Theor Appl Genet (2004) 108:873–880
DOI 10.1007/s00122-003-1490-y

ORIGINAL PAPER

G. T. Slavov · G. T. Howe · I. Yakovlev · K. J. Edwards · K. V. Krutovskii · G. A. Tuskan · J. E. Carlson · S. H. Strauss · W. T. Adams

Highly variable SSR markers in Douglas-fir: Mendelian inheritance and map locations

Received: 17 March 2003 / Accepted: 25 September 2003 / Published online: 19 November 2003

Abstract Twenty-two highly variable SSR markers were developed in Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] from five SSR-enriched genomic libraries. Fifteen PCR primer pairs amplified a single codominant locus, while seven primer pairs occasionally amplified two loci. The Mendelian inheritance of all 22 SSRs was confirmed via segregation analyses in several Douglas-fir families. The mean observed heterozygosity and the mean number of alleles per locus were 0.855 (SE=0.020) and 23 (SE=1.6), respectively. Twenty markers were used in genetic linkage analysis and mapped to ten known linkage groups. Because of their high polymorphism and unambiguous phenotypes, 15 single-locus markers were selected as the most suitable for DNA fingerprinting and parentage analysis. Only three SSRs were sufficient to achieve an average probability of exclusion from paternity of 0.998 in a Douglas-fir seed orchard block consisting of 59 parents.

Communicated by O. Savolainen

G. T. Slavov (✉) · G. T. Howe · S. H. Strauss · W. T. Adams
Department of Forest Science,
Oregon State University,
Corvallis, OR 97331-5752, USA
e-mail: gancho.slavov@orst.edu
Fax: +1-541-7371393

I. Yakovlev
Norwegian Forest Research Institute (Skogforsk),
Hogskoleveien 12, 1432 As, Norway

K. J. Edwards
Functional Genomics Group, OB120,
School of Biological Sciences,
University of Bristol,
Bristol, BS8 1UG, UK

K. V. Krutovskii
Institute of Forest Genetics, USDA Forest Service,
Pacific Southwest Research Station,
Davis, CA 95616, USA

G. A. Tuskan
Oak Ridge National Laboratory,
MS-6422, P.O. Box 2008, Oak Ridge, TN 37830, USA

J. E. Carlson
School of Forest Resources,
Pennsylvania State University,
University Park, PA 16802-5805, USA

Introduction

Highly polymorphic genetic markers such as simple sequence repeats (SSRs) can radically improve the precision of pollen contamination and gene flow estimates. We are developing a paternity exclusion procedure to measure pollen contamination in seed orchards of Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] using SSRs. This class of markers has been used to infer paternity and estimate gene flow through genotypic exclusion in a number of tree species. In each case, only a few SSR loci (4–6) were needed to achieve high exclusionary power (Dow and Ashley 1998; Streiff et al. 1999; Lian et al. 2001). Unfortunately, the development of SSR markers is still inefficient, time-consuming, and resource-intensive (Zane et al. 2002), particularly in organisms with large and complex genomes, such as conifers. Many attempts to develop SSR markers for conifers have yielded just a handful of useful marker loci (Pfeiffer et al. 1997; Hicks et al. 1998; Soranzo et al. 1998).

Fifty SSR markers for Douglas-fir were reported by Amarasinghe and Carlson (2002). We characterized 22 additional markers, 15 of which produce robust banding patterns and segregate as single codominant loci. We show that this set of 15 single-locus SSRs is a valuable tool for genotype identification, parentage analysis, and genome mapping. The remaining seven SSRs can be used for certain applications, but additional optimization of PCR conditions is needed to obtain clear, single-locus banding patterns for all samples.

Materials and methods

Plant materials and DNA extraction

DNA used to construct genomic libraries was extracted from the needles of a single Douglas-fir seedling using the DNeasy Plant Maxi Kit (QIAGEN, Valencia, Calif.). Seeds for segregation analysis were collected from 18 putatively unrelated Douglas-fir trees growing in a grafted seed orchard in western Oregon, USA. Haploid (1*n*) megagametophyte tissue was obtained by removing the seed coats of each seed, then separating the megagametophyte from the embryo. DNA was extracted from 7–8 megagametophytes of each of the 18 parents using a modified CTAB protocol (http://www.fsl.orst.edu/pnwtirc/research/CTAB_protocol_Df_seed.pdf). Diploid (2*n*) winter buds were collected from the same 18 parents to confirm the inheritance of SSR alleles identified in megagametophytes. We also collected winter buds from an additional 134 ramets (total=152 ramets from 59 genotypes) in the orchard and from 38 trees in native stands located within one kilometer of the orchard. DNA was isolated from winter buds using a protocol developed at Oregon State University (http://www.fsl.orst.edu/tgerc/dnaext.htm).

The mapping population was a three-generation outbred pedigree described by Jermstad et al. (1994). Ninety-two of the F_2 progeny were genotyped for the SSR markers reported in this paper. Needle tissue from parents, grandparents and progeny was harvested, ground in liquid nitrogen to a coarse powder, stored at −80°C, and then used for DNA isolation as described by Jermstad et al. (1998).

