Theor Appl Genet (1998) 97 : 735–743

D. J. Perry · J. Bousquet

Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers

Received: 10 April 1998 / Accepted: 22 April 1998

Abstract Sequence-tagged-site primers, previously developed based upon black spruce (*Picea mariana*) cDNA sequences, were tested for their ability to direct specific amplification in two individuals of each of 12 additional conifer species. Nearly all (95–97%) of the primers functioned well in congeneric trials, while a lower proportion (21–33%) scored positively in other Pinaceae genera. Outside of the Pinaceae, amplification of homologous products was not achieved. Products from the various species often differed in size from their homologs in black spruce. In one case a large difference in size was due to the lack of an intron in a jack pine product while in several other cases the differences were due to the presence or absence of large direct repeats in the DNA sequences. Length polymorphism was occasionally evident between the two individuals examined of a given species. We investigated marker polymorphism in detail in a panel of 15 white spruce (*Picea glauca*) trees. Allelic segregation among haploid megagametophytes was revealed directly at 16 loci by standard agarose-gel electrophoresis without any additional manipulation of amplification products. Polymorphisms observed at 12 of these loci were exclusively co-dominant. For this subset of 12 loci, the average number of alleles was 3.2 and the average observed heterozygosity was 0.37.

Key words Co-dominant molecular markers · Expressed sequence tags · Conifer genes · Heterologous PCR primers · Cross-species amplification

Communicated by P. M. A. Tigerstedt

D. J. Perry (✉) · J. Bousquet
Centre de recherche en biologie forestière,
Pavillon Charles–Eugène–Marchand, Université Laval,
Sainte-Foy, Québec G1K 7P4, Canada
Tel.: + 418 656-2131, ext 4598
Fax: + 418 656-7493
E-mail: dperry@rsvs.ulaval.ca

Introduction

In recent years, polymerase chain reaction (PCR)-based molecular markers have gained wide popularity in population genetics and genome mapping. Of the various classes of PCR-based markers currently available, those which reveal co-dominant polymorphisms in specifically targeted sequences are the most informative because they can be scored unambiguously in diploid tissues and concern about the homology of markers among individuals or populations is greatly reduced. In an effort to obtain markers having these characteristics for use in black spruce [*Picea mariana* (Mill.) B.S.P.], we have developed a set of cDNA-based STS primers (Perry and Bousquet 1998). Many of the markers generated by using these primers have polymorphisms that may be revealed directly by standard agarose-gel electrophoresis and ethidium bromide staining, without the need for further manipulation of PCR products. At many loci, the polymorphisms observed have been exclusively co-dominant.

As with any PCR-based markers that target specific sequences, considerable cost is incurred during the development of cDNA-based STS markers. Sequence data must first be obtained, followed by primer design, synthesis and testing. If the primers generated for one species could be used in others, the time and resources required to launch studies involving "new" species would be significantly reduced. In general, the wider the taxonomic range across which primers are transferable, the more economically attractive primer development becomes.

The transferability of cDNA-based STS primers among plant species has not been well studied. Perhaps the most extensive trial to-date was carried out by Tsumura et al. (1997) using primers developed for the conifer *Cryptomeria japonica* screened against a panel of 14 different coniferous species. Under stringent (specific) amplification conditions, about 47% of the *C. japonica* primers directed amplification of single

fragments from the closely related *Taxodium distichum* and, although there was considerable variation among species, this number averaged about 18% within the Taxodiaceae. Outside of the Taxodiaceae, amplification success was generally much reduced.

To our knowledge, no comparable studies have been carried out using a primer source from within the Pinaceae, a family that includes many of the ecologically and economically important tree species worldwide. Here we present the results of an investigation of the degree to which cDNA-based STS primers that were originally developed for black spruce are useful in other conifers. We examined cross-species amplification in four other spruces, five additional Pinaceae genera, and two phylogenetically more distant conifers. Successful amplification is only one measure of utility; we also screened a panel of 15 white spruce (*Picea glauca*) trees to assess the amount and nature of polymorphism revealed.

