Sequence-Tagged-Site (STS) Markers of Arbitrary Genes: Development, Characterization and Analysis of Linkage in Black Spruce

Daniel J. Perry and Jean Bousquet

Centre de Recherche en Biologie Forestière, Université Laval, Ste-Foy, Québec, Canada G1K 7P4

Manuscript received November 11, 1997
Accepted for publication February 23, 1998

ABSTRACT

Sequence-tagged-site (STS) markers of arbitrary genes were investigated in black spruce [*Picea mariana* (Mill.) B.S.P.]. Thirty-nine pairs of PCR primers were used to screen diverse panels of haploid and diploid DNAs for variation that could be detected by standard agarose gel electrophoresis without further manipulation of amplification products. Codominant length polymorphisms were revealed at 15 loci. Three of these loci also had null amplification alleles as did 3 other loci that had no apparent product-length variation. Dominant length polymorphisms were observed at 2 other loci. Alleles of codominant markers differed in size by as little as 1 bp to as much as an estimated 175 bp with nearly all insertions/deletions found in noncoding regions. Polymorphisms at 3 loci involved large (33 bp to at least 114 bp) direct repeats and similar repeats were found in 7 of 51 cDNAs sequenced. Allelic segregation was in accordance with Mendelian inheritance and linkage was detected for 5 of 63 pairwise combinations of loci tested. Codominant STS markers of 12 loci revealed an average heterozygosity of 0.26 and an average of 2.8 alleles in a range-wide sample of 22 trees.

OUR ability to evaluate genetic parameters in individuals or populations is directly related to our ability to detect polymorphisms at multiple genetic loci. Currently, several molecular marker technologies are available to reveal variation in nuclear genomes. Markers based on the polymerase chain reaction (PCR) are attractive because they may be essentially unlimited in numbers and require mere nanogram quantities of DNA, permitting analysis of single megagametophytes and embryos in conifers (Bousquet *et al.* 1990).

Of PCR-based markers, random amplified polymorphic DNAs (RAPDs; Welsh and McClelland 1990; Williams *et al.* 1990) and simple sequence repeats (SSRs, also known as microsatellites; Tautz 1989; Weber and May 1989) have received much attention. RAPDs are simple to develop, but equally migrating amplification products from different individuals (or species) may not represent the same locus, making it difficult to compare or combine linkage maps. In population studies, the dominant nature of RAPDs can be problematic; estimates of population genetic parameters may be unreliable if RAPDs are surveyed in diploid material (Isabel *et al.* 1995; Szmidt *et al.* 1996).

SSR markers represent single specific loci and are often highly variable with multiple codominant alleles, but their development is rather complex, often requiring enrichment cloning steps. Nonetheless, primer sequences are now available for some nuclear SSR markers in a few conifer species (Echt *et al.* 1996; Pfeiffer *et al.* 1997; Smith and Devey 1994; van de Ven and McNicol 1996). The assessment of allelic variation of SSR markers often requires high resolution, labor intensive techniques such as polyacrylimide gel electrophoresis followed by silver staining. Also, a high mutation rate, including backward mutations, and a limited range of SSR allele sizes may have a homogenizing effect, limiting the potential for divergence of SSR loci among populations (Nauta and Weissing 1996).

We are considering different approaches for obtaining PCR-based markers in black spruce [*Picea mariana* (Mill.) B.S.P.]. In this paper, we investigate sequence-tagged-site (STS) markers having polymorphisms that may be observed without manipulation of amplified products. Such markers combine the technical simplicity of RAPDs with the specificity of SSRs and, as we demonstrate, they may often be codominant. In addition to designing STS primers for black spruce genes, we characterize observed polymorphisms at the DNA sequence level and examine allelic segregation in megagametophyte arrays of individual trees, confirming Mendelian inheritance and in a few instances demonstrating linkage between locus pairs.

MATERIALS AND METHODS

cDNA sequencing: The black spruce cDNA library (provided by B. Rutledge, Natural Resources Canada) derived from an embryonic cell culture of a single diploid genotype. Reverse transcription had been initiated with a *Not*I primer-adapter ($5'$AATTCGCGGCCGC$(T)_{15}$), facilitating the inclusion of the $3'$-untranslated region (UTR) and directional cloning into

Corresponding author: Daniel J. Perry, Centre de Recherche en Biologie Forestière, Pavillon Charles-Eugène-Marchand, Université Laval, Ste-Foy, PQ, G1K 7P4 Canada. E-mail: dperry@rsvs.ulaval.ca

λgt22A. We plated the library with *Escherichia coli* Y1090 (Promega, Madison, WI) following standard procedures. Arbitrarily selected plaques were each transferred to 1 ml of SM buffer (Sambrook *et al.* 1989) containing one drop of chloroform. Inserts were amplified directly using primers GT11-F (5′ATTGGTGGCGACGACTCCTGGAG) and GT11-R (5′CAG ACCAACTGGTAATGGTAGCG) in PCRs containing 0.1 μm each primer, 0.2 mm each dNTP, 1 μl of a plaque suspension (in a 50 μl reaction), 0.025 units/μl *Taq* DNA polymerase (Pharmacia Biotech Inc., Piscataway, NJ) and 1× of the supplied reaction buffer (included 1.5 mm $MgCl_2$). PCR was carried out for 35 cycles (94°, 1 min; 55°, 1 min; 72°, 2 min) followed by 10 min at 72° in a DNA Thermal Cycler (Perkin-Elmer, Norwalk, CT). Products were examined by gel electrophoresis (1.2% agarose in TAE) and ethidium bromide staining.

cDNAs were candidates for sequencing if a single product of size 600 to 1350 bp was present in the insert amplification. Sequencing templates were PCR-amplified as above, purified using a QIAquick PCR purification kit (QIAGEN, Chatsworth, CA) and, starting with primer GT11-F, sequenced using dideoxy dye terminator cycle sequencing analyzed on a Perkin-Elmer-ABI model 373 automated DNA sequencer. With larger inserts of more than 600 bp, a second sequencing run was initiated from a clone-specific internal primer that was chosen such that its position would also allow it to be used as a forward amplification primer (see below). The deduced amino acid sequences of cDNAs were compared to the nonredundant protein sequence databases using BLASTX (Altschul *et al.* 1990) accessed at the National Center for Biotechnology Information/BLAST server (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast). When BLASTX results were negative, BLASTN was used to compare the insert sequences with the nonredundant nucleotide sequence databases. We also examined the cDNA sequences for the presence of large repeats using the program REPEAT of the Genetics Computer Group (GCG) Wisconsin Package v8.1 (Devereux *et al.* 1984).

Selection and testing of amplification primers: For each sequenced cDNA, our aim was to select a reverse primer in the presumed 3′-UTR and a forward primer about 350 to 600 bp upstream within the coding region. Occasionally, possible intron locations were identified by examining genomic sequences of similar plant genes that were found in BLASTX searches and, when possible, the forward primer was located such that the predicted amplification product would include one or two introns. Typically, 21 mers with G+C contents near 50%, minimal secondary structure and no significant inter-primer complementarity were selected.

