P. contorta*. *P. taeda* EST-SSRs were also used to test cross-species transferability on two pine species: Ponderosa pine (*P. ponderosa* Dougl.), represented by two individuals and Scots pine (*P. sylvestris* L.), represented by three individuals (Table 3). First, presence or absence of microsatellite PCR products was scored on 2% agarose gels. When microsatellite fragments were found to be present, they were tested for polymorphism on 6% (Long Ranger) polyacrylamide gels using a LiCor 4200 automated sequencer (LiCor, Lincoln, Neb.). Microsatellite products were detected by M13 tailed primer (Oetting et al. 1995) or end-labelled primer. To determine the success of cross-species transferability of *P. taeda* SSR markers, the three classes of microsatellite markers were classified according to the transfer criteria described by Shepherd et al. (2002) as follows: (1) polymorphic, (2) monomorphic, and (3) polymorphic but with low product yield or non-specific amplification.

Table 1 Primer sets of 14 *Pinus taeda* L. expressed sequence tag microsatellite (EST-SSR) markers. T_a Annealing temperature, derived from the Oligo 6.0 software program

Locus	Forward primer (5′–3′)	Reverse primer (5′–3′)	T_a (°C)	Repeat motif	Accession number
LOP1	GGCTAATGGCCGGCCAGTGCT	GCGATTACAGGGTTGCAGCCT	55	$(TA)_{10}$	AI812473
LOP2	GTCTCCAGCCAGTTCACCTGC	CTTCACCACGTAGGCCCGCTC	55	$(TA)_9$	AW010960
LOP3	GTCTCCAGCCAGTTCACCTGC	CAGTGGATCTGTCACCTCCTC	48	$(TA)_9$	AA556662
LOP4	GCCTCATCATATGAAAAGCAA	CATTGTTCTCACTACGAATGC	49	$(TA)_{20}$	AW888197
LOP5	AGCCGTAAAAGCTATCTTGTG	GGCATACTTACATTTTAATAA	45	$(TA)_{33}$	AW758812
LOP11	CCAGAAGGCTATAGTACAC	CAACAATACAAGTAGCAATAC	45	$(TA)_2T(TA)_{12}$	AA739689
LOP12	AGGACAGTCCTTACTGCCCAA	CATGTTTTCCCATGGTTTTCC	45	$(TA)_{26}$	AW888197
LOP13	GGCTGGAAAGTGGTCTTTGTT	ACATAAAATGCATAATAAACG	46	$(TA)_{21}$	BE187296
LOP14	GGTCCATCTGGTTATATATTG	AGGAATTTCGCCACTTCACTG	50	$(TA)_{12}$	BF517779
LOP6	AGTTTTATCCATGCTGCACAG	ACCTAAAGCCCAATATCCACA	55	$(AAT)_7$	AA556221
LOP7	CGGGGAATTGATAGTGTG	TCATCGTCCTCAGCTGCAAGT	54	$(AAG)_6$	AW981642
LOP8	TATCCACCAGAAGGGCATC	CGGGAGCTTTAATGATCTTGA	50	$(CCT)_6$	AI725303
LOP9	GGATTCTCGTTGTGGCTGG	TTGCCTTTGCACATAATATCT	55	$(GGC)_6$	AI813163
LOP10	CTCTCCTCCGGCTATTTGCAG	CGGCGAAGCTCTTCATTCCT	59	$(AAG)_6$	BE520122

Table 2 Results of transferable microsatellite loci from *P. taeda* L. to *P. contorta* ssp. *latifolia*