Genomic libraries and isolation of SSRs

Four SSR-enriched genomic libraries were constructed by Genetic Identification Services, Chatsworth, Calif., (GIS; http://www.genetic-id-services.com) using a magnetic bead-capture approach (Peacock et al. 2002). Biotin-$(CA)_{15}$, biotin-$(GA)_{15}$, biotin-$(AAT)_{12}$, and biotin-$(ATG)_{12}$ were used as capture molecules for the four libraries, respectively. A fifth library was constructed at the University of Bristol, UK using membrane hybridization enrichment for a mixture of SSR motifs (Edwards et al. 1996).

Vector inserts were amplified using PCR and a subsample of each PCR product was used to determine the length of the insert via electrophoresis in 2% agarose gels. A second subsample was denatured and a single aliquot was spotted and bound to a positively charged nylon membrane (Roche Boehringer Mannheim Diagnostics, Basel, Switzerland) to measure the relative copy number of each insert sequence in the Douglas-fir genome. Total genomic DNA from the seedling used to construct the genomic libraries was labeled with digoxigenin, then hybridized to the dot-blot membranes using the hybridization conditions described by Pfeiffer et al. (1997). Hybridization signals were quantified using the LabWorks Analysis Software (Ultra-Violet Products, Upland, Calif.).

Plasmids containing low-copy inserts longer than 400 bp were purified using the QIAprep Spin Miniprep Kit (QIAGEN), then sequenced on an ABI Prism 3700 DNA analyzer [Applied Biosystems (ABI), Foster City, Calif.] using the BigDye Terminator v. 3.0 Ready Reaction Sequencing Kit (ABI). Redundant sequences were identified by pairwise BLAST analyses and eliminated from further consideration.

Primer design and detection of putative SSR loci

Primers targeting the SSR flanking sequences were designed using the PRIMER program (version 0.5, Whitehead Institute for Biomedical Research, Cambridge, Mass.). Initial screening of the candidate markers was done by amplifying 10 ng of template DNA in 15 μl of PCR mix including 2.5 mM of $MgCl_2$, 0.67 mg ml^{-1} BSA, 0.17 mM of each dNTP, 0.5 μM of the respective forward and reverse primers, 1× PCR buffer, and 1 unit of *Taq* DNA polymerase (Invitrogen). We added *Taq* polymerase following a hot start at 94°C for 10 min. The program proceeded with seven cycles of touchdown PCR: 95°C for 30 s, empirically determined optimal annealing temperature (T_a)+7°C for 30 s, then 72°C for 45 s. The T_a was decreased by 1°C for each of the six subsequent touchdown cycles. Following touchdown PCR, the program continued with 32 cycles of 95°C for 30 s, T_a°C for 30 s, 72°C for 45 s, and a final extension of 72°C for 20 min. After electrophoresis in 2% agarose gels, primer pairs that produced variable patterns of bands of the expected size were tested in reactions containing 1.7 μM fluorescent deoxynucleotides (R110 [F]dNTP, ABI), and then detected on an ABI Prism 377 DNA sequencer. DNA samples from one megagametophyte and five adult trees were used for these preliminary screening steps. Forward primers of the best-performing candidate markers were end-labeled with the fluorescent dyes Fam, Hex, or Ned (ABI) and the resulting PCR products were detected on an ABI Prism 3100 Genetic Analyzer. The putative alleles were sized using the GeneScan software (ABI), scored using the Genotyper software (ABI), and then individually verified by visual inspection.

Allelic variability, inheritance, and map locations of the SSRs

After genotyping trees located inside and outside of the orchard, the number of alleles, the frequency of the most common allele, the frequency of null alleles, and the observed and expected heterozygosities for each SSR locus were determined using the CERVUS program (Marshall et al. 1998). This program was also used to calculate the cumulative average probability of exclusion from parentage (PE) provided by the SSR markers. We used a chi-square test to detect deviations from a 1:1 segregation ratio of alleles in megagametophytes from heterozygous mother trees (Adams and Joly 1980). The map locations of the SSRs were determined using the JoinMap software (v. 2.0, Stam and Van Ooijen 1995).

Results and discussion

Molecular characterization of SSRs

We screened 1,452 insert-containing colonies from the five SSR-enriched genomic libraries that we obtained (Table 1). Inserts were PCR-amplified and their approximate sizes were measured after electrophoresis in agarose gels. We used dot-blot hybridization to determine the relative number of copies of the insert sequences in the Douglas-fir genome. In this assay, a weak hybridization signal suggests that the target sequence has a relatively low genome copy number (Pfeiffer et al. 1997; Scotti et al. 2002). Based on visual assessment of relative hybridization signal intensity, we selected and sequenced 517 low-copy inserts that were longer than 400 bp. We also quantified the hybridization signals and compared the mean hybridization signal (as a proportion of the average for each membrane) for the selected colonies with the mean hybridization signal for all colonies. As expected, the mean hybridization signal for the selected colonies was considerably lower (29%) and this difference was statistically significant (one-sided $P<0.001$).