The performance of primer pairs was tested in amplifications of the original plaque suspension and of genomic DNA of the same genotype from which the cDNA library was cloned (reference DNA). Genomic amplifications were conducted using 50 ng of reference DNA in 15 μl reaction mixtures having the same composition as that used for cDNA templates above. All PCR of genomic templates was carried out for 40 cycles (94°, 1 min; 55°, 2 min; 72°, 3 min) followed by 10 min at 72°. The ramp time to annealing and extension temperatures was 4 sec/degree.

Screening of DNA panels for polymorphism: Primer pairs that performed satisfactorily in initial tests were used to screen range-wide haploid and diploid panels of black spruce DNAs. The diploid panel (provenance trees) consisted of one tree from each of 22 provenances distributed across the species range. Needle samples were collected near Quebec City in a provenance test established in 1975 (Beaulieu *et al.* 1989). DNA was extracted from 50–75 mg of needles following Bousquet *et al.* (1990) with an additional chloroform extraction. About 50 ng of this DNA was used per 15 μl PCR with reaction conditions as for reference DNA amplifications above. The haploid panel consisted of one megagametophyte from each of the 22 provenances, usually from the same trees included on the diploid panel. Seedcoats and embryos were removed and DNA isolation from individual megagametophytes followed Bousquet *et al.* (1990) modified to include a phenol:chloroform:isoamyl alcohol (25:24:1) extraction and precipitation with ethanol. About 1–5 ng DNA was used per 15 μl PCR. Amplification products were subjected to electrophoresis through thin (3 mm) gels (1.2 or 2% agarose in TAE, see results) followed by ethidium bromide staining.

Marker segregation analyses: Polymorphic markers were examined in 22 to 30 megagametophytes from each heterozygote among 11 individuals that were a subset of the provenance trees. For each locus, goodness-of-fit to a 1:1 ratio of alternate alleles was tested using a G-test (Sokal and Rohlf 1981). In the absence of heterogeneity, data of heterozygotes with like alleles were pooled.

Linkage was examined between pairwise combinations of loci using a method equivalent to the double backcross (Bailey 1961; Narain 1990). This involved the calculation of three χ^2 statistics; two for testing segregation at individual loci (χ^2_A and χ^2_B) and the third for testing linkage (χ^2_L). When data were available from more than one double-heterozygote for a pairwise combination, heterogeneity χ^2's were obtained following Narain (1990). If heterogeneity was found for the linkage component, tests for linkage were conducted using individual tree data. Otherwise, data were pooled. For each pair of loci demonstrating linkage, the recombination frequency (y) and its standard error (SE_y) were estimated following Bailey (1961). The possibility of heterogeneous recombination frequencies among trees was also investigated using a χ^2 test (Adams and Joly 1980).

Genomic sequences of alternative alleles: Sequencing templates of individual alleles were amplified from single megagametophytes and purified using QIAquick PCR or gel purification kits (QIAGEN). Sequencing was from the same forward primer used for template amplification. Allelic sequences were aligned manually.

Production of synthetic heterozygotes: Amplification products of heterozygote genotypes that were not represented in the panel of provenance trees were simulated by template mixing. Approximately 0.25 ng of each of two allelic sequencing templates were combined and amplified by PCR. Alternatively, approximately equal quantities of allelic products from separate PCRs of haploid megagametophyte DNAs were combined and subjected to five additional thermal cycles as used in amplification. The validity of these procedures was confirmed by constructing several synthetic heterozygotes corresponding to genotypes for which comparison with true heterozygotes was possible.

RESULTS

Characterization of cDNAs: Upon examination of amplification products of each of 100 plaques, 71 cDNA clones satisfied the requirements for sequencing. Of these, 51 were sequenced. With one exception (SB66), sequencing was full length.

Many (78%) of the sequenced cDNAs encoded products similar to those of genes previously characterized in other organisms (Table 1). Of the 11 sequences that did not produce positive BLASTX results, two (SB07 and SB08) shared about 70% nucleotide identity with *Arabidopsis thaliana* sequences in the GenBank expressed sequence tag (EST) division (accessions ATTS1819 and AA394640, respectively). Two pairs of cDNAs

TABLE 1

Black spruce cDNAs with similarity to known genes of other organisms

Clone	Putative identification	Source organism and GenBank accession no.		Amino acid identity
SB01	Aquaporin	*Spinacia oleracea*	L77969	83 (264)
SB06	Acyl-CoA oxidase homolog	*Phalaenopsis* sp.	U66299	73 (223)
SB09	Thioredoxin	*Nicotiana tabacum*	X58527	58 (115)
SB11	Ribosomal protein L15	*Rattus norvegicus*	X78167	68 (204)
SB12	Nucleolysin TIA-1	*Homo sapiens*	M77142	51 (49)
SB13	DNA-binding protein CROC-1B	*Homo sapiens*	U39361	49 (136)
SB14	NADH-glutamate synthase	*Medicago sativa*	L01660	74 (147)
SB15	Calcium dependent protein kinase	*Vigna radiata*	U08140	82 (264)
SB16	Ribosomal protein L13a	*Cyanophora paradoxa*	Y09971	68 (177)
SB18	Glutathione *S*-transferase	*Nicotiana tabacum*	X56266	53 (139)
SB21	Fibrillarin	*Schizosaccharomyces pombe*	X69930	75 (236)
SB23	Ribosomal protein S15	*Arabidopsis thaliana*	Z23161	89 (139)
SB25	Protein kinase	*Spinacia oleracea*	Z30332	52 (98)
SB29	ATAF1	*Arabidopsis thaliana*	X74755	65 (46)
SB30	Rac-like protein	*Arabidopsis thaliana*	U43501	85 (198)
SB31	Actin	*Striga asiatica*	S68003	100 (53)
SB32	Mitotic cyclin (S13-7)	*Glycine max*	X62303	59 (224)
SB34	Regulatory protein preg	*Neurospora crassa*	L07314	42 (117)
SB35	GASA5	*Arabidopsis thaliana*	U53221	76 (72)
SB38	Ribosomal protein L17	*Nicotiana tabacum*	L18915	96 (133)
SB40	Heat shock protein 82	*Oryza sativa*	Z11920	89 (196)
SB41	Mitochondrial import site protein ISP42	*Saccharomyces cerevisiae*	X56885	36 (60)
SB42	Ribosomal protein L31	*Nicotiana glutinosa*	U23784	78 (120)
SB46	KIAA0107	*Homo sapiens*	D14663	58 (232)
SB48	BTF3-like transcription factor	*Nicotiana plumbaginifolia*	Y09106	56 (111)
SB49	YGL010w	*Saccharomyces cerevisiae*	Z72532	54 (53)
SB50	C01F1.3	*Caenorhabditis elegans*	U58761	26 (129)
SB51	Ribosomal protein L3	*Oryza sativa*	D12630	88 (86)
SB52	Glutathione *S*-transferase	*Nicotiana tabacum*	X56266	50 (148)
SB53	Ubiquitin-activating enzyme 2	*Arabidopsis thaliana*	U40566	44 (49)
SB55	Ubiquitin-conjugating enzyme	*Saccharomyces cerevisiae*	U18839	51 (152)
SB56	Phosphoglycerate kinase (cytosolic)	*Nicotiana tabacum*	Z48976	89 (146)
SB58	Ribosomal protein L5	*Solanum melongena*	AB001583	62 (88)
SB62	Ribosomal protein L15	*Rattus norvegicus*	X78167	68 (204)
SB65	Proteasome subunit HsC10-II	*Homo sapiens*	D26598	62 (112)
SB66	Defender against cell death 1 homolog	*Malus domestica*	U68560	82 (113)
SB67	Late embryogenesis abundant protein	*Picea glauca*	L47117	45 (77)
SB68	Pyruvate dehydrogenase E1 beta subunit	*Synechocystis* sp.	D90906	75 (285)
SB71	TAT-binding protein homolog TBP10	*Dictyostelium discoideum*	L16579	83 (175)
SB72	Ribosomal protein L27A	*Arabidopsis thaliana*	X91959	78 (56)