Materials and methods

PCR conditions

Primers were based upon arbitrarily selected black spruce cDNA sequences as previously described (Perry and Bousquet 1998). Thirty nine pairs of primers that had successfully directed amplification of genomic targets in black spruce were selected for this study. PCR with these STS primers was carried out using a standard set of conditions as previously described for genomic amplification (Perry and Bousquet 1998), and the products were subjected to electrophoresis through thin (3-mm) gels (2% agarose in TAE, except for *Sb01* products for which 1.2% agarose in TAE was used) followed by ethidium bromide staining.

Cross-species amplification

Needle samples for all multi-species panels were collected locally from arboreta or wild trees. Isolation of DNA from needles followed Bousquet et al . (1990) with an additional chloroform extraction. The suitability of DNA samples for PCR was tested in reactions using degenerate primers that were based on *copia*-like retrotransposon reverse transcriptases (see Voytas et al . (1992) for primer sequences and PCR conditions). These amplification trials were positive for all but the two *Taxus canadensis* samples which were subsequently replaced by DNA that was extracted using a DNeasy Plant Mini Kit (Qiagen) and amplified well.

Each pair of black spruce-derived STS primers was tested for its ability to direct amplification with two individuals from each of 12 coniferous species. This multi-species panel included four spruces (*Picea abies*, *P. glauca*, *Picea pungens* and *Picea rubens*), six additional members of the Pinaceae (*Abies balsamea*, *Larix laricina*, *Pinus banksiana*, *Pinus strobus*, *Pseudotsuga menziesii* and *Tsuga canadensis*) and two other conifers (*Taxus canadensis* of the Taxaceae and *Thuja occidentalis* of the Cupressaceae). Amplification was tentatively judged as positive if a single product was clearly predominant for both representatives of a species, or if it gave a single predominant band in one and produced none, or a banding pattern consistent with a heterozygous genotype, in the other. DNA sequences were obtained for a selection of amplification products (using a Perkin–Elmer–ABI model 373 automated DNA sequencer), with an emphasis on products for which homology was suspect owing to sizes that differed considerably from corresponding products in other species. For *Taxus canadensis* and *T. occidentalis*, sequence data were obtained in all cases judged tentatively positive regardless of product size. Sequences were manually aligned and compared to those previously obtained from black spruce.

In a second phase of cross-species trials, primer pairs that had scored positively in amplifications of *A. balsamea*, *L. laricina*, *P. banksiana* or *P. strobus* were tested against additional species. Primers that functioned in *A. balsamea* were tested against *Abies concolor* and *Abies lasiocarpa*. Where *L. laricina* had scored positively, *Larix gmelinii* and *Larix sibirica* were also tested. Additional hard pine (subgenus *Pinus*) trials were carried out with *Pinus mugo*, *Pinus nigra* and *Pinus sylvestris*, while for soft pines (subgenus *Strobus*), *Pinus cembra* and *Pinus koraiensis* were used. Two replicates of a single individual from each additional species were examined alongside the congeners initially tested.

Polymorphism within white spruce

Needles were collected from white spruce clones that originated from provenances in Ontario and Quebec and were established in the Cap Tourmente breeding orchard (Natural Resources Canada). We selected one clone from each of 15 provenances. Needle DNA was isolated following Bousquet et al . (1990), with an additional chloroform extraction. These 15 diploid DNAs were screened with each of the 39 pairs of STS primers for variation that was revealed directly by electrophoresis.

Markers that were polymorphic in the diploid white spruce panel were investigated further using haploid DNA of seeds from the 15 white spruce clones. Seedcoats and embryos were removed and DNA isolation from individual haploid megagametophytes followed Bousquet et al. (1990) modified to include a phenol : chloroform : isoamyl alcohol (25 : 24 : 1) extraction and precipitation with ethanol. Six megagametophyte DNAs were included per clone giving a total haploid panel size of 90. Examination of this haploid DNA panel served two purposes: (1) the observation of allelic segregation among haploid megagametophytes of heterozygous trees facilitated the interpretation of polymorphisms as dominant or co-dominant, and (2) null alleles that were undetected in a heterozygous state could be uncovered. With six megagametophytes, the probability of observing both alleles of a heterozygote is $1 - (1/2)^5 = 0.97$ if segregation is 1 : 1.