SSRs were found in 385 (74%) of the 517 sequenced inserts. We selected 110 SSRs whose flanking sequences were not redundant and were long enough to design pairs of compatible primers free of sequences capable of

Table 1 Development of markers from five SSR-enriched genomic libraries

Library enriched for SSR motif	Colonies processed	Colonies sequenced	Colonies with an SSR (%)[a]	Primer pairs designed	Markers developed	Repeat units per marker (ave.)	Efficiency[b] (%)
$(CA)_n$ [c]	864	322	292 (91)	81	18	44.4	5.6
$(GA)_n$ [c]	182	62	58 (94)	17	2	34.5	3.2
$(AAT)_n$[c]	96	35	9 (26)	0	0	–	0.0
$(ATG)_n$[c]	96	50	14 (28)	4	0	–	0.0
$(GA)_n$+$(GT)_n$[d]	214	48	12 (25)	8	1	21	2.0
Total (mean)	**1452**	**517**	**385 (74)**	**110**	**21** [e]	**(33.3)**	**(4.1)**

[a] Percent of sequenced colonies containing an SSR
[b] Efficiency=100×(number of markers developed)/(number of colonies sequenced)
[c] Library developed by Genetic Identification Services
[d] Library produced at the University of Bristol
[e] One additional marker, PmOSU_783, was developed using a cDNA sequence downloaded from the GenBank database (accession number AA701783). Thus, the total number of markers developed was 22

forming internal secondary structures. These 110 primer pairs were tested by amplifying template DNA from one megagametophyte (1*n*) and five adult trees (2*n*). Nine primer pairs failed to amplify products of the expected size, 17 produced a monomorphic banding pattern, 62 produced a complex banding pattern indicating that two or more loci were amplified, and 22 produced a pattern indicative of a single polymorphic locus (e.g., Fig. 1A). Fifteen of these 22 primer pairs produced a robust, single-locus pattern, whereas seven (the last seven in Table 2) occasionally amplified a second locus or produced phenotypes with excessive band 'stuttering'. The latter seven primer pairs may require different PCR conditions for different samples to achieve uniform quality of data (i.e., obtaining clear, single-locus phenotypes for some samples may require varying the annealing stringency). The highest efficiency (5.6%) for developing single-locus markers was achieved from the $(CA)_n$-enriched library (Table 1). Efficiency could be increased by checking the length of the flanking sequences prior to sequencing each insert as described by Rafalski et al. (1996). For example, we were unable to design primers for 132 of the 385 SSR-containing sequences because the flanking sequences at the 5′-end were too short. Therefore, we recommend selecting only inserts with at least 50 bp of flanking sequence on both sides of the SSR. Oligonucleotides identical to those used for library enrichment could be used in combination with vector primers to amplify and determine the length of the SSR flanks prior to purifying the plasmids. This step would also allow one to confirm the presence of an SSR within each insert (Rafalski et al. 1996).

Genotyping error is a major concern when molecular markers are used for parentage analysis. If not accounted for, genotyping error may lead to considerable bias in the estimated parameters (SanCristobal and Chevalet 1997; Marshall et al. 1998). This problem is greater when using markers with inherently high variability, such as SSRs (Pemberton et al. 1995; Robinson and Harris 1999). Minimizing the rate of mistyping and avoiding markers with high frequencies of null alleles may be crucial for obtaining unbiased estimates of gene flow and pollen contamination. Although markers which simultaneously detect two or more loci can be useful for some applications (Fisher et al. 1998; Amarasinghe and Carlson 2002), their use in parentage analysis is likely to lead to increased rates of mistyping and false inferences. Therefore, our main goal was to develop SSR markers with strong and consistent single-locus banding patterns, and low frequencies of null alleles.

Prior to developing the 22 markers characterized in this paper, we tested the 49 SSR-amplifying primer pairs for Douglas-fir reported by Amarasinghe and Carlson (2002). Although we experimented with a variety of PCR conditions (e.g., a wide range of T_a and $MgCl_2$ concentrations), and even redesigned the primers for certain SSR sequences, we had no success obtaining markers with the strong and consistent single-locus phenotypes that we desired. The best performing primer pairs (BCPsmAC5, BCPsmAC8, BCPsmAG38 and BCPsmAG39; Table 3 in Amarasinghe and Carlson 2002), were tested with T_a between 48 and 59°C. Only two of these primer pairs (BCPsmAG38 and BCPsmAG39) produced single-locus banding patterns, but the strengths and consistencies of their banding phenotypes were unsatisfactory. In contrast, the 15 single-locus SSR markers we obtained work well under a range of T_a. We attribute the better performance of our markers to the greater length of the sequences flanking the SSRs that we were able to isolate. Longer flanking sequences provide better chances to design GC-rich, compatible, and highly specific primers, which are not prone to forming internal secondary structures during PCR. The relatively short flanking sequences in the clones isolated by Amarasinghe and Carlson (2002) may have restricted their ability to design such primers far enough from the SSR sequences to enable consistent amplification of single loci.

Mendelian inheritance and polymorphism

Primer sequences, T_a, and other properties of the 22 SSR markers are shown in Table 2. We surveyed the allelic variability of the markers by genotyping an average of 32