Marker	Repeat motif	Source of library[a]	T_a	Allele size (bp)		Quality class[d]
				P. taeda[b]	*P. contorta*[c]	
LOP1	$(TA)_{10}$	EST	55	161	153	1
LOP5	$(TA)_{33}$	EST	45	209	168	1
LOP8	$(CTT)_6$	EST	50	369	370	1
LOP9	$(GCC)_6$	EST	55	142	131	1
LOP11	$(TA)_2T(AT)_{12}$	EST	45	254	243	1
LOP12	$(TA)_{26}$	EST	45	191	154	1 (N)
PtTX2146	$(GCT)_4GCC(GCT)_7GCC(GCT)_8$	EST	55	180	196	1
LOP3	$(TA)_9$	EST	48	220	209	2
LOP4	$(TA)_{20}$	EST	49	207	208	3 (M)
LOP6	$(AAT)_7$	EST	55	228	259	3 (M)
LOP14	$(TA)_{12}$	EST	50	300	320	3 (M)
PtTX2123	$(AGC)_8$	G	55	202	200	1
PtTX2128	$(GAC)_8$	G	55	245	228	1
PtTX2183	$(CAA)_{18}$	G	55	205	100	3 (M)
PtTX3011	$(GAA)_5(A)_6(GAA)_3 \ldots (GAT)_{15}$	LC	55	186	178	1
PtTX3025	$(CAA)_{10}$	LC	59	266	266	1
PtTX3029	$(GCT)_5 \ldots (GCT)_8 \ldots (GCT)_5$	LC	61	255	274	1 (M)
PtTX3030	$(TA)_4 \ldots (GGT)_{10}$	LC	59	327	320	1
PtTX3034	$(GT)_{10}(GA)_{13}$	LC	55	207	207	1
PtTX3049	$(TG)_{16}$	LC	55	311	302	1
PtTX3052	$(ATC)_8 \ldots (ATC)_4$	LC	55	242	239	1
PtTX3107	$(CAT)_{14}$	LC	55	182	177	1
PtTX3127	$(CAA)_{10}$	LC	55	183	187	1
PtTX2003	$(ACC)_8$	LC	61	122	125	2
PtTX2082	$(GT)_{14}(GAGT)_7(GA)_{13}$	LC	61	208	200	2
PtTX3002	$(GAG)_6 \ldots (GAG)_4AA(GAG)_4$	LC	65	194	191	2
PtTX3091	$(GTT)_{10}T_{13}GGT_{10}CT_5$	LC	64	229	172	2
PtTX2034	$(TTTG)_9$	LC	55	170	123	3 (M)
PtTX2037	$(GTGA)_8GT_{14}$	LC	58	177	145	3 (N)
PtTX3020	$A_{16}(CAA)_9$	LC	61	183	183	3 (M)
PtTX3045	$(CA)_{12}$	LC	55	226	210	3 (M)
PtTX3055	$(GAT)_5 \ldots (GAT)_8 \ldots (GAT)_6$	LC	59	402	375	3 (N)
PtTX3090	$(CAC)_4(CAT)_{24}CAC(CAT)_{11}$	LC	57	259	285	3 (W)
PtTX3098	$(GTT)_8$	LC	61	187	185	3 (M)
PtTX3105	$(GTT)_9$	LC	55	258	170	3 (N)
PtTX3116	$(TTG)_7 \ldots (TTG)_5$	LC	55	146	122	3 (N)
PtTX4046	$(TA)_3(TG)_{13}$	UM	55	363	331	1
PtTX4054	$(GA)_{21}$	UM	55	179	292	1
PtTX4056	$(GA)_{17}$	UM	65	436	409	1
PtTX4058	$(CA)_3(GA)_{20}$	UM	55	188	158	1
PtTX4139	$(CT)_{21}$	UM	59	153	113	1
PtTX4009	$(CA)_3TA(CA)_{14}$	UM	63	280	252	2
PtTX4050	$(CA)_6 \ldots (CA)_3$	UM	57	188	185	2
PtTX4090	$(CTT)_6 \ldots (CTT)_8$	UM	55	188	183	2
PtTX4112	$(AT)_6(GT)_6$	UM	57	463	424	2
PtTX4146	$(GAA)_8$	UM	59	126	115	2
PtTX4004	$(GT)_{14} \ldots (T)_4(GT)_2(T)_7$	UM	55	175	168	3 (M)
PtTX4092	$(GAA)_{21}$	UM	57	162	150	3 (M)
PtTX4100	$(GAA)_8$	UM	61	207	210	3 (W)
PtTX4137	$(GAA)_{21}$	UM	61	139	145	3 (M)

[a] Source of microsatellite markers: *EST* expressed sequence tag, *G* genomic library, *LC* low-copy library, *UM* undermethylated library
[b] Expected size based on the Conifer Microsatellite Handbook (Auckland et al. 2002)
[c] Allele size based on most common allele
[d] Quality class of microsatellites: *1* polymorphic, *2* monomorphic, *3* poor amplification, *M* multiple alleles, *N* null alleles, *W* weak amplification

DNA isolation and polymerase chain reaction optimization

Genomic DNA was isolated from individual vegetative buds or young seedlings (1 week old) using a modification of the cetyltrimethyl ammonium bromide method (Doyle and Doyle 1987). PCR reactions were carried out in 10 μl (final volume) using an MJ Research PTC-100 thermal cycler (Watertown, Mass.) based on the protocol proposed by Elsik et al. (2000) with modifications. Each reaction was composed of 50 ng total genomic DNA; 1 pmol of each primer; 0.5 mM each of dATP, dCTP, dGTP, and dTTP; 1× buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3) (Roche, Laval, QC); and 0.25 U of *Taq* DNA (Roche). When the tailed primer was used for amplification, 0.3 pmol of M13 Infrared Label (LiCor) was added to PCR reactions. Samples were amplified as follows: 5 min at 94°C, followed by 30 cycles of 1 min at 94°C, 1 min at the annealing temperature (see Auckland et al. 2002), 1 min at 72°C, followed by a long denaturation cycle of 3 min at 72°C. PCR conditions were optimized for loci having complex

Table 3 Allele size and number of alleles of seven EST-SSR loci in four different pine species