Results

Cross-species amplification

Nearly all (95–97%) of the black spruce-derived STS primers functioned well in congeneric trials (Table 1), with only three loci (*Sb52*, *Sb53* and *Sb66*) scoring negatively in one or more spruce species. Although all spruces scored positively for *Sb18*, one of the two *P. pungens* individuals was a null homozygote. These four loci are among the six at which null alleles were apparent in black spruce (Perry and Bousquet 1998).

Amplification success rates were lower with non-*Picea* templates. The proportion of primers scoring positively was remarkably consistent among *A. balsamea*, *L. laricina*, *P. menziesii* and *Tsuga canadensis* (31–33%) but the primers comprising these successful subsets differed considerably among the genera (Table 1). For *Pinus*, success rates dropped to 21–23%,

Table 1 Amplification trials of 39 black spruce-derived STS markers in 12 conifer species. Each species-primer combination was scored in two individuals for success (+) or failure (−) of putative or confirmed homologous amplification, putative length polymorphism (p), or the presence of a null homozygote (n). Products that were investigated by DNA sequencing are also indicated (s)

Locus	*Picea abies*	*Picea glauca*	*Picea pungens*	*Picea rubens*	*Abies balsamea*	*Larix laricina*	*Pseudotsuga menziesii*	*Tsuga canadensis*	*Pinus banksiana*	*Pinus strobus*	*Taxus canadensis*	*Thuja occidentalis*
Sb01	+p	+	+	+	+	+	+p	+	+s	+s	−	−
Sb06	+p	+	+	+	−	−	−	−	−	−	−	−
Sb07	+	+	+	+	+	−	+	−	−	−	−	−
Sb08	+	+	+	+	−	−	−	−	−	−	−	−
Sb11	+	+	+	+	−	−	−	−	−	−	−	−
Sb12	+	+	+	+	−	+s	+	+	−	−	−	−
Sb14	+	+	+	+	−	+	−	+	−	−	−	−
Sb16	+	+	+	+	−	−	−	−	−	−	−	−
Sb17	+	+	+	+	−	−	−	−	−	−	−	−
Sb18	+	+	+n	+	−s	−	−	−	−	−	−	−
Sb19	+	+	+	+	−	−	−	−	−	−	−	−
Sb21	+	+	+	+	−	−	+	−s	−	+	−s	−s
Sb24	+	+p	+	+	−	−	−	−	−	−	−	−
Sb28	+	+	+p	+	+	−	−	−	−	−	−	−
Sb29	+p	+p	+	+	+	−	+	−	−	−	−	−
Sb31	+	+	+	+	−	+	+	+	−	−	−	−
Sb32	+p	+ps	+	+p	−	−	−	+s	+p	+s	−s	−
Sb34	+	+	+p	+	+	+p	+	+	+	−	−	−
Sb35	+	+p	+p	+	−	−	−s	−	−	−	−	−
Sb36	+	+	+	+	+s	−	−	−	−	−	−	−
Sb41	+	+	+	+	−	+	+	+	−	+	−	−
Sb42	+p	+	+	+p	+	−	−	−	−	−	−	−
Sb46	+	+	+	+	+	+	−	+	+p	+	−	−
Sb49	+	+	+	+	+	+	+	+	+s	+	−	−
Sb50	+	+	+	+	−	−	−	−	−	−	−	−
Sb51	+	+	+	+	−	+	−	−	−	−	−	−
Sb52	−	−	+	−	−	−	−	−	−	−	−	−
Sb53	+n	−	−	+	−	−	−	−	−	−	−	−
Sb56	+p	+	+	+	+	+	+	−	−	−	−	−
Sb58	+	+	+	+	−	−	−	−	+	+	−	−
Sb60	+	+	+	+	+	+	−	+s	−	−	−	−
Sb62	+	+	+	+	+	+	+	−	−	−	−s	−
Sb64	+	+	+	+	−	−	+	+	+	−	−	−
Sb66	+	+	+	−	−	−	−	−	−	−	−s	−
Sb67	+	+	+	+	−	−	−	−	−	−	−	−
Sb68	+	+	+	+	−	−	−	−	+s	+	−	−
Sb70	+	+	+	+	+p	−	+	+	−	+	−	−
Sb71	+	+	+	+	−	−	−	−	−	−	−	−
Sb72	+	+	+	+p	−	−	−	−	−	−	−	−
Total positive	38	37	38	37	13	12	13	12	8	9	0	0

and outside of the Pinaceae amplification of homologous targets was not achieved.