Putative identification was determined from the highest scoring BLASTX alignment. Amino acid identity shows the percentage of identical amino acids, with the length of the alignment in parentheses.

(SB18 and SB52; Sb11 and SB62) encoded similar products. These pairs had 83.2% and 81.3% nucleotide identities, respectively, within protein coding regions, but their 3′-UTRs appeared completely divergent. All remaining cDNAs were unique.

Large (38 bp to 106 bp) direct repeats were found in noncoding regions of seven cDNAs (SB06, SB08, SB13, SB24, SB42, SB49 and SB52). In SB13 and SB52, repeat elements were notably decayed (about 85% identity). A 38 bp direct repeat in SB08 was itself interrupted by another of 24 bp.

Two other peculiarities were noted. SB40 consisted entirely of open reading frame, encoding 206 amino acids (a.a.), and no 3′-UTR. And, a BLASTX search with the SB25 sequence suggested similarity to a protein kinase, but the similarity existed in what we inferred to be the 3′-UTR of SB25. This may reflect a rearrangement, perhaps a cloning artifact.

Selection and testing of amplification primers: Amplification primers were selected for each of the sequenced cDNAs, excepting SB40 owing to its lack of 3′-UTR sequence. All 50 primer pairs produced cleanly amplified products of predicted sizes from corresponding plaque suspensions. Based upon amplification trials using reference genomic DNA, 39 pairs (78%) were judged suitable for screening of haploid and diploid panels. Of these, 18

pairs did not reveal polymorphism. With SB41 primers, a double-banded pattern was obtained for some trees, but single invariant products were amplified from megagametophytes of those trees. We did not investigate this putative locus further. Markers generated using the remaining 20 primer pairs could be classified into four general groups based on the types of polymorphisms revealed: (1) those with null amplification alleles, but no length variants evident (three loci); (2) loci with null alleles and length variants (three loci); (3) loci at which only codominant, length variants were observed (12 loci); and (4) markers revealing dominant, length polymorphisms (two loci).

Loci with null amplification alleles: Null amplification alleles were apparent at six loci (*Sb16*, *Sb17*, *Sb18*, *Sb52*, *Sb53* and *Sb66*). Segregation of null alleles appeared consistent with the expected 1:1 ratio but replication of results was at times problematic. In some trials, alleles first characterized as null were better described as low amplification alleles and occasionally, a range of product concentrations was present. Also, the misclassification of occasional failed reactions as nulls became evident upon repetition.

Three of these loci (*Sb17*, *Sb18* and *Sb52*) also had codominant length polymorphisms that segregated in accordance with a 1:1 ratio among megagametophytes of heterozygous trees. Differences in sizes of alleles at these loci were small, likely less than 5 bp in most cases. Since the presence of null alleles would limit their potential value as markers in population studies, they were not characterized further. Codominant alleles were initially suspected at *Sb66* too, because amplification of diploid DNAs of some trees resulted in the production of additional bands consistent with heteroduplex products. However, no length variation was detected among products from corresponding megagametophyte DNAs.

Loci at which all observed polymorphisms were codominant: Codominant markers were obtained for 12 loci at which there was no evidence of null alleles (Figure 1). All were resolved on 2% agarose gels excepting *Sb01* for which 1.2% gels were used. Size differences among alleles ranged from 1 bp to an estimated 175 bp. The presence of slower migrating heteroduplex DNA made it possible to detect heterozygotes on the diploid panel using short gels (10 cm), even when differences in sizes among alleles were small. But, when alleles differed in length by less than 10 bp, and for *Sb01*, long gels (22 cm) were necessary to assess allelic segregation among megagametophytes and to assign diploid genotypes with confidence. Genotypes of *Sb21* were the least resolved of these 12 loci. Three alleles could be distinguished in homozygotes or haploid megagametophytes, but *Sb21-473/474* heterozygotes appeared no different than *Sb21-474/474* homozygotes and *Sb21-471/473* heterozygotes were essentially indistinguishable from *Sb21-471/474* heterozygotes (Figure 1g). Pooling of alleles *Sb21-473* and *Sb21-474* will be required in population studies if electrophoretic conditions are similar to ours.

The DNA sequences of all observed allelic products were determined for each of these loci except *Sb01*. Of the codominant markers, the amplification products of *Sb01* were the largest and, with five alleles and an observed heterozygosity of 0.77 among the range-wide panel of 22 trees, they were also the most variable. We inferred that the *Sb01* polymorphisms were likely within an intron corresponding to intron 2 of the three introns in similar plant genes (Guerrero and Crossland 1993; Kaldenhoff *et al.* 1993). Amplification using a reverse primer within the coding region (SB01-Rb) and the original forward primer (SB01-F) was adopted because it excluded the apparently invariant intron 3 and downstream sequence, thereby shortening product lengths by about 400 bp and improving resolution of alleles.

DNA sequencing revealed that most of the remaining polymorphisms were due to one or more small ($\leq$15 bp) insertions/deletions in introns (*Sb07*, *Sb11*, *Sb31*, *Sb62*), the 3′-UTR (*Sb70*, *Sb72*) or both (*Sb08*, *Sb21*). Two markers are noteworthy because the observed polymorphisms involved large tandem direct repeats of 3′-UTR sequences. At *Sb06*, the common allele (*Sb06-539*) had two copies of a 70 bp element where *Sb06-609* had three. At *Sb24*, the two observed alleles differed by the presence or absence of a 33 bp repeat. Polymorphisms at *Sb29* were unique in that they were located within the protein coding sequence. Relative to the common allele (*Sb29-574*), the *Sb29-553* product would have a deletion of 7 a.a. and the *Sb29-580* product would have an insertion of 2 a.a.