Locus	*P. taeda*		*P. ponderosa*		*P. contorta* ssp. *latifolia*		*P. sylvestris*	
	Allele size (bp)	No. of alleles	Allele size (bp)	No. of alleles	Allele size (bp)	No. of alleles	Allele size (bp)	No. of alleles
LOP1	153, 155, 157	3	155, 157	2	153, 155, 159	3	158, 162	2
LOP3	208	1	209, 215	2	209	1	209, 213, 217	3
LOP5	173, 213, 217, 241	4	174, 176, 178, 192	4	168, 178, 186, 216	4	166–190	5
LOP8	369, 375	2	370	1	370, 373	2	367	1
LOP9	135, 138	2	132	1	128, 131, 137	3	135	1
LOP11	243–269	6	243	1	243, 253, 272	3	235	1
LOP12	154, 156	2	152, 154, 156	3	150, 154, 160, 168	4	156, 160	2
PtTX2146	163, 172, 175	3	163, 178, 187	3	169, 190, 193, 196	4	181, 193, 211, 220	4

Table 4 Polymorphisms, indels (insertions/deletions), and base-substitution rates (*BSR*) in flanking regions (relative to *P. taeda* sequences) in *P. contorta* ssp. *latifolia* at eight EST-SSR loci and 16 polymorphic PtTX microsatellite loci that were successful in cross-species amplification. H_o Observed heterozygosities, H_e expected heterozygosities

Marker[a]	Allele size range (bp)[b]	No. of alleles[b]	H_o[b]	H_e[b]	Insertion[c]	Deletion[c]	BSR[c] (%)	Flanking sequence length (bp)	Change in SSR structure[c]	Accession number
LOP1	153–163	4	0.261	0.278	0	0	0.70	141	YES	AY330148
LOP3	209	1	–	–	0	1	2.00	201	No	AY330152
LOP5	166–252	17	0.826	0.946	0	1	0.65	118	No	AY330155
LOP8	370–373	2	0.217	0.198	1	0	0.84	358	No	AY330158
LOP9	128–137	3	0.043	0.273	0	0	0.00	116	No	AY330161
LOP11	243–271	14	1.000	0.908	0	0	0.44	225	No	AY330164
LOP12	150–176	10	0.333	0.899	NA[d]	NA	NA	NA	NA	NA
PtTX2146	160–208	12	0.913	0.824	0	0	2.54	117	No	AY330133
Mean		7.88	0.449	0.541			1.02	182		
PtTX2123	194–203	4	0.652	0.63	2	1	3.61	178	No	AY330131
PtTX2128	228–237	4	0.565	0.654	1	3	3.27	214	No	AY330132
PtTX3011	151–211	14	0.696	0.892	0	0	0.81	103	YES	AY330134
PtTX3025	254–305	7	0.636	0.653	0	0	0.42	236	YES	AY330135
PtTX3029	253–304	11	0.739	0.885	0	1	0.00	166	No	AY330136
PtTX3030	318–325	4	0.304	0.704	0	8	4.72	223	YES	AY330137
PtTX3034	201–221	11	0.652	0.708	0	2	0.63	159	No	AY330138
PtTX3049	300–332	11	0.696	0.885	0	0	14.11	276	No	AY330139
PtTX3052	239–254	5	0.478	0.446	1	1	2.34	199	No	AY330140
PtTX3107	156–177	6	0.348	0.760	0	0	1.43	138	No	AY330141
PtTX3127	169–202	7	0.409	0.689	0	0	0.66	158	No	AY330142
PtTX4046	324–342	6	0.500	0.766	0	6	2.90	314	No	AY330143
PtTX4054	268–302	14	0.826	0.913	0	0	0.00	254	No	AY330144
PtTX4056	427–453	10	0.864	0.864	1	2	1.51	399	No	AY330145
PtTX4058	128–160	13	0.870	0.901	0	0	1.02	100	No	AY330146
PtTX4139	113–147	9	0.739	0.844	1	0	0.93	106	No	AY330147
Mean		8.50	0.623	0.762			2.40	201		

[a] LOP series were derived from EST, and PtTX series were derived from a genomic library, a low-copy library, and an undermethylated library
[b] Based on *P. taeda* microsatellite loci that were successful in amplifying PCR products in *P. contorta* ssp. *latifolia*
[c] Based on the comparison between *P. contorta* ssp. *latifolia* and *P. taeda*
[d] *NA* No data at this locus

banding patterns or low yield by changing annealing temperature and/or primer concentration. Two microliters of stop dye buffer (LiCor) were added to each PCR reaction tube, and PCR reactions were kept at −20°C in the dark until electrophoresis.

Scoring and estimating genetic polymorphism

Polymorphisms of transferable *P. taeda* EST-SSR and PtTX SSR markers were tested on *P. contorta* ssp. *latifolia*. All genotypes were scored using SAGA software (LiCor). The allelic diversity of eight *P. taeda* EST-SSR loci and 16 polymorphic PtTX SSR loci (Table 4) were based on an evaluation of 24 individuals sampled across the natural range of *P. contorta* ssp. *latifolia*.