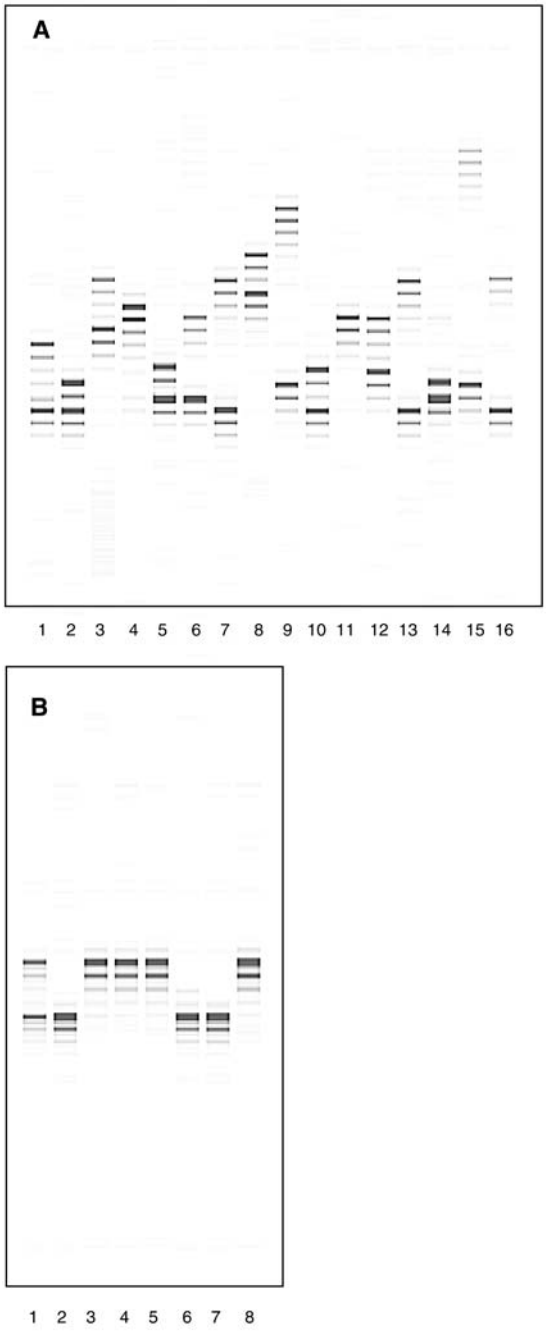

Fig. 1 Phenotypes demonstrating the variability (**A**) and the Mendelian inheritance (**B**) of SSR marker PmOSU_3G9. **A** Each of the 16 lanes contains SSR DNA amplified from diploid bud tissue from a different Douglas-fir tree. **B** The leftmost lane contains SSR DNA amplified from diploid maternal bud tissue, whereas the remaining lanes contain SSR DNA amplified from seven haploid megagametophytes extracted from seeds of the mother tree

(range=16–38) of the 38 trees sampled from natural Douglas-fir stands surrounding the seed orchard. The mean number of alleles per locus was 23 (standard error (SE)=1.6), the mean observed heterozygosity was 0.855 (SE=0.020), and the mean expected heterozygosity (H_e) was 0.923 (SE=0.013). The mean frequency of the most common allele was 0.163 (SE=0.079), and the mean frequency of null alleles was 0.036 (SE=0.039). The frequency of null alleles was estimated assuming that deviations from Hardy-Weinberg equilibrium were entirely due to the presence of null alleles (Summers and Amos 1997). The mean number of alleles per locus was 31 (SE=2.00) in a larger sample of trees (mean=78 trees/locus; range=60–95), which included trees located within the seed orchard and in the surrounding stand that were genotyped for 21 of the SSR markers (Table 3).

Our estimates of heterozygosity and mean number of alleles per locus are among the highest reported for dinucleotide SSRs in conifers (Table 3). The markers described in this paper have an H_e that is 37% higher, and a mean number of alleles per locus (based on $\bar{N}$=32) that is 188% higher than those of the 50 SSRs previously reported for Douglas-fir (Amarasinghe and Carlson 2002). We sampled 33% more individuals, and we may have sampled a more polymorphic population, but the higher levels of polymorphism probably resulted from the longer SSR sequences that we isolated. The average number of dinucleotide repeats for the 50 SSR sequences reported by Amarasinghe and Carlson (2002) (Table 3) was 20. In contrast, the average number of repeats for the $(CA)_n$ and $(GA)_n$ markers reported in this study was 39, nearly twice as large. The number of repeats is positively associated with SSR mutation rates and, therefore, with SSR marker polymorphism (reviewed in Estoup and Cornuet 1999). Only two of our 22 markers are based on perfect SSRs (i.e., single uninterrupted repeat motifs), whereas six are based on compound (two or more adjacent SSR domains with different repeat motifs), two on interrupted (two or more SSR domains interrupted by short sequences that do not fit the repeat structure), and 12 on compound interrupted SSRs (Table 2). There was a weak, but statistically significant, negative correlation between dot-blot hybridization signal and mean number of alleles per locus for the developed SSR markers (r=–0.477, two-sided P=0.033, n=21).

We confirmed the Mendelian inheritance of all 22 SSRs by analyzing the haploid segregation of alleles in seed megagametophytes from known mothers. In all cases, a bud sample from the parent tree was genotyped and run side-by-side with megagametophyte samples (Fig. 1B). For each SSR, we analyzed 5–8 megagametophytes from 6–17 heterozygous mothers. No significant deviations from a 1:1 segregation ratio were detected after pooling the data over the mother trees (data not shown). Although the small sample sizes within each mother precluded a formal test for segregation heterogeneity among mothers, there were no obvious indications of segregation heterogeneity.