Without exception, segregation of alleles among megagametophytes from heterozygous trees was consistent with the expected 1:1 ratio indicative of Mendelian inheritance. In the range-wide sample of 22 trees, these 12 loci showed an average observed heterozygosity (H_o) of 0.26 and an average of 2.8 alleles (Table 2).

Markers with dominant length polymorphisms: Amplification of *Sb35* from diploid DNA of provenance trees produced either a single band of 440 bp or a 440 bp product and a 496 bp product. This two-banded phenotype behaved in a dominant manner (Figure 2a). We sequenced both products amplified from a single megagametophyte and found the 496 bp product to be identical to the 440 bp product except that it was extended by 56 bp of sequence that was composed of 35 bp of additional 3′-UTR sequence (as seen in cDNA SB35) plus the 21 bp of primer SB35-R. There was no sequence at this location in cDNA SB35 bearing any resemblance to a SB35-R priming site and it is not clear how the alleles differed such that priming also occurred at this distal site in some genotypes.

When segregation at *Sb42* was examined among megagametophytes, either a 582 bp or a 766 bp product predominated. However, in a diploid state, the larger product was dominant (Figure 2b). Amplification prim-

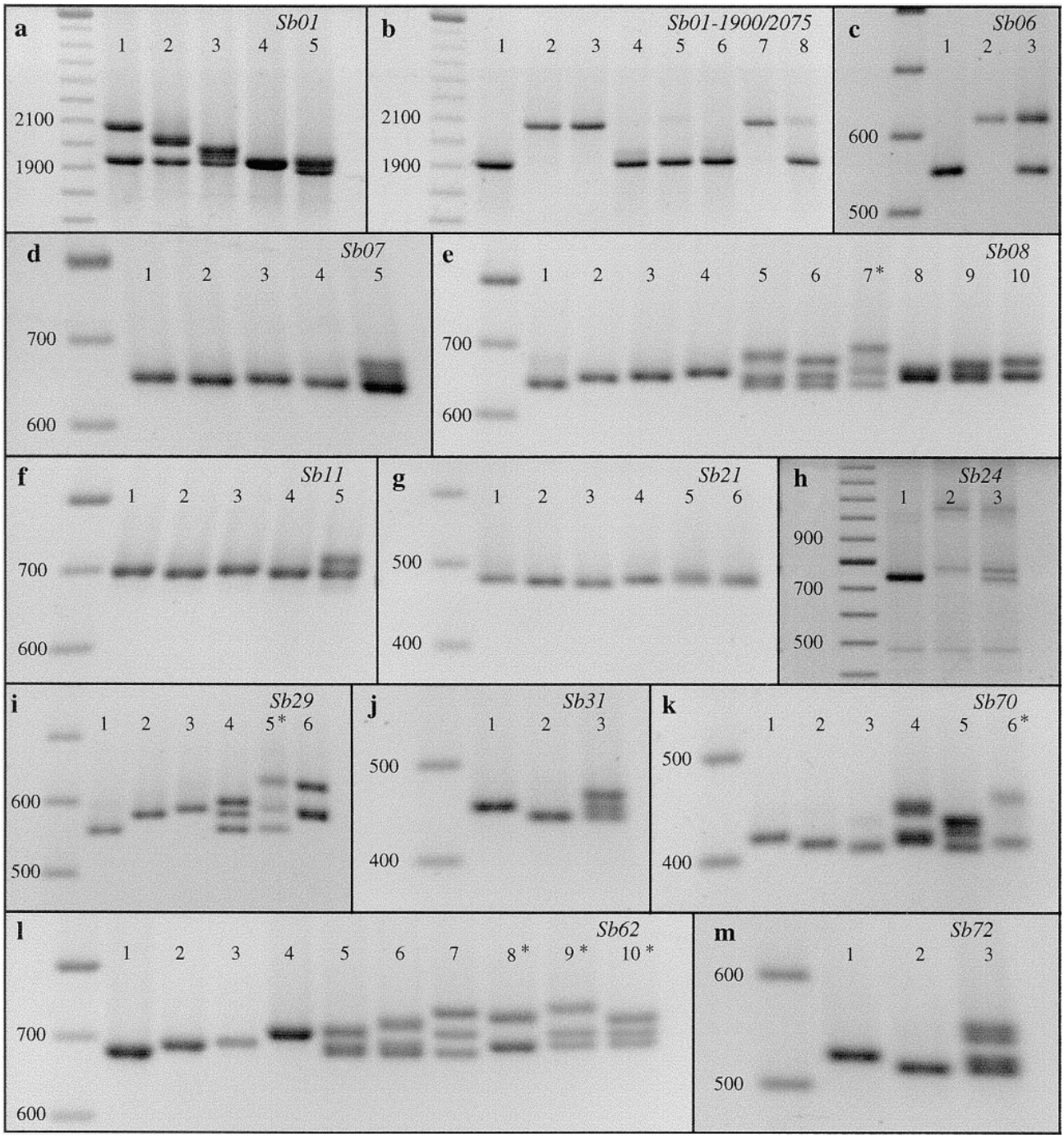

Figure 1.—Codominant sequence-tagged-site (STS) markers of black spruce genes. Polymorphisms were observed on ethidium bromide-stained agarose gels without further manipulation of amplification products. Negative images are shown. Size markers (left-hand lanes) are fragments of a 100-bp ladder (Pharmacia). Lanes with numbers marked by asterisks contain synthetic heterozygote products representing genotypes not found on the provenance tree panel (see materials and methods). (a) Lanes 1–5, genotypes *Sb01-1930/2075*, *Sb01-1930/2010*, *Sb01-1930/1960*, *Sb01-1930/1930* and *Sb01-1900/1930*. (b) Allelic segregation among eight megagametophytes of an *Sb01-1900/2075* heterozygote. (c) Lanes 1 and 2, alleles *Sb06-539* and *Sb06-609*; lane 3 an *Sb06-539/609* heterozygote. (d) Lanes 1–4, alternating alleles *Sb07-648* and *Sb07-645*; lane 5, an *Sb07-645/648* heterozygote. (e) Lanes 1–4, alleles *Sb08-634*, *Sb08-645*, *Sb08-646* and *Sb08-653*; lanes 5–10, heterozygotes *Sb08-634/645*, *Sb08-634/646*, *Sb08-634/653*, *Sb08-645/646*, *Sb08-645/653* and *Sb08-646/653*. (f) Lanes 1–4, alternating alleles *Sb11-695* and *Sb11-691*; lane 5, an *Sb11-691/695* heterozygote. (g) Lanes 1–3, alleles *Sb21-474*, *Sb21-473* and *Sb21-471*; lanes 4–6, heterozygotes *Sb21-473/474*, *Sb21-471/474* and *Sb21-471/473*. (h) Lanes 1 and 2, alleles *Sb24-738* and *Sb24-771*; lane 3 an *Sb24-738/771* heterozygote. (i) Lanes 1–3, alleles *Sb29-553*, *Sb29-574* and *Sb29-580*; lanes 4–6, heterozygotes *Sb29-553/574*, *Sb29-553/580* and *Sb29-574/580*. (j) Lanes 1 and 2, alleles *Sb31-449* and *Sb31-439*; lane 3 an *Sb31-439/449* heterozygote. (k) Lanes 1–3, alleles *Sb70-417*, *Sb70-410* and *Sb70-404*; lanes 4–6, heterozygotes *Sb70-410/417*, *Sb70-404/417* and *Sb70-404/410*. (l) Lanes 1–4, alleles *Sb62-681*, *Sb62-689*, *Sb62-691* and *Sb62-706*; lanes 5–10, heterozygotes *Sb62-681/689*, *Sb62-681/691*, *Sb62-681/706*, *Sb62-689/691*, *Sb62-689/706* and *Sb62-691/706*. (m) Lanes 1 and 2, alleles *Sb72-523* and *Sb72-515*; lane 3, an *Sb72-515/523* heterozygote.