Amplification of *Sb12* produced a double-banded pattern in both examples of *A. balsamea* tested and, in compliance with our judging criteria (see Materials and methods), this species-primer combination was given a negative score. Attempts to amplify *Sb12* in *A. concolor* and *A. lasiocarpa* also produced identical double-banded patterns, suggesting that the targeted gene was duplicated in *Abies*. A double-banded pattern was also observed for *Sb32* in both examples of *P. banksiana* tested (Fig. 1); however, we gave this combination a positive score because in other hard pines a single band was obtained for *Sb32*, suggesting that the pattern seen in *P. banksiana* was the result of a polymorphism at a single locus. For several other markers, polymorphisms were evident between the two individuals representing a species (Table 1). Except for *P. glauca* (see below), we did not examine these polymorphisms in haploid material, and therefore cannot comment on allelic segregation.

We investigated the homology of 18 PCR products by DNA sequencing (Table 1). Although one locus in *T. occidentalis* and four loci in *Taxus canadensis* were tentatively scored as positive, sequences of the amplified products bore no apparent similarity to the sequences originally targeted in black spruce. In general, the sizes of these products differed considerably from those amplified from spruces, but in one case (*Taxus canadensis Sb32*), a product was cleanly amplified and

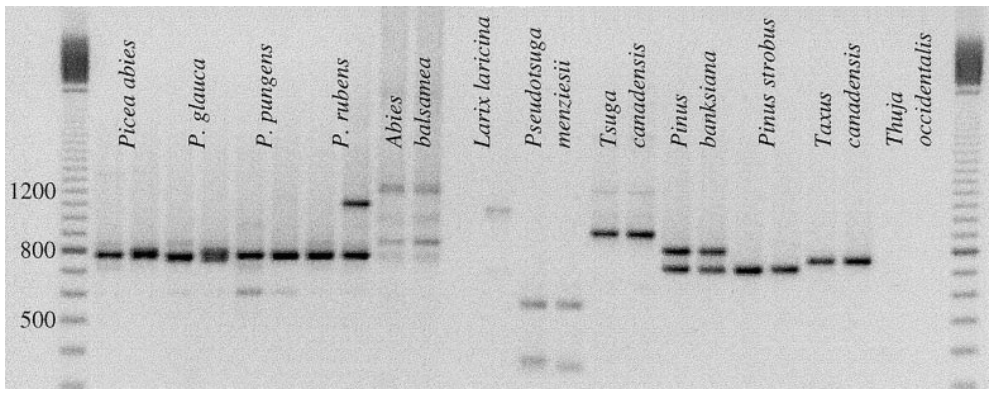

Fig. 1 Results of amplification trials of black spruce-derived STS locus *Sb32* in a panel of 12 conifers. A negative image of an ethidium bromide-stained agarose gel is shown. The size markers in the outside lanes are fragments of a 100-bp ladder (Pharmacia)

co-migrated with spruce *Sb32* products (Fig. 1). Even in this case, the *Taxus canadensis* amplification product was not homologous to the gene targeted in black spruce. Three species-primer combinations within the Pinaceae (*A. balsamea Sb18*, *P. menziesii Sb35* and *Tsuga canadensis Sb21*) were initially given positive scores, but the amplification products were later found to be nonhomologous to the targets in black spruce. In the first two of these cases, these results were not viewed as surprising since the products were several hundred bp larger than those obtained from spruces, and no other non-*Picea* panel members had scored positively. Conversely, a smaller size of the *Tsuga canadensis Sb*21 product had also alerted us to the possibility that it was not homologous to the fibrillarin gene targeted in black spruce. The genomic *Sb21* product of black spruce (471–474 bp) included a 92–93-bp intron and 95–97 bp of 3′-untranslated region (3′-UTR). If homologous, the *Tsuga canadensis Sb21* product (approximately 400 bp) would necessarily have lacked a large proportion of this noncoding sequence. In all other cases, sequencing confirmed the homology of amplified products and we are reasonably confident that the remaining untested products are true homologs of the sequences targeted in black spruce.