Sequence verification and comparison

To verify the cross-species amplification of microsatellite fragments, the amplification products of 16 polymorphic PtTX SSR markers were sequenced from *P. contorta* ssp. *latifolia*. Seven *P. taeda* EST-SSR markers (LOP nos. 1, 3, 5, 8, 9, 11, PtTX2146) were also tested for sequence variation in *P. contorta* ssp. *latifolia*, *P. ponderosa*, and *P. sylvestris*. For each of these markers, two homozygous PCR products were purified using Qiaquick PCR

amplification kits (QIAGEN, Mississauga, Ont.) and sequenced using SequiTherm EXCEL II Long-Read DNA sequencing Kits-LC (Epicentre Technologies) on a LiCor 4200 automated sequencer. DNA sequences were aligned using ESEE3S software (Cabot and Beckenbach 1989) and then edited manually. To determine similarity in flanking regions and repeat motifs, sequences of each locus were compared to the original *P. taeda* sequences from GenBank (http://www.ncbi.nlm.nih.gov/BLAST/). The frequency of base substitution in the flanking regions was calculated as a percentage of the proportion between the total number of base substitutions and the total number of base pairs in the flanking regions. The insertion/deletion (indel) was also counted by comparing the sequence data to *P. taeda*. A total of 36 sequences indicated with following accession numbers were conducted (Tables 4, 6).

Results

A total of 98 EST sequences with repeat motifs $n \geq 8$ for dinucleotides and $n \geq 6$ for trinucleotides were identified among 55,000 *P. taeda* EST sequences, of which 62 sequences contained dinucleotide motifs (AT, AG, and GT), and 36 contained trinucleotide motifs (mostly AAG and GGC) (Fig. 1).

Of the 14 *P. taeda* EST-SSRs developed, three primer pairs (LOP4, LOP6, and LOP14) produced multiple bands and four primer pairs (LOP2, LOP7, LOP10, and LOP13) were not successful in amplification of SSR products in *P. taeda*. The remaining seven primer pairs (LOP1, LOP3, LOP5, LOP8, LOP9, LOP11, and LOP12) were successful in amplification of SRR products in *P. taeda*. These seven LOP loci and one PtTX2146 locus amplified products of the expected size in all species tested (Tables 2, 3). Of the eight primer pairs that produced clear reproducible bands, four pairs (LOP1, LOP5, LOP12, and PtTX2146) were polymorphic in all species and three pairs (LOP8, LOP9, and LOP11) were polymorphic in *P. taeda* and *P. contorta* ssp. *latifolia*, but were monomorphic in *P. ponderosa* and *P. sylvestris* (Table 3). The remaining pair (LOP3) was polymorphic in *P. ponderosa* and *P. sylvestris*, but not in *P. taeda* and *P. contorta* ssp. *latifolia* (Table 3). Allele size and number of alleles detected by the EST-SSRs within the four pine species are shown in Table 3. The polymorphisms of those eight EST-SSRs tested on *P. contorta* are shown in Table 4. Allele number ranged from 1 to 17 with a mean of 7.88 per locus, while observed heterozygosities (H_o) ranged from 0.043 to 1.000 (Table 4).

Of the 99 PtTX microsatellite loci tested, 39 loci amplified products (Table 2) and the remaining 60 loci gave no amplification products in *P. contorta* ssp. *latifolia*. Twenty-five of the 39 loci produced single and clear

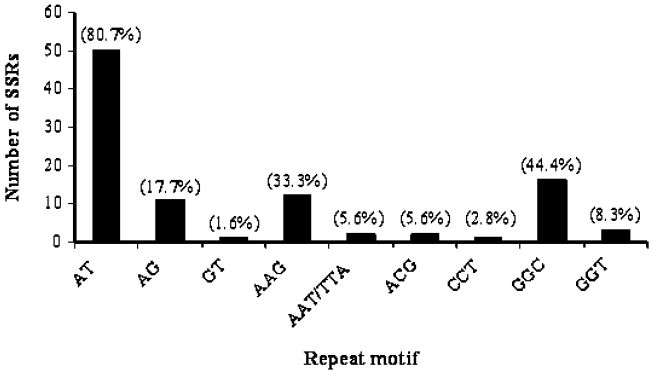

Fig. 1 Distribution of the di- and trinucleotide expressed sequence tag microsatellites (EST-SSRs). The frequency of SSRs observed within each repeat classes (%) are shown in *parentheses*

amplification products, but the remaining 14 loci produced weak, null and/or non-specific amplifications (Table 2). Of the 25 single-banded loci, 16 produced polymorphic bands and the remaining nine were monomorphic (see Table 4 for the description of these 16 polymorphic loci). The allele number ranged from 4 to 14 with a mean of 8.50 per locus, while H_o ranged from 0.304 to 0.870 (Table 4). Null alleles, observed in both EST-SSRs and PtTX SSRs (Table 2), were inferred from markers giving no amplification products and excess homozygotes.

The transferability successes and polymorphisms in *P. contorta* ssp. *latifolia* based on the sources of SSR are shown in Table 5. The success rates of transferability of EST, genomic, low-copy, and undermethylated SSRs were 100, 29, 23, and 30%, respectively, with the expected heterozygosities (H_e) from the respective sources as 0.54, 0.64, 0.74, and 0.85. However, there were no significant differences in average H_e between the sources of SSR marker (one-way ANOVA F value = 1.84, P value = 0.17, Table 5).