TABLE 2

Allelic length polymorphisms observed at 12 codominant STS loci in black spruce

Locus	Putative identification	Allele frequencies[a]	Observed heterozygosity
Sb01	Aquaporin	0.34, 0.45, 0.05, 0.11, 0.05	0.77
Sb06	Acyl-CoA oxidase homolog	0.95, 0.05	0.09
Sb07	unknown	0.02, 0.98	0.05
Ab08	unknown	0.05, 0.27, 0.55, 0.14	0.59
Sb11	Ribosomal protein L15	0.11, 0.89	0.23
Sb21	Fibrillarin	0.18, 0.82[b]	0.18
Sb24	unknown	0.84, 0.16	0.32
Sb29	ATAF1	0.14, 0.84, 0.02	0.23
Sb31	Actin	0.11, 0.89	0.23
Sb62	Ribosomal protein L15	0.80, 0.05, 0.07, 0.09	0.32
Sb70	unknown	0.02, 0.02, 0.95	0.09
Sb72	Ribosomal protein L27A	0.02, 0.98	0.05

Putative identification refers to gene products identified in BLASTX searches (Table 1). Polymorphisms were directly observed on agarose gels without further manipulation of amplification products. Estimates of allele frequencies and observed heterozygosities are based on a range-wide sample of 22 black spruce trees.

[a] Alleles are listed in order of increasing sizes (see Figure 1).

[b] Alleles *Sb21-473* and *Sb21-474* were pooled (see results).

ers were positioned to include a large (106 bp) repeat found in cDNA SB42, but the structure of this repeat did not vary among the allelic products. Rather, *Sb42-766* had an additional large direct repeat of at least 114 bp that included the SB42-R primer site. Although the first element of this additional repeat had a site exactly complementary to SB42-R, amplification from the distal site was favored. The distal site was either not present or was not favored in amplifications of the common allele (*Sb42-582*). The mechanism of suppression of amplification from the proximal site in heterozygotes is unknown.

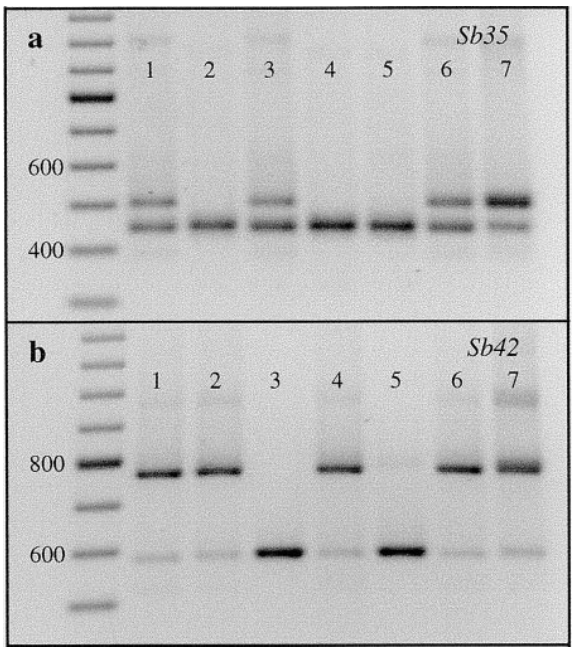

Figure 2.—Segregation of dominant length polymorphisms of sequence-tagged-site (STS) markers among megagametophytes of heterozygous trees. Polymorphisms were observed on ethidium bromide-stained agarose gels without further manipulation of amplification products. Negative images are shown. Size markers (left-hand lanes) are fragments of a 100 bp ladder (Pharmacia). (a) Lanes 1–6, segregation of alleles *Sb35-440* and *Sb35-440*&*496* among six megagametophytes of an *Sb35-440/440*&*496* heterozygote (lane 7). (b) Lanes 1–6, segregation of alleles *Sb42-582* and *Sb42-766* among six megagametophytes of an *Sb42-582/766* heterozygote (lane 7).

Analysis of linkage: All codominant markers and the dominant markers of *Sb35* and *Sb42* (a total of 17 loci) were included in the analysis of linkage. We examined 63 of 136 possible two-locus combinations; five were indicative of linkage (Table 3) with no heterogeneity of recombination frequencies detected among trees. It may be appropriate to exclude one member of linked pairs in analyses that require an assumption of independence among loci, retaining those having higher heterozygosities. However, the results for the *Sb07/Sb62* and

TABLE 3

Linkage of STS markers of black spruce genes

Loci	Sample sizes: Trees	Sample sizes: Megagametophytes	y	SE_y
Sb01/Sb06	2	46	0.17	0.06
Sb01/Sb29	3	76	0.28	0.05
Sb07/Sb62	1	23	0.26	0.09
Sb11/Sb24	1	30	0.23	0.08
Sb11/Sb62	2	60	0.33	0.06

Estimates of recombination frequencies (y) and standard errors (SE_y) are given for pairs of loci for which significant linkage was detected.

Sb11/*Sb24* combinations should be viewed as tentative because only one doubly heterozygous tree was available for each.

DISCUSSION

Allelic variation that can be detected directly on agarose gels without additional manipulation of PCR products is reasonably common among STS markers of black spruce genes. Out of 39 markers screened, 12 showed codominant length polymorphisms suitable for use in population studies. Codominant markers were also found for three additional loci, but these are less suited to population studies owing to the presence of null alleles which could cause erroneous heterozygosity estimates. However, these three markers and dominant length polymorphisms identified at two additional loci should be well suited to applications such as genome mapping.

Most (78%) of the black spruce cDNAs sequenced here were similar to previously characterized genes. This high number probably reflects the fact that we made no effort to avoid abundantly expressed messages that are more likely to be already represented in sequence databases. To obtain markers of a wider variety of types of genes, techniques such as cold-plaque screening (Hodge *et al.* 1992) could be used to identify clones of rarely expressed mRNAs. Markers could also be tailored to represent different classes of genes by using libraries derived from specific tissues, developmental stages, or environmental treatments. Also, for some species, an increasingly large variety of precharacterized sequences are becoming available in publicly accessible databases.