Some of the sequenced products (*P. glauca Sb32*, *P. strobus Sb32 and A. balsamea Sb36*) did not differ markedly in size from those amplified from black spruce and homology was clearly anticipated. However, in several cases homology was also confirmed among products that differed considerably in size. The most extreme case was *P. banksiana Sb01* which was 220 bp in comparison to about 1900 bp for *Sb01* in black spruce. Although the *P. banksiana Sb01* product closely resembled the aquaporin gene originally targeted in black spruce (92% nucleotide identity), it completely lacked an intron where one was predicted based upon sequences of similar genes in other plants (Guerrero and Crossland 1993; Kaldenhoff et al. 1993). *P. strobus Sb01* had a 382-bp intron at the expected location and the corresponding products of all soft pines were of very similar size. Larger *Sb01* product sizes in all of the other Pinaceae members examined (Fig. 2) were suggestive of much longer introns (>1500 bp).

The *Sb01* intron characterized in *P. strobus* contained a complex pattern of repeated sequences; one element of a direct repeat of 72 bp was interrupted by a tandem triplication of 26 bp. Several of the other sequenced products also contained large direct repeats that accounted for some or all of the difference in size compared to homologous black spruce products. The *L. laricina Sb12* product had a direct repeat of at least 49 bp that included the reverse primer site. *Tsuga canadensis Sb32*, *Tsuga canadensis Sb60* and *P. banksiana Sb68* products contained tandem repeats of 68 bp, 179 bp and 86 bp, respectively.

Some species had smaller *Sb49* products than did the spruces. The 3′-UTR of the *Sb49* sequence targeted in black spruce included a large direct repeat of at least 100 bp and the reverse primer (SB49-R) had been placed within the second element of this repeat at a position where the sequence differed slightly from the first element. In conjunction with primer SB49-F, the PCR yielded a 323-bp product in black spruce (Perry and Bousquet 1998) and we obtained similarly, perhaps identically, sized products from *L. laricina*, *P. menziesii* and each of the additional spruces tested. However, for *A. balsamea*, *Tsuga canadensis*, *P. banksiana* and *P. strobus*, *Sb49* products were about 170 bp. Sequencing of this smaller product in *P. banksiana* revealed that it was homologous to the black spruce target, but lacked the large repeat. Either the large repeat is absent in these species, or the sequences differ in other ways such that amplification from within the distal element is inhibited. Amplification of additional congeners indicated that *Sb49* product sizes were consistent within genera.

In general, any primers that scored positively in one member of a genus (or subgenus in the case of pines) were functional in all additional members of that genus (or subgenus) tested. Exceptions noted were: *Sb36* was positive in *A. balsamea* but did not function in *A. concolor* or *A. lasiocarpa*; *Sb58* was positive in all hard pines tested except for *P. mugo*; *Sb60* scored positively in *A. balsamea* and *A. lasiocarpa*, but not in *A. concolor*;