Sequencing of PCR products confirmed the presence of microsatellite repeats at all loci. The repeat structures of seven *P. taeda* EST-SSRs and 16 PtTX microsatellites were highly conserved in *P. contorta* ssp. *latifolia*, with the exception of LOP1, PtTX3025, PtTX3011, and PtTX3030. Single-base pair substitution caused interrupt and compound repeats in *P. contorta* ssp. *latifolia* at locus LOP1 $[(TA)_3CA(TA)_3]$ and locus PtTX3025 $[(CAA)_5(CAAAAA)_3]$, respectively. A single-base pair substitution in the repeat of the SSR at locus PtTX3011 also altered the length of poly$(A)_n$ from $(A)_6$ to $(A)_3$ $[(GAA)_6(A)_3(GAA)_3 \ldots (GAT)_{13}]$. A 54-bp insertion was

Table 5 Transferability successes and polymorphisms of SSR markers from *P. taeda* to *P. contorta* ssp. *latifolia* based on the sources of library. A one-way ANOVA indicated no significant difference in average H_e based on polymorphic markers (P-value = 0.17)

Source of library	No. of loci tested	No. of amplified loci	No. of polymorphic loci	H_e
EST	8	8	7	0.54±0.39
G	7	2	2	0.64±0.02
LC	56	13	9	0.74±0.14
UM	36	10	5	0.85±0.06

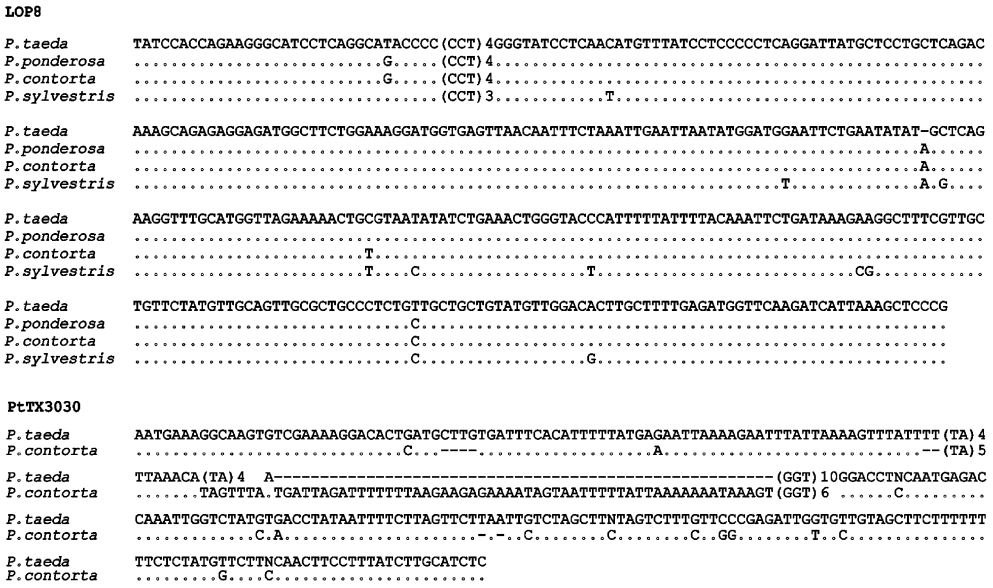

Fig. 2 Nucleotide sequence comparison of two *Pinus taeda* (PtTX) microsatellite loci (LOP8 and PtTX 3030). The *dots* indicate conserved nucleotides (relative to *P. taeda*)

Table 6 Repeat structures of SSRs and indels in the flanking regions (relative to *P. taeda* sequences) between four pine species at seven EST-SSR loci

Locus	Species	Repeat motif	Insertion	Deletion	Accession number
LOP1	*P. taeda*	$(TA)_{10}$	0	0	
	P. ponderosa	$(TA)_{10}$	0	0	AY330149
	P. contorta	$(TA)_3CA(TA)_3$	0	0	
	P. sylvestris	$(TA)_2T(TA)_7$	0	0	AY330150
LOP3	*P. taeda*	$(TA)_9$	0	0	
	P. ponderosa	$(TA)_4$	0	1	AY330151
	P. contorta	$(TA)_4$	0	1	
	P. sylvestris	$(TA)_4$	0	1	AY330153
LOP5	*P. taeda*	$(TA)_{33}$	0	0	
	P. ponderosa	$(TA)_{18}$	0	1	AY330154
	P. contorta	$(TA)_9$	0	1	
	P. sylvestris	$(TA)_{10}$	0	1	AY330156
LOP8	*P. taeda*	$(CCT)_4$	0	0	
	P. ponderosa	$(CCT)_4$	1	0	AY330157
	P. contorta	$(CCT)_4$	1	0	
	P. sylvestris	$(CCT)_3$	1	0	AY330159
LOP9	*P. taeda*	$(GGC)_6$	0	0	
	P. ponderosa	$(GGC)_5$	0	0	AY330160
	P. contorta	$(GGC)_5$	0	0	
	P. sylvestris	$(GGC)_6$	0	0	AY330162
LOP11	*P. taeda*	$(TA)_2T(TA)_{12}$	0	0	
	P. ponderosa	$(TA)_2TTATG(TA)_4$	0	0	AY330163
	P. contorta	$(TA)_2T(TA)_7$	0	0	
	P. sylvestris	NA[a]	NA	NA	NA
PtTX2146	*P. taeda*	$(GCT)_4GCC(GCT)_7GCC(GCT)_8$	0	0	
	P. ponderosa	$(GCT)_4GCC(GCT)_5GCC(GCT)_{12}$	0	0	AY489266
	P. contorta	$(GCT)_4GCC(GCT)_{10}GCC(GCT)_{10}$	0	0	
	P. sylvestris	$(GCT)_3CCT(GCT)_8CCT(GCT)_{15}$	0	0	AY489267