STS markers have been developed in other plants (Bradshaw *et al.* 1994; Ghareyazie *et al.* 1995; Talbert *et al.* 1994; Tragoonrung *et al.* 1992) including the conifer *Cryptomeria japonica* (Tsumura *et al.* 1997). In general, the proportion of directly observable length polymorphisms has been low and digestion of amplification products with restriction enzymes (PCR-RFLP) has routinely been used. Also, the source of sequence information for previous STS marker development has often been genomic clones rather than cDNAs. However, there are scattered reports of allelic length polymorphisms of plant genes (Bradshaw *et al.* 1994; Davis and Yu 1997; Perry and Furnier 1996; Tragoonrung *et al.* 1992). A low frequency of directly observable length polymorphism may be a reflection of the screening panels that have been used; small panels, or panels with a restricted genetic base, may have encompassed little of the total genetic diversity. When 15 pairs of STS primers were screened against a diverse panel of 40 rice varieties, six (40%) revealed length polymorphisms (Ghareyazie *et al.* 1995), a proportion similar to that found here for black spruce (15/39, 38%), suggesting that potential success rates may be reasonably high for a wide range of plant species. But, we also note that in cases where interspecific crosses have been used to create presumably highly heterozygous mapping populations, the amount of length polymorphism has remained low (Bradshaw *et al.* 1994; Slabaugh *et al.* 1997).

Polymorphisms, and length polymophisms in particular, are most likely to occur in noncoding regions. Therefore, when possible intron locations were identified based upon similar gene sequences in other plants, we placed the amplification primers such that one or two introns would be included in genomic products. To ensure that noncoding DNA was included even if no introns were present, reverse amplification primers were placed in the 3′-UTR. This strategy was also intended to increase specificity when primers were based upon one member of a gene family, a concern of particular importance in conifers where large gene families are common (Ahuja *et al.* 1994; Kinlaw *et al.* 1994; Perry and Furnier 1996). Our results indicate that very similar members of a gene family are generally sufficiently divergent in their 3′-UTRs that PCR can be directed toward single genes.

In addition to large gene families, another interesting feature of conifer genomes is an abundance of large tandem direct repeats. Large repeats are common in noncoding regions of jack pine (*Pinus banksiana*) alcohol dehydrogenase (*Adh*) genes (Perry and Furnier 1996) and, in that same study, similar repeats were identified in five of seven genomic sequences of conifer genes found in GenBank. In the present study, large direct repeats ranging in size from 38 to 106 bp were found in seven of 51 cDNAs. Considering the smaller noncoding component of cDNAs, it is not unexpected that this frequency is lower than that reported for genomic gene sequences.

In codominant STS markers of three *Adh* loci in jack pine (Perry and Furnier 1996), alleles differed by the presence or absence of large repeats. With this in mind, when a large repeat was present in a black spruce cDNA, PCR primers were positioned, when possible, to include the repeat in the amplified products. In one case (*Sb06*), this strategy was successful and resulted in a codominant marker with alleles differing in size by 70 bp. However, the polymorphism was not due to the presence or absence of the repeat as anticipated, rather, alleles differed by having either a duplication or a triplication of the sequence. In both other cases where primers were positioned to flank a large repeat in the cDNA (*Sb24* and *Sb42*), polymorphisms were found but they involved additional large repeats rather than the elements originally targeted. The presence of the targeted repeats was apparently fixed. As illustrated by *Sb42*, additional repeats may lead to unpredictable results, including dominant length polymorphisms when a primer site is duplicated.

Of the 11 codominant markers characterized at the DNA sequence level, only those of *Sb06* and *Sb24* in-

volved large repeated sequences. The remainder were based on relatively small insertions or deletions with net differences among alleles ranging from 1 bp to 27 bp. In nearly all cases, each possible heterozygote could be identified unambiguously, even when differences among alleles were small. *Sb21* was an exception where pooling of alleles may be necessary. In many cases, classification of heterozygous genotypes was simplified by the presence of genotype-specific heteroduplex bands. For example, the alleles *Sb62-689* and *Sb62-691* were very similar in size, but *Sb62-681/689* and *Sb62-681/691* heterozygotes were readily discriminated by their distinctive heteroduplex products (Figure 1l). Moreover, we have demonstrated that it is possible to predict the heteroduplex banding patterns of hitherto unseen genotypes by construction of synthetic heterozygotes via template mixing. In some cases, template mixing may also be a useful tool to ensure that rare homozygotes are properly identified when the possible genotypes would give products of similar size, *e.g., Sb62-689/689* and *Sb62-691/691*, and, owing to their low frequencies, examples of both are not available for direct comparison.

With codominant length polymorphisms revealed by 15 of a total of 50 pairs of primers synthesized, our overall success rate may be similar to that of finding SSR polymorphisms in conifers. An intensive effort to develop SSR markers has been directed toward eastern white pine (*Pinus strobus*; Echt *et al.* 1996). Primer pairs were selected from 77 SSR containing clones and of those, 16 pairs amplified well and revealed polymorphisms in a panel of 16 trees. A similar success rate has been reported for SSR marker development in Norway spruce (*Picea abies*), with 7 of 36 primer pairs amplifying single polymorphic loci (Pfeiffer *et al.* 1997). An overall success rate similar to that for developing SSR markers is perhaps unexpected since, unlike SSR markers, STS markers do not target specific sequences that are expected to promote polymorphism. Therefore, they do not entail the added effort and expense of isolating and identifying regions containing such sequences.

However, SSR markers will likely surpass STS markers having directly observed polymorphisms in terms of heterozygosity and numbers of alleles per locus. Average heterozygosities of 0.515 and 0.79, and averages of 5.4 and 13 alleles per locus were reported for the polymorphic SSR markers in 16 white pine and 18 Norway spruce, respectively (Echt *et al.* 1996; Pfeiffer *et al.* 1997), compared to an observed heterozygosity of 0.26 and 2.8 alleles per locus for codominant STS markers in 22 black spruce. The amount of variation revealed by these STS markers appears more in line with that of RAPD and allozyme loci in black spruce (Boyle and Morgenstern 1987; Isabel *et al.* 1995). As with SSR markers, the total information per PCR may be increased by multiplexing. Indeed, we have conducted successful trials employing several two-set combinations of STS primers (data not presented).

STS markers may be useful when incorporated into linkage maps. Placement of known genes on maps would add to our knowledge of conifer genome organization and assist in combining maps from different individuals. Although RAPD-based maps are commonly constructed for conifers, it is often difficult to use the same RAPD markers in different trees (Devey *et al.* 1995). Plomion *et al.* (1995) have suggested the use of protein polymorphisms revealed by 2-D electrophoresis to aid in establishing the correspondence of RAPD linkage groups among trees. STS markers may be a more convenient choice for this purpose since they use the same technology as RAPDs and gene identifications may be more easily determined. However, owing to relatively low levels of heterozygosity, few of the markers described here are likely to be shared among maps if mapped individuals are selected arbitrarily with respect to these loci. Our efforts were focused on a low sensitivity screening of an extensive sampling of genes in a diverse panel of individuals. More sensitive (and more laborious) detection techniques, *e.g.*, PCR-RFLP or single-strand conformation polymorphism, may be warranted for some applications such as genome mapping. Primer pairs producing products that appear monomorphic under current conditions may be a valuable resource in such endeavors.