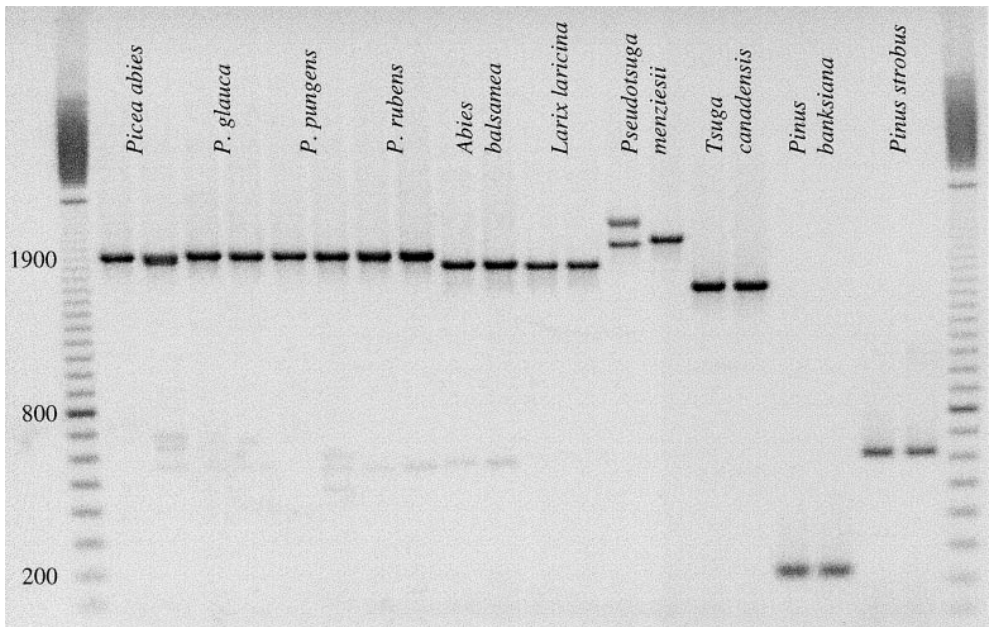

Fig. 2 Results of amplification trials of black spruce-derived STS locus *Sb01* in the ten Pinaceae members of a multi-species panel. A negative image of an ethidium bromide-stained agarose gel is shown. The size markers in the outside lanes are fragments of a 100-bp ladder (Pharmacia)

and *Sb01* was amplified in *P. banksiana*, but not in other hard pines.

Polymorphism in white spruce

Although two loci (*Sb52* and *Sb53*) were previously given negative scores in white spruce (see above), we screened all 39 black spruce-derived STS markers against the diploid white spruce panel. *Sb53* failed to amplify satisfactorily in any of the 15 panel members, while *Sb52* was successfully amplified from one individual. Null alleles were also apparent at an otherwise invariant *Sb66* where a single null homozygote was observed.

Based upon screening of the diploid panel, length polymorphisms were suspected at 18 loci. When corresponding products were examined in the haploid megagametophyte DNA panel, we were unable to detect length differences among those derived from apparent *Sb28* or *Sb46* heterozygotes, but allelic segregation was clearly evident for the others. In one case (*Sb16*) we departed from standard PCR reaction conditions and used an annealing temperature of 49°C rather than 55°C. This improved the intensity of bands obtained from megagametophytes but did not alter the patterns observed.

At 12 loci, all observed polymorphisms were co-dominant (Table 2, Fig. 3). In general, more slowly migrating heteroduplex products made it possible to detect heterozygotes using short gels (10 cm), but long gels (22 cm) were necessary to view allelic segregation and to assign diploid genotypes when differences among alleles were small. *Sb01* was highly variable, having a large number of co-dominant alleles. Although the largest differed from the smallest by about 150 bp, size differences among many *Sb01* alleles were slight and when all were ordered on the basis of length, the gradient appeared rather continuous, making an exact determination of the number of alleles problematic under current electrophoretic conditions (Fig. 4). With the number of alleles at *Sb01* estimated as ten, the average number of alleles at these 12 loci for which only co-dominant polymorphisms were evident was 3.2 and the average observed heterozygosity was 0.37. When

Table 2 Co-dominant length polymorhpisms revealed at 12 loci in a panel of 15 white spruce trees using black spruce-derived STS primers. Polymorphisms were directly observed on agarose gels without further manipulation of amplification products

Locus	Putative identification[a]	No. of alleles	Observed heterozygosity
Sb01	Aquaporin	~ 10[b]	0.80
Sb08	Unknown	2	0.20
Sb12	Nucleolysin TIA-1	2	0.27
Sb16	Ribosomal protein L13a	4	0.47
Sb17	Unknown	3	0.40
Sb21	Fibrillarin	2	0.07
Sb29	ATAF1 homolog	2	0.40
Sb32	Mitotic cyclin	3	0.60
Sb41	ISP42 homolog	2	0.60
Sb58	Ribosomal protein L5	2	0.13
Sb60	Unknown	3	0.33
Sb62	Ribosomal protein L15	3	0.13

[a] Putative identification was based on similarity to gene products identified in searches of public databases (Perry and Bousquet 1998)
[b] An exact determination of the number of alleles was problematic (see Fig. 4)