Twenty of the 22 SSR loci also segregated in the progeny ($2n$) of a single controlled cross that was previously used for linkage mapping (Jermstad et al. 1998). One of the parents was heterozygous for 19 of the 20 segregating loci and the other for 17. At the 5% level, there were two statistically significant deviations from a

Table 2 Primer sequences and properties of 22 SSR markers in Douglas-fir (GenBank accession numbers at http://www.fsl.orst.edu/pnwtirc/research/DFSSRs.htm)

Locus	Forward primer 5′-3′	Reverse primer 5′-3′	Repeat motif	Optimal T_a °C (tested range)	N[a]	A[b]	Allele size (bp)	H_o/H_e[c]	f_{max}[d]	f_0[e]
PmOSU_1C3	CTCCCCTCCAGATTTTACTC	TGGCGTAACAAATAAGAGAAA	$(TC)_{25}(AC)_{32}...(TC)_4$	57 (55-57)	28	28	166–232	0.929/0.968	0.094	0.012
PmOSU_1F9	CCTCATGCATTGGACACTC	GGATTCTTGAGCAGGTAGG	$(AG)_{34}$	55 (52–57)	35	33	201–319	0.943/0.973	0.089	0.008
PmOSU_2C2	TAAATCCGCAGCTCATAGAATC	GGGTGGTGGCTAGGGAAAC	$(AC)_{32}...(CT)_4$	60 (58–62)	38	12	142–200	0.711/0.752	0.487	0.004
PmOSU_2C3	AAAGACAACATTATGAAAGG	GTAATGGTTCGAAAAATAATG	$(TC)_{24}(AC)_{18}$	50 (48–51)	35	25	163–251	0.943/0.955	0.121	0.000
PmOSU_2D4	TTATTGCACATGAGTATTATGA	CAGATGTTGTTTTTTATACCAC	$(AT)_4...(TG)_{18}(AG)_{26}$	50 (48–53)	34	30	108–194	0.912/0.968	0.109	0.022
PmOSU_2D6	GGAAAATATACATCTCACGAC	AAGCATGCGTACTAGGTG	$(AC)_5...(AC)_4...(GC)_8(AC)_{13}...(AC)_7$	54 (N/A)	34	30	162–264	0.912/0.975	0.061	0.026
PmOSU_2D9	TCGATTTACGCTTTTTTCTC	TGTTTATCCCCAGTCTCAAG	$(TC)_{32}(AC)_{15}$	57 (54–57)	16	8	125–181	0.688/0.806	0.312	0.072
PmOSU_2G12	CAAGGACTCATATGGGAAA	AACATCAGTAATAACCTTTT	$(AC)_{11}...(AC)_{19}...(GCAC)_5$ $...(GCAC)_4(AC)_7...(AC)_6$	51 (48–51)	34	16	244–310	0.824/0.914	0.177	0.047
PmOSU_3B2	CTTTGGAGTTCTTAATATAG	GATAATAGCACCCCACCATA	$(TG)_{22}(CG)_7$	49 (46–49)	32	27	88–176	0.938/0.962	0.116	0.005
PmOSU_3B9	TGTGTAAAAATGTCTAATCC	ACTACTATTCGAGGTTTTCT	$(CG)_6(CA)_6...(AC)_6...(AC)_5...AC)_6$	47 (46–49)	30	25	119–223	0.900/0.930	0.204	0.008
PmOSU_3D5	GGCATCCTATTTTTCATTTT	GTGATTACCTAACTTGTGC	$(TG)_{16}(AG)_{26}$	50 (48–51)	35	19	125–193	0.943/0.931	0.172	0.000
PmOSU_3F1	GACTAGATCATCGCAACTT	GGTATTCTTATGGTTTTTAT	$(TG)_6...(TG)_7(AG)_{27}...(AC)_4$	50 (48–52)	27	20	144–246	0.741/0.936	0.159	0.108
PmOSU_3G9	ATTCCTTTTGAGACCTACTT	CTTCAAAAATTCCTACAACA	$(TG)_{12}(AG)_{28}$	51 (48–52)	35	22	110–192	0.857/0.926	0.137	0.034
PmOSU_4A7	TTGTAAAAATTCCCATGTAT	AAGTGGGGGAGTGTGTAAT	$(TG)_5...(TG)_5...(CG)_7(TG)_4$ $...(TG)_{29}...(ATC)_5$	48 (48–54)	34	30	196–340	0.912/0.960	0.142	0.018
PmOSU_4G2	ATTTTTTGTATTGTGCTTG	TGGATATATTTGCATTTTAC	$(AT)_6...(AG)_{29}$	48 (48–51)	30	16	138–168	0.900/0.920	0.143	0.003
PmOSU_2B6[f]	TTGTTGGGTATAATTTTCA	TAATAAAATAGCTCTAACCC	$(TG)_{19}(AG)_{31}...(AT)_4$	49 (48–49)	32	28	134–346	0.813/0.957	0.133	0.075
PmOSU_2G4[f]	ATGCATTCTTGAAAGTAAA	ATAATATGCAAGTGAATCCC	$(TC)_{24}(AC)_{29}...(AC)_{12}$	51(50–51)	27	19	180–272	0.778/0.937	0.097	0.087
PmOSU_3E3[f]	TGCTTCAATTTCATATCTA	TAACATTTCAATCTATTCAC	$(TG)_5...(TG)_4...(TG)_{15}(AG)_{28}$	48 (46–49)	29	31	126–266	0.897/0.969	0.090	0.031
PmOSU_3H4[f]	TTTGCCGTCACATTTTTATTG	GCATCTTTCAGGCATAGTCT	$(GC)_9(AC)_{22}$	55 (48–57)	32	25	170-256	0.875/0.957	0.116	0.035
PmOSU_4E9[f]	GTTGGTTGTGTATATTCAGTTT	GCCTCTTCTTGGTTTGGT	$(AC)_{36}$	54 (48–55)	34	24	120–218	0.853/0.923	0.233	0.024
PmOSU_5A8[f]	CATTTTTGGATTCTGGTTTTG	ATGCCTCAAGCTATGTAATC	$(TG)_{11}...(TG)_{10}$	54 (50–55)	37	7	166–190	0.595/0.805	0.241	0.142
PmOSU_783[f]	GAGCTGATGCCTTGAAGACT	CAAGTCAGTTCACAATTCCT	$(AT)_5...(AT)_5$	57 (56–59)	33	15	205–303	0.939/0.879	0.261	0.000
Mean					**32**	**23**		**0.855/0.923**	**0.163**	**0.036**