[a] *NA* No sequence data available

observed between the compound microsatellite repeat of PtTX3030 in *P. contorta* ssp. *latifolia* (Fig. 2). The presences of indels in the flanking regions of *P. contorta* ssp. *latifolia* (compared to the *P. taeda* sequences) were lower for EST-SSRs (43%) than for PtTX microsatellite markers (56%) (Table 4). In addition, EST-SSRs seem to have a small number of base pairs inserted or deleted when compared to PtTX microsatellite markers. The average base substitution for each locus was 1.02 and 2.40 per 100 bp in EST-SSRs and PtTX microsatellite markers, respectively (Table 4). A *t*-test indicated no significant difference in the average base substitution between two sources of microsatellite markers (two-tailed *t*-tests, *t* value of −1.49, *P* value of 0.15). The lengths of flanking sequences of EST-SSRs and PtTX microsatellites are shown in Table 4 indicating that frequency of indels and base substitution are not dependent on the flanking sequence length.

PCR products at seven EST-SSR loci from *P. ponderosa* and *P. sylvestris* were also sequenced. Table 6

Table 7 The percentage of base substitution in the flanking regions between species

Species	Locus	*P. ponderosa*	*P. contorta*	*P. sylvestris*
P. taeda	LOP1	0.70	0.70	0.70
	LOP3	1.52	2.00	1.52
	LOP5	0.65	0.65	1.31
	LOP8	0.56	0.84	2.80
	LOP9	0.00	0.00	0.00
	LOP11	0.88	0.44	NA[a]
	PtTX2146	2.54	2.54	4.24
P. ponderosa	LOP1		1.40	1.40
	LOP3		0.51	0.00
	LOP5		0.00	0.65
	LOP8		0.28	2.80
	LOP9		0.00	0.00
	LOP11		0.44	NA
	PtTX2146		0.00	1.69
P. contorta	LOP1			1.40
	LOP3			0.51
	LOP5			0.65
	LOP8			2.51
	LOP9			0.00
	LOP11			NA
	PtTX2146			1.69

[a] No sequence data available

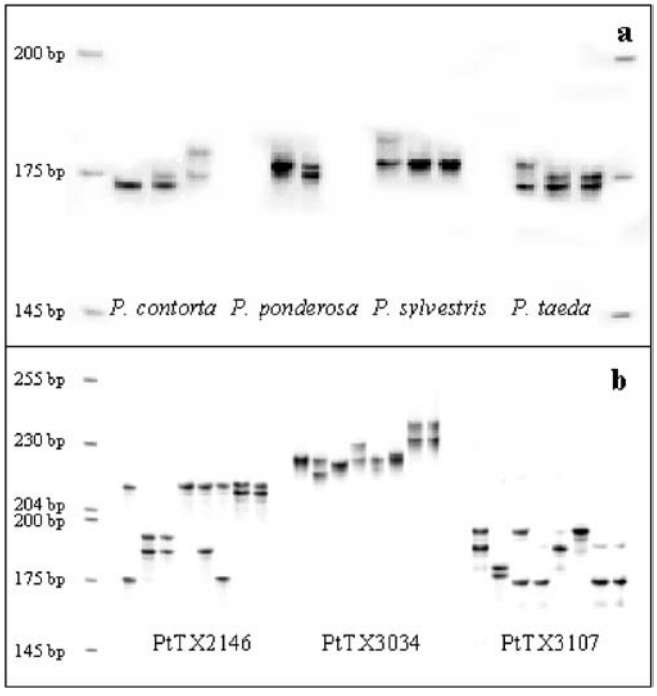

Fig. 3 **a** Transferability of *P. taeda* EST-SSR at locus LOP1 on different pine species. **b** An example of cross-transferability of microsatellite markers from *P. taeda* to *P. contorta* at one EST-SSR marker (PtTX2146) and two PtTX markers (PtTX3034 and PtTX3107). Note all allelic sizes have Licor primer tails