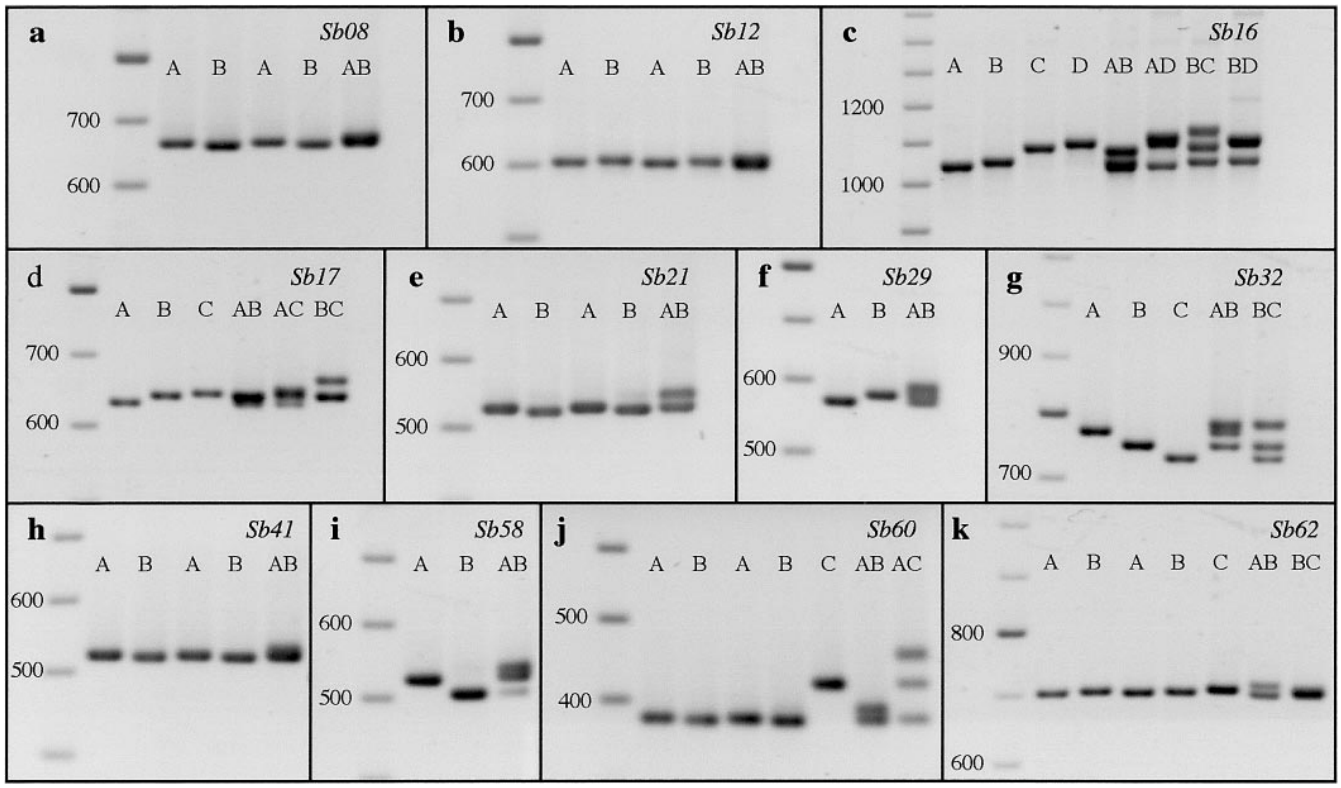

Fig. 3a–k White spruce alleles and heterozygous genotypes observed at 11 of the 12 black spruce-derived STS loci at which only co-dominant polymorphisms were evident in a panel of 15 white spruce trees. Products of individual alleles were PCR–amplified from DNA of single megagametophytes, while products of heterozygotes were amplified from DNA of needles. Polymorphisms were revealed by electrophoresis without further manipulation of PCR products. Negative images of ethidium bromide-stained agarose gels are shown. The assignment of alphabetic designations to alleles was arbitrary. The size markers in the left-hand lanes are fragments of a 100-bp ladder (Pharmacia)

Sb01 was excluded, these values were reduced to 2.5 and 0.33, respectively. A similar set of 12 loci in black spruce had an average of 2.8 alleles and an average observed heterozygosity of 0.26 in a range-wide sample of 22 trees (Perry and Bousquet 1998).