[a] N is the number of trees genotyped. Because seed orchards may differ in gene diversity from natural populations, we used data for 38 trees sampled in a natural Douglas-fir population adjacent to the seed orchard. Exception is locus PmOSU_2D9, for which we used data from the parents genotyped for segregation analysis

[b] A is the number of alleles detected in a sample of N trees

[c] H_o and H_e are observed and expected heterozygosities, respectively

[d] f_{max} is the frequency of the most common allele

[e] f_0 is the estimated frequency of null alleles, based on deviations from Hardy-Weinberg equilibrium

[f] Primers for these loci may amplify two loci and need additional optimization

Table 3 Polymorphism and diversity of dinucleotide SSRs in some conifers

Species	No. of SSRs	N[a]	A[b]	H_o (H_e)[c]	Reference
Pseudotsuga menziesii	21[d]	78	31	0.864	This study, pooled data
	21[d]	46	26	0.864	This study, trees inside the orchard
	21[d]	32	23	0.863	This study, trees outside the orchard
	50	24	8	(0.673)	Amarasinghe and Carlson 2002
Pinus sylvestris	7	13[e]	6.7	(0.850)	Soranzo et al. 1998
Picea abies	7	18	13	(0.789)	Pfeiffer et al. 1997
Pinus halepensis/P. brutia	7	50/47	2.9	0.586	Keys et al. 2000
Picea glauca	15	14	10.2	0.520	Hodgets et al. 2001
Pinus strobus	16	16	5.4	0.515	Echt et al. 1996

[a] N is the mean number of diploid individuals genotyped
[b] A is the mean number of alleles per locus
[c] H_o is the observed heterozygosity and H_e is the expected heterozygosity for studies in which observed heterozygosity was not reported
[d] Data for PmOSU_2D9 were not used because only some trees (N=16) located within the orchard were sampled
[e] Megagametophytes were sampled in this study

1:1 segregation ratio (data not shown). In both cases, the segregation distortion was limited to only one of the two parents. We detected five null alleles (6.25%) among the 80 alleles (20 loci×2 parents×2 alleles) that we sampled. This finding, although based on a small sample of alleles, confirmed our expectations for the occurrence of a small percentage of null alleles based on population deviations from Hardy-Weinberg equilibrium (reported above and in Table 2). All 20 loci were successfully mapped to ten existing Douglas-fir linkage groups at LOD=5 (Table 4). The SSRs that we mapped are well dispersed throughout the Douglas-fir genome. All markers mapped at least 10 cM apart, except for loci PmOSU_2G4 and PmOSU_5A8, which mapped 4 cM apart. These mapping distances are unlikely to lead to non-independent association of alleles in multilocus gametes in large outcrossed populations (Epperson and Allard 1987). Therefore, this set of markers is appropriate for fingerprinting and parentage analysis in Douglas-fir.

Applications in tree improvement

Molecular markers have a variety of applications in tree improvement (Adams 1983; Wheeler and Jech 1992). Because of their high polymorphism, our 15 single-locus SSR markers will be valuable tools in the testing, breeding, and seed orchard phases of tree improvement programs. Compared to using isozyme markers, less effort will be needed to verify genotypes and controlled crosses between selected parents. SSRs should allow pollen and seed contamination in seedlots to be measured cheaply and precisely. It will also be easy to measure the success of seed orchard management techniques such as bloom delay and supplemental mass pollination. Finally, the high polymorphism of these markers can be used to directly determine the relative maternal and paternal contributions in open-pollinated seedlots from seed orchards. Because of their low variability, this is difficult to achieve using isozyme markers. Until recently, maximum likelihood modeling methods were the only feasible way to obtain this information (Adams 1992).

For example, we used our SSR markers to identify parental genotypes in the sampled seed orchard block. For 51 of the genotypes within the block, we sampled 2–3 ramets (total 145). For the remaining seven genotypes, we sampled the only ramet available. We used three of our more variable SSRs (PmOSU_2C3, PmOSU_3B2, and PmOSU_2G12) to genotype all 152 ramets. Assuming Hardy-Weinberg equilibrium, linkage equilibrium, and allele frequencies equal to those estimated from our pooled sample of trees (inside and outside of the seed orchard), no three-locus genotype is expected to occur at a frequency greater than 1.8×10^{-6} (i.e., no three-locus genotype is expected to occur more than once in 555,555 individuals). As expected, all parental genotypes had distinct three-locus SSR genotypes. In all but one case, the ramets that were labeled as belonging to the same genotype matched for all three loci. The only exception was one ramet whose genotype differed at all three loci from the other two putative ramets of the same genotype. This ramet did not appear to be identical to any of the genotypes in the orchard block. Therefore, it was included as an additional 59th genotype in further analyses. This ramet was probably intended to be included in a different orchard block and was misplaced due to a labeling error made during orchard establishment.