represents the repeat structure and the presence of indels (compared to the *P. taeda* sequences). Table 7 represents pairwise comparison of the percentage of base substitution in the flanking regions between pine species tested. The average base substitution across loci of each species were 0.98, 1.02, and 1.76 per 100 bp for *P. ponderosa*, *P. contorta* ssp. *latifolia*, and *P. sylvestris*, respectively (compared to the *P. taeda* sequences). The lowest average base substitution across loci was observed between *P. ponderosa* and *P. contorta* ssp. *latifolia* (0.20 per 100 bp). There were no changes in the flanking sequences observed at locus LOP9. Size homoplasy was observed at locus LOP3 and LOP8. At locus LOP3 *P. ponderosa*, *P. contorta* ssp. *latifolia*, and *P. sylvestris* shared the same allele sizes at 209 bp; however, a single-base pair substitution in the flanking region (T→G) was observed in *P. contorta* ssp. *latifolia*. Similarly, an identical allele size of 370 bp was observed at locus LOP8 in *P. ponderosa* and *P. contorta* ssp. *latifolia*, but sequencing revealed a single-base pair substitution in the flanking region (C→T) in *P. ponderosa* (Fig. 2).

Discussion

Prevalence of microsatellites in ESTs

Among dimeric repeats, the motifs AT (80.6%) and AG (17.7%) were most common in databases, whereas TG motifs were rarely detected (1.6%) (Fig. 1). In contrast to other crop species, the AT motifs were the rarest (Temnykh et al. 2000; Thiel et al. 2003). Among trimeric repeats, AAG (33.3%) and GGC (44.4%) were the most abundant (Fig. 1). Similarly, the motifs AAG and GGC appeared common in *Arabidopsis* (Cardle et al. 2000), and GGC was most abundant in maize (Chin et al. 1996). However, it is difficult to compare the prevalence of a particular motif in different plant species due to a difference in minimal motif-repeat criteria.

Transferability of *P. taeda* microsatellites

Most primers amplified genomic regions of the expected size in the majority of species tested (Table 2). The reason for this can be explained by the conservation of the sequence repeats and flanking regions in a majority of pine species. *P. taeda* EST-SSRs had a high transfer rate (100%) across different subsections of pine species (Fig. 3a). In this study, the success rate of transferability of EST-SSRs was also higher than that of other types of SSRs in *P. contorta* (Table 5). Similarly, the ability of transferring EST-SSRs among closely related genera has been reported in crop species (Chen et al. 2002; Cordeiro et al. 2001; Decroocq et al. 2003). In contrast, the level of transferability of genomic and low-copy SSRs observed in this study was lower than that of transferability from *P. taeda* to *P. elliottii* var. *elliottii* and *P. caribaea* var. *hondurensis* (Shepherd et al. 2002). This result can be explained by the fact that the transferability success reduces as the evolutionary distance between the source and target species increases. *P. elliottii* var. *elliottii* and *P. caribaea* var. *hondurensis* are found to be more closely

related to *P. taeda* than *P. contorta* ssp. *latifolia* (Little and Critchfield 1969).

EST-SSR markers with trinucleotide repeats were less polymorphic than dinucleotide repeats and had low polymorphisms when compared to markers from other sources. Similarly, the low variability of trimeric EST-SSR loci was reported in *Oryza sativa* L. (Cho et al. 2000). Dimeric EST-SSR markers with high numbers of repeats seem to have high polymorphisms, as do genomic SSRs. The relationship between polymorphisms and the number of repeats has been reported for dimeric, trimeric and tetrameric EST-SSRs in barley (*Hordeum vulgare* L.) (Thiel et al. 2003). However, the analysis of a large set of primer pairs will be required to confirm this result, especially for trimeric and tetrameric EST-SSRs. On average, the level of genetic polymorphisms was lower for EST-SSRs than for other types of microsatellite (Table 5; Fig. 3). The results of this study were similar to previous studies in rice (*O. sativa* L.) and barley (*H. vulgare* L.). In rice, microsatellites derived from ESTs had a lower level of polymorphism than those derived from genomic library (83.8 vs 54.0) (Cho et al. 2000). The mean level of polymorphism is also lower for EST-SSRs (0.45) than for genomic SSRs (0.58) in barley (Thiel et al. 2003).

Successful cross-amplification of microsatellites also has been reported among pine species that diverged over 100 million years ago (Kutil and Williams 2001). In their study, Kutil and Williams (2001) found that microsatellites from hard pine (*P. taeda* L.) had trans-specific amplification in both hard and soft pines. They also suggested that perfect trinucleotide-repeat SSRs seem to cross-amplify better than do compound SSRs.

Analysis of sequence variation at the microsatellite loci

Sequence analyses confirmed that the level of conservation in the microsatellite motif and flanking regions was higher for EST-SSRs than for other types of microsatellites. Most of the variation in allele length among pine species was mainly due to changes in the number of repeat motifs in the microsatellite region, combined with indel and base substitution in flanking regions. The percentage of the base substitution of EST-SSRs was not strongly supported the phylogenetic study in pine species by Krupkin et al. (1996). *P. taeda* and *P. ponderosa* were close relatives, and *P. contorta* emerged as a sister group to this pair and more closely related to *P. ponderosa*. *P. sylvestris* was highly differentiated as an outgroup. However, our results show that the lowest base substitution across loci was observed between *P. ponderosa* and *P. contorta* ssp. *Latifolia*. This result is not in concordance with the previous pine phylogeny because EST-SSRs are short and highly conserved sequences and, thus, might not differentiate the phylogenetic relationship in closely related pine species. However, there was a tendency of the base substitution to increase when genetic distance between species increased, such as in locus LOP8. Similar results to our study were observed by Karhu et al. (2000), who studied the evolution of microsatellites in pines.