Various combinations of allele types were evident at the remaining polymorphic loci (co-dominant and null alleles, *Sb18*; co-dominant and dominant length variants, *Sb24* and *Sb35*; dominant length variant, *Sb71*) (Fig. 5). Again, *Sb18* was among the six loci at which null alleles were apparent in black spruce (Perry and Bousquet 1998). *Sb24* was polymorphic in black spruce, but only co-dominant alleles were observed in that species. In white spruce, allele *Sb24-C* was dominant over *Sb24-A*, but not over *Sb24-B*; however, in *Sb24-B*/*Sb24-C* heterozygotes, almost all of the *Sb24-B* allele product appeared to be incorporated into heteroduplex (Fig. 5 b). A dominant length polymorphism had also been observed at *Sb35* in black spruce, but the appearance of the dominant allele was unlike the one found in white spruce. Finally, although we classified the polymorphisms at *Sb71* as dominant, it was evident that the dominance of the smaller allele was not quite complete (Fig. 5 d). The difference in size between *Sb71* alleles was quite remarkable, about 1100 bp.

Discussion

Almost all of the STS primers that we developed for black spruce function well in other *Picea*. We considered four spruce species, including *P. glauca*, which was among the most genetically distant from black spruce of the 31 North American and Eurasian spruces previously included in a phylogenetic study of *Picea* (Sigurgeirsson and Szmidt 1993). Therefore, we expect our results to be representative for *Picea* in general. A reasonably high proportion (21–33%) of the primers also directed amplification in other Pinaceae genera. Since primers that functioned well in one member of a genus (or subgenus in the case of pines) were generally also successful in other members of that genus (or subgenus), our scoring of primer performance (Table 1) may provide a useful starting point when initiating studies involving any of a large number of "new" species.

In similar trials using cDNA-based STS primers of *C. japonica* (Tsumura et al. 1997), some primers were reported to be successful in amplifications across a much broader taxonomic range than was

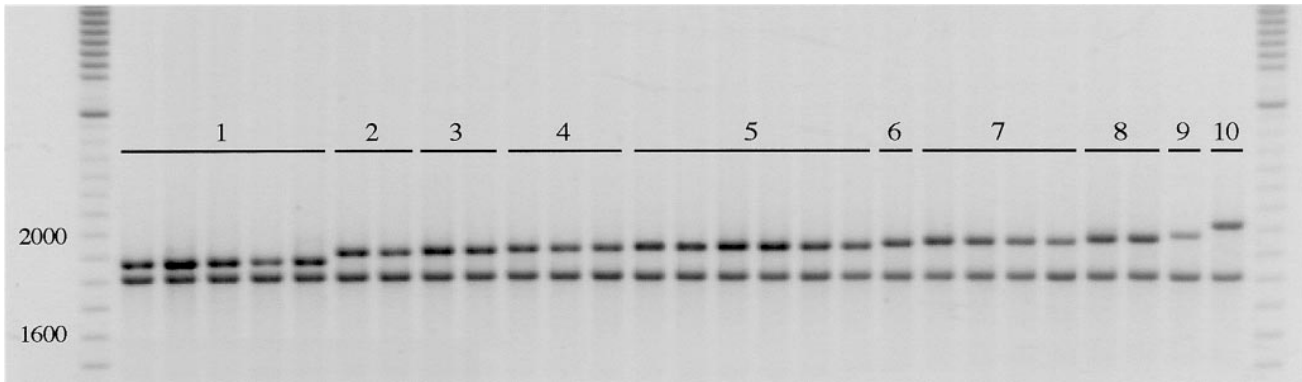

Fig. 4 A negative image of an ethidium bromide-stained agarose gel showing amplification products of black spruce-derived STS locus *Sb01* from 27 haploid white spruce megametophytes. Both alleles from each of 12 heterozygotes and a single allele from each of three homozygotes found in a panel of 15 white spruce trees are ordered according to length. The lower band in each lane is a size standard loaded in advance to facilitate comparison across the gel. *Numbered bars* represent the classification of alleles. The size markers in the outside lanes are fragments of a 100-bp ladder (Pharmacia)