We also evaluated the degree of resolution in parentage analysis provided by the developed set of SSRs. The cumulative average probability of exclusion is the expected proportion of unrelated potential parents that can be excluded from parentage in a finite population using a given set of markers. Along with the number of possible parents, PE is a key determinant of the proportion of offspring that would be assigned unambiguous parentage based on genotypic exclusion (Chakraborty et al. 1988). We calculated PEs for different numbers of loci based on the SSR genotypes of the 59 parental genotypes identified in the orchard block. For example, using the three SSR loci described above, the estimated PE was 0.991 for analyses in which nothing is known about the

Table 4 Linkage map locations of 20 SSR markers in Douglas-fir (LOD=5)

Linkage group	SSR[a] markers and terminal[b] markers on the same linkage group	Position (cM)[c]	Distance between adjacent SSRs (cM)
1	rapdUBC_BC_590_975	0.0	
1	**PmOSU_2D9**	68.3	15.6
1	**PmOSU_3B2**	83.9	
1	rflpPmIFG_1246_a	158.3	
4	rflpPtIFG_2413_a	0.0	
4	**PmOSU_2G4**	91.4	4.0
4	**PmOSU_5A8**	95.4	
4	estPtINR_COMT1	146.2	
5	estPmIFG_c519_c	0.0	
5	**PmOSU_1C3**	12.0	
5	estPtIFG_8415_e	157.1	
6	rflpPmIFG_1052_d	0.0	
6	**PmOSU_4E9**	70.1	30.6
6	**PmOSU_783**	100.7	
6	estPmIFG_c572	186.7	
8	rflpPmIFG_1548_a	0.0	
8	**PmOSU_3G9**	17.4	53.4
8	**PmOSU_4G2**	70.8	
8	rflpPmIFG_1591_a	108.8	
11	rflpPmIFG_1278_b	0.0	
11	**PmOSU_2B6**	9.7	59.6
11	**PmOSU_2D4**	69.3	37.5
11	**PmOSU_2G12**	106.8	
11	Ugpp-1	164.7	
13	estPmIFG_73–6-130-E12	0.0	
13	**PmOSU_3F1**	9.2	
13	Idh	29.8	
19	**PmOSU_3H4**	0.0	
19	**PmOSU_4A7**	10.3	10.3
19	**PmOSU_3D5**	32.0	21.7
19	**PmOSU_3E3**	49.9	17.9
19	**PmOSU_2C2**	72.5	22.6
19	estPaTUM_PA66	96.1	
21	**PmOSU_3B9**	0.0	
21	rapdUBC_BC_304_450	18.9	
22	rflpPmIFG_1124_a	0.0	
22	**PmOSU_2D6**	5.9	

[a] SSRs are shown in bold; the complete names of the SSR markers appear in the DENDROME database (http://dendrome.ucdavis.edu) and contain "OSUPCT_ssr" preceding the SSR names used in this paper (e.g., OSUPCT_ssrPmOSU_2B6 is the complete name for marker PmOSU_2B6)
[b] Other markers used to construct the map are described elsewhere (Jermstad et al. 1998, 2001a, 2001b; Krutovskii et al., in preparation)
[c] Kosambi distance in cM from the marker mapped at position 0.0

two parents of a given offspring (e.g., if seeds are collected without keeping track of the mother trees). For cases in which one of the parents could be determined based on other data (e.g., known mother for each seed), the estimated PE was 0.998. Figure 2 shows how PE changes when loci with similar variability are added consecutively. For comparison, 10–15 typical isozyme loci would be needed to exceed a PE of 0.900 (Adams 1992).

The reliability and exclusionary power of our 15 single-locus SSRs make them the most efficient genetic markers available for Douglas-fir. Only three SSRs were enough to measure the success of a supplemental mass pollination experiment in which pollen from a single father tree was used to fertilize three different female parents (unpublished data). Based on preliminary results from computer simulations, we expect that a small number of loci (<10) will provide enough genetic resolution for measuring pollen contamination through

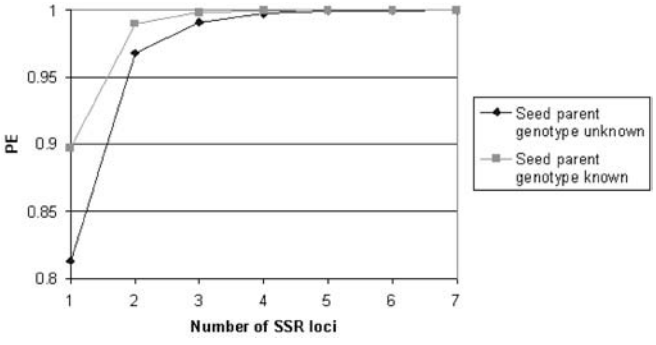

Fig. 2 Cumulative average probability of exclusion (PE) (defined in text) provided by SSR loci in Douglas-fir. PE was recalculated after adding each of the following seven SSR loci: PmOSU_2C3, PmOSU_3B2, PmOSU_2G12, PmOSU_3G9, PmOSU_2D6, PmOSU_3B9, and PmOSU_4A7. *Seed parent genotype unknown* corresponds to a situation in which seeds are collected without keeping track of the source mothers. *Seed parent genotype known* corresponds to a situation in which the genotype of the mother tree of each seed is known

paternity exclusion in any currently existing seed orchard of Douglas-fir.