The codominant STS markers developed here provide an additional means to explore natural genetic variation in black spruce populations. It remains to be determined to what extent these primers can be used in other spruces and conifers. Preliminary results indicate that primers producing invariant products in black spruce may reveal polymorphisms in related species. Clearly, the wider the range of taxa in which primers are useful, the more attractive future STS marker development will be.

B. Rutledge (Natural Resources Canada) kindly provided the cDNA library, J. Beaulieu and N. Isabel (Natural Resources Canada) provided seeds, F. Larochelle and M. Perron helped with tissue collections, and D. Fournier, I. Gamache, G. Pelletier and P. Perry provided much assistance in the laboratory. This work was supported by grants to J.B. from Fonds pour la Formation de Chercheurs et l'Aide à la Recherche of Québec, Natural Sciences and Engineering Research Council of Canada and Network of Centres of Excellence in Sustainable Forest Management. The sequences reported in this article have been deposited in the GenBank database (accession nos. AF051202–AF051252 and AF051733–AF051765).

LITERATURE CITED

Adams, W. T., and R. J. Joly, 1980 Linkage relationships among twelve allozyme loci in loblolly pine. J. Hered. **71:** 199–202.

Ahuja, M. R., M. E. Devey, A. T. Groover, K. D. Jermstad and D. B. Neale, 1994 Mapped DNA probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers. Theor. Appl. Genet. **88:** 279–282.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Bailey, N. T. J., 1961 *Introduction to the Mathematical Theory of Genetic Linkage*. Oxford University Press, London.

Beaulieu, J., A. Corriveau and G. Daoust, 1989 Phenotypic stability and delineation of black spruce breeding zones in Quebec. Forestry Canada Information Report LAU-X-85E.

Bousquet, J., L. Simon and M. Lalonde, 1990 DNA amplification from vegetative and sexual tissues of trees using polymerase chain reaction. Can. J. For. Res. **20:** 254–257.

Boyle, T. J. B., and E. K. Morgenstern, 1987 Some aspects of the population structure of black spruce in central New Brunswick. Silvae Genet. **36:** 53–60.

Bradshaw, H. D., Jr., M. Villar, B. D. Watson, K. G. Otto, S. Stewart *et al.*, 1994 Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. Theor. Appl. Genet. **89:** 167–178.

Davis, T. M., and H. Yu, 1997 A linkage map of the diploid strawberry, *Fragaria vesca*. J. Hered. **88:** 215–221.

Devereux, J., P. Haeberli and O. Smithies, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12:** 387–395.

Devey, M. E., A. Delfino-Mix, B. B. Kinloch and D. B. Neale, 1995 Random amplified polymorphic DNA markers tightly linked to a gene for resistance to white pine blister rust in sugar pine. Proc. Natl. Acad. Sci. USA **92**: 2066–2070.

Echt, C. S., P. May-Marquardt, M. Hseih and R. Zahorchak, 1996 Characterization of microsatellite markers in eastern white pine. Genome **39:** 1102–1108.

Ghareyazie, B., N. Huang, G. Second, J. Bennett and G. S. Khush, 1995 Classification of rice germplasm. I. Analysis using ALP and PCR-based RFLP. Theor. Appl. Genet. **91:** 218–227.

Guerrero, F. D., and L. Crossland, 1993 Tissue-specific expression of a plant turgor-responsive gene with amino acid sequence homology to transport-facilitating proteins. Plant Mol. Biol. **21:** 929–935.

Hodge, R., W. Paul, J. Draper and R. Scott, 1992 Cold-plaque screening: a simple technique for the isolation of low abundance, differentially expressed transcripts from conventional libraries. Plant J. **2:** 257–260.

Isabel, N., J. Beaulieu and J. Bousquet, 1995 Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. Proc. Natl. Acad. Sci. USA **92:** 6369–6373.

Kaldenhoff, R., A. Kolling and G. Richter, 1993 A novel blue light- and abscisic acid-inducible gene of *Arabidopsis thaliana* encoding an intrinsic membrane protein. Plant Mol. Biol. **23**: 1187–1198.

Kinlaw, C. S., S. M. Gerttula and M. C. Carter, 1994 Lipid transfer protein genes of loblolly pine are members of a complex gene family. Plant Mol. Biol. **26:** 1213–1216.

Narain, P., 1990 *Statistical Genetics*. John Wiley and Sons, New York.

Nauta, M. J., and F. J. Weissing, 1996 Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics **143:** 1021–1032.

Perry, D. J., and G. R. Furnier, 1996 *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. Proc. Natl. Acad. Sci. USA **93:** 13020–13023.

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

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

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: a Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, New York.

Slabaugh M. B., G. M. Huestis, J. Leonard, J. L. Holloway, C. Rosato, *et al.*, 1997 Sequence-based genetic markers for genes and gene families: single-strand conformational polymorphisms for the fatty acid synthesis genes of *Cuphea*. Theor. Appl. Genet. **94:** 400–408.

Smith, D. N., and M. E. Devey, 1994 Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome **37:** 977–983.

Sokal, R. R., and F. J. Rohlf, 1981 *Biometry*, Ed. 2. W. H. Freeman, San Francisco.

Szmidt, A. E., X. R. Wang and M. Z. Lu, 1996 Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* (L.) using haploid tissue analysis. Heredity **76:** 412–420.

Talbert, L. E., N. K. Blake, P. W. Chee, T. K. Blake and G. M. Magyar, 1994 Evaluation of "sequence-tagged-site" PCR products as molecular markers in wheat. Theor. Appl. Genet. **87:** 789–794.

Tautz, D., 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl. Acids Res. **17:** 6463–6471.

Tragoonrung, S., V. Kanazin, P. M. Hayes and T. K. Blake, 1992 Sequence-tagged-site-facilitated PCR for barley genome mapping. Theor. Appl. Genet. **84:** 1002–1008.

Tsumura, Y., Y. Suyama, K. Yoshimura, N. Shirato and Y. Mukai, 1997 Sequence-tagged-sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers. Theor. Appl. Genet. **94:** 764–772.

van de Ven, W. T. G., and R. J. McNicol, 1996 Microsatellites as DNA markers in Sitka spruce. Theor. Appl. Genet. **93:** 613–617.

Weber, J. L., and P. E. May, 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44:** 388–396.

Welsh, J., and M. McClelland, 1990 Fingerprinting genomes using PCR with arbitrary primers. Nucl. Acids Res. **18:** 7213–7218.

Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey, 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. **18:** 6531–6535.