Functional roles of EST microsatellites

There have been few reports of the functional roles of microsatellites located near or within coding regions in plants. However, the typical example of the characteristics of microsatellites in regulatory genes has been well conducted in yeast (*Saccharomyces cerevisiae*). In the yeast genome mono-, di-, and tetranucleotide repeats were located primarily in adjacent regions (intergenic regions), whereas trinucleotide repeats were often found in open reading frames (ORFs) and were related to cellular regulation (Richard and Dujon 1996; Young et al. 2000). Young et al. (2000) reported that certain types of trinucleotide repeats were overrepresented in ORFs and encoded a biased set of amino acids. They suggested that negative selection might act against certain trinucleotide repeats at different levels. Recent study in the feature of microsatellites within transcribed regions of rice and *Arabidopsis* was reported by Fujimori et al. (2003). In their study, microsatellites in the transcribed regions of rice and *Arabidopsis* were frequently found in the 5′UTRs than in coding regions or 3′UTRs, suggesting that they can potentially act as factors in regulating gene expression. The *waxy* gene of rice containing GA/CT repeats in the 5′UTR was found to be associated with amylose content (Ayres et al. 1997). Microsatellites $(CCG)_n$ in 5′UTRs of some ribosomal protein genes of maize might be related to the regulation of fertilization (Dresselhaus et al. 1999).

Implications for forest genetic research

Published ESTs and SSR markers from related species proved to be valuable resources for SSR marker development in *P. contorta*. The transferability success reduces as the evolutionary distance between the source and target species increases. According to phylogenetic relationship in pines (Krupkin et al. 1996), it can be expected that transferability success of PtTX microsatellites should be higher in *P. ponderosa* than in *P. sylvestris*. Our 16 polymorphic PtTX microsatellites are a good source for testing cross-species transferability in particular for genetic study in *P. ponderosa*. Our results also suggest that the level of polymorphisms in *Pinus* EST-SSRs also might depend on the number of repeats, and EST sequences with dinucleotide repeats of ten or more can be useful for the development of informative microsatellites in *Pinus* spp. Although EST-SSRs have low levels of polymorphisms, their ability to amplify across species/genera is high, and they might be associated with genes of known function. The variability of flanking-region sequences of SSRs can be used to elucidate the genetic history of pine species from the standpoint of evolution. In addition, the analysis of these microsatellites will

provide an important new tool for addressing questions related to conservation and tree improvement programs. Many parameters, such as genetic diversity in natural and breeding populations, gene flow, pollen and/or seed dispersal, and mating systems, are important for the conservation of genetic resources. In tree improvement programs, microsatellites can be used for QTL mapping, clone identification, estimating pollen flow/contamination, and determining male parentage of seeds produced in seed orchards.

Acknowledgements This work was funded through NSERC grants to K.R. We sincerely thank D.B. Neale for providing loblolly pine EST sequences, and C.H. Newton for contributing the DNA of 24 individual lodgepole pines. We also would like to thank two anonymous reviewers and O. Savolainen for helpful comments on the manuscript.

References

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

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a sin1gle-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences with Esee. Comput Appl Biosci 5:233–234

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

Chen X, Cho YG, McCouch SR (2002) Sequence divergence of rice microsatellites in *Oryza* and other plant species. Mol Genet Genomics 268:331–343

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866–873

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

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

Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape microsatellite markers across taxa. Theor Appl Genet 106:912–922

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh tissue. Phytochem Bull 19:11–15

Dresselhaus T, Cordts S, Heuer S, Sauter M, Lorz H, Kranz E (1999) Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. Mol Gen Genet 261:416–427

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

Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S, Tomita M (2003) A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Lett 554:17–22

Hackauf B, Wehling P (2002) Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed 121:17–25

Hicks MD, Adams D, O'Keefe S, Macdonald E, Hodgetts R (1998) The development of RAPD and microsatellite markers in lodgepole pine (*Pinus contorta* var. *latifolia*). Genome 41:797–805

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

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

Koch P (1996) Lodgepole pine in North America. Forest Product Society, Madison, Wis.

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

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

Little EL Jr, Critchfield WB (1969) Subdivisions of the genus *Pinus* (pines). USDA For Serv Misc Publ 1144, Washington, D.C.

Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA, King RA (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30:450–458

Richard GF, Dujon B (1996) Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. Gene 174:165–174

Scott KD (2001) Microsatellites derived from ESTs and their comparison with those derived from by other methods. In: Henry RJ (ed) Plant genotyping: the DNA fingerprinting of plants. CABI, New York, pp 225–237

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

Shepherd M, Cross M, Maguire TL, Dieters MJ, Williams CG, Henry RJ (2002) Transspecific 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, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR markers in the barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Young ET, Sloan JS, Van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. Genetics 154:1053–1068

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