Acknowledgements This study was funded by the Pacific Northwest Tree Improvement Research Cooperative at Oregon State University (OSU). We thank Stephen DiFazio, Craig Echt, and Michele Morgante for valuable advice on developing SSR markers. Materials, laboratory equipment, and technical support were generously provided by OSU's Tree Genetic Engineering Research Cooperative, Oak Ridge National Laboratory, the University of Bristol, and the University of British Columbia.

References

Adams WT (1983) Application of isozymes in tree breeding. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, part A. Elsevier, Amsterdam, pp 381–400

Adams WT (1992) Gene dispersal within forest tree populations. New For 6:217–240

Adams WT, Joly RJ (1980) Genetics of allozyme variants in loblolly pine. J Hered 71:33–40

Amarasinghe V, Carlson JE (2002) The development of microsatellite markers for genetic analysis in Douglas-fir. Can J For Res 32:1904–1915

Chakraborty R, Meagher T, Smouse P (1988) Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118:527–536

Dow BD, Ashley MV (1998) High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. J Hered 89:62–70

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996) Characterization of microsatellite markers in eastern white pine. Genome 39:1102–1108

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20:758–760

Epperson BK, Allard RW (1987) Linkage disequilibrium between allozymes in natural populations of lodgepole pine. Genetics 115:341–352

Estoup A, Cornuet J-M (1999) Microsatellite evolution: inferences from population data. In: Goldstein DB, Schlötterer C (eds) Microsatellites: evolution and applications. Oxford University Press, Oxford, pp 49–65

Fisher PJ, Richardson TE, Gardner RC (1998) Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. Theor Appl Genet 96:969–979

Hicks M, 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

Hodgetts RB, Aleksiuk MA, Brown A, Clarke C, Macdonald E, Nadeem S, Khasa D (2001) Development of microsatellite markers for white spruce (*Picea glauca*) and related species. Theor Appl Genet 102:1252–1258

Jermstad KD, Reem AM, Henifin JR, Wheeler NC, Neale DB (1994) Inheritance of restriction fragment length polymorphisms and random amplified polymorphic DNAs in coastal Douglas-fir. Theor Appl Genet 89:758–766

Jermstad KD, Bassoni DL, Wheeler NC, Neale DB (1998) A sex-averaged linkage map in coastal Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco var '*menziesii*') based on RFLP and RAPD markers. Theor Appl Genet 97:762–770

Jermstad KD, Bassoni DL, Jech KS, Wheeler NC, Neale DB (2001a) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. Theor Appl Genet 102:1142–1151

Jermstad KD, Bassoni DL, Wheeler NC, Anekonda TS, Aitken SN, Adams WT, Neale DB (2001b) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. Theor Appl Genet 102:1152–1158

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 *P. halepensis* and *Pinus brutia* Ten. Mol Ecol 9:2157–2159

Lian C, Miwa M, Hogetsu T (2001) Outcrossing and paternity analysis of *Pinus densiflora* (Japanese red pine) by microsatellite polymorphism. Heredity 87:88–98

Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Peacock M, Kirchoff VS, Merideth SJ (2002) Identification and characterization of nine polymorphic microsatellite loci in the North American pika, *Oncotona princeps.* Mol Ecol Notes 2:360–362

Pemberton J, Slate J, Bancroft DR, Barrett JA (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. Mol Ecol 4:249–252

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

Rafalski JA, Vogel JM, Morgante M, Powell W, Andre C, Tingey SV (1996) Generating and using DNA markers in plants. In: Birren B, Lai E (eds) Nonmammalian genomic analysis–a practical guide. Academic Press, San Diego, pp 75–134

Robinson JP, Harris SA (1999) Amplified fragment length polymorphism and microsatellites: a phylogenetic perspective. In: Which DNA Marker for Which Purpose? Final Compendium of the Research Project Development, Optimization and Validation of Molecular Tools for Assessment of Biodiversity in Forest Trees http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

SanCristobal M, Chevalet C (1997) Error tolerant parent identification from a finite set of individuals. Genet Res 70:53–62

Scotti I, Paglia GP, Magni F, Morgante M (2002) Efficient development of dinucleotide microsatellite markers in Norway spruce (*Picea abies* Karst.) through dot-blot selection. Theor Appl Genet 104:1035–1041

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

Stam P, Van Ooijen JW (1995) JoinMap™ version 2.0: Software for the calculation of genetic linkage maps. Centre for Plant Breeding and Reproduction Research (CPRO-DLO), Wageningen, The Netherlands

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. Mol Ecol 8:831–841

Summers K, Amos W (1997) Behavioral, ecological, and molecular genetic analyses of reproductive strategies in the Amazonian dart-poison frog, *Dendrobates ventrimaculatus*. Behav Ecol 8:260–267

Wheeler N, Jech K (1992) The use of electrophoretic markers in seed orchard research. New For 6:311–328

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1–16