Communicating editor: A. H. D. Brown

APPENDIX

STS primer sequences

Clone	Forward primer	Reverse primer	Product size (bp)	
			cDNA	Genomic
SB01	GCGTTCCAGAAATCCTACTAC	CCAAATGCACCATAAATACAG[a]	220	1900–2075
SB06	TAAGGCAATTCTTCGGCTCAC	ACTAAGACAACCATTCTCTCC	539	539–609
SB07	AACAATGGGTTGGAGATCGTC	CGCTTGACAGGTCTTGGTAAC	382	645–648
SB08	TTCGATGCTAGGTCTTGAGTC	CAGAAATTGGAAGTAAGAACG	382	634–653
SB09	CGTGTTTGCATGTCACTCTAC	CAGCATATCCACACCGACATA	449	—
SB11	GTATTACCCAGCTCAAGTTCC	AACTATCCCACCACTCCTGTC	469	691–695
SB12	TTATTGAGGATGTCCGTGTTC	AGAGGTAGACCATCTAGTCAC	497	600
SB13	AATAGGCGATGGAAGTGTCAG	CAGAAGCAAGTACAGATGAGC	502	—
SB14	TACTTCGAGTGTCTCTCATTG	GCTGTCAGAGTTTGTAACATC	446	446
SB15	ACAAACTGGAGCGTGAAGAGC	ACTATGACGGGAGCCAAGTTG	602	—
SB16	GATTCCACACAAAACCAAGCG	CAAAGTATACCCCTTGAACAC	553	1050[b]
SB17	GAGGGATGAATATGGTCTACG	AATAACGCCAAATGCCTCCAC	515	640[bc]
SB18	TCCTTATGACCGAGCCATTGC	AACACGGTGAGAACTGATAGC	597	720[bc]
SB19	TTTGCGGAAGAGGAGACTATC	CGCTGCCCTATATCTTGTATG	397	397
SB21	CAGATCAGGCACGCATTGTTG	GTCCATCAGGGCTCATGTTTG	381	471–474
SB23	GGTTTGAAGAGGCAGCCAATG	TTGGGAAAGGCGGCACTAATG	412	—
SB24	CAGTATGTGGGTTCATGTTAG	TTTGATAGCAGAGACCACTTC	474	738–771
SB25	ACTTACCTGACATGCCCATTG	CTCCGAGCTGTTTCATATGTG	351	—
SB26	TTGGGGAAGCTACAGAGATAC	GCGAACTAAGAGACAGCAGAC	366	—
SB28	CCAACAAGAAAGCCACGTCAG	ACCAACAAACGCCCTCTTCAC	453	550
SB29	AGCGGCATTGAACAGAGTAAC	AATGGAAATGAAGGCAGACTC	574	553–580
SB30	CAGTGCTGTGCCTATAACAAC	GGCTGATGTGATTCCAGAAAG	525	—
SB31	TTGGCATCTCTCAGCACATTC	TAGGTTTCTGGTCACGTCTAC	311	439–449
SB32	TGCTGTCTACACTGCTCAATG	CAGAAGCCTGAGGATGTTACC	529	760
SB34	TATCCATCGCCTGCTTCTCAC	TGTAGTCAGTCCGAATGTACC	498	1160
SB35	AGTATGGCGAGGGCAGTCTTC	TTCACTCCCGATCCACTCATC	327	440–496
SB36	TTCAGATCCATTGCCTGTGAG	TGAGGACAAGCAACCACAGAC	429	429
SB38	GATATGGTCATGGCTACAGTG	GCAAAAACTGGACCTTATTTC	433	—
SB41	GCTGAGGGGAAGGATTGATAC	GCTTCGACAGGCATATTACAG	404	520
SB42	GAAGCTTAACAAGGCCGTATG	CCCAAACATAGGCAATAATCC	582	582–766
SB46	GGCTGTCAATACAAGTCATTC	TCACGTTGTTATTGTTGTCAC	599	2320
SB48	TCGTAAACCCAAAAGTTCAAG	ACCGCATATCCTAAAGGTAAC	504	—
SB49	AGGTCCTCCAAAAGTTCTGTG	GCCTCATGTTCCCAAAGTCTC	323	323
SB50	GCGGAACCTTACAGGAATTTG	GCATTTAGACCCCGAGGACAG	451	451
SB51	TGAAACAGACTTCTCGTACTG	TTCTTACGTAGCTGCTCTAAC	358	358
SB52	—[d]	AAATCATCGCACATAGCTACAG	756	900[bc]
SB53	CTGATCCTCCAGAAGAAACTC	AGTCCGATGGTTGCTTATGTG	569	800[b]
SB55	TCCAATGTGGTCAGTCTCAAG	GCAGATTGAAAGATTCCAGTC	544	—
SB56	CTTTGGACACAACTAAGACTG	TAGTGTCACTCCATCTGAAAC	372	475
SB58	CCGACAATCAAATACACTGAG	TACCAGACCAGACCTTCAATG	392	520
SB60	TGGGAGAATGACTAGATTGTG	AAGCCTTGACAATAGTAAGTG	378	378
SB62	—[e]	ACAGTACGCCGCAGACAAATG	424	681–706
SB64	AGGAGGATATAGCTCGGATAC	TGTTACACCGAAACTGTTCAG	510	510
SB65	CGAGTCTTGTTTCTGCGTTAC	CTTCGTCCATTGAAAGACTTG	600	—
SB66	AGGTTGTGTACATGGCAATAG	AGGCAAGGAGAATTAACAAAG	382	740[b]
SB67	CAATTCAAGTTCTTCAATCTC	GTCAATGGATTTAATGAGTTC	417	417
SB68	ACGTCCAGGTGCAGATGTAAC	TCGAACAATGTTGAGATCAAG	502	502
SB70	AAATGGCGGTGTCATCTCTTC	AAAATGAGTTCCCTGCCAATC	417	404–417
SB71	AGTATAGGATCTGCTCGAATG	CCAATATGAAACACACGGTAG	536	2500
SB72	GCTCAGGAATCACTATCATTG	CAAAGATACCAACCGATTAAG	523	515–523

Primers sequences are given 5′ to 3′. Clone refers to the cDNA from which the primer sequences were selected. Sizes of cDNA products were inferred from sequences of cDNA clones. For genomic products, sizes were estimated from relative electrophoretic mobilities, or inferred from sequences of different alleles. Multiple alleles are represented by a range of genomic product sizes. Dashes in the genomic size column indicate that genomic amplifications were unsatisfactory.

[a] Sequence shown is of SB01-Rb, a primer in the coding region within the presumed exon 3. The original reverse primer located in the 3′-UTR (SB01-R) was CAACAGAATCAGCAGCATAAG.

[b] Null amplification allele(s) also detected.

[c] Amplification products that differed slightly in size were observed but not characterized.

[d] Amplification of *Sb52* was performed using the forward primer of *Sb18* (SB18-F).

[e] Amplification of *Sb62* was performed using the forward primer of *Sb11* (SB11-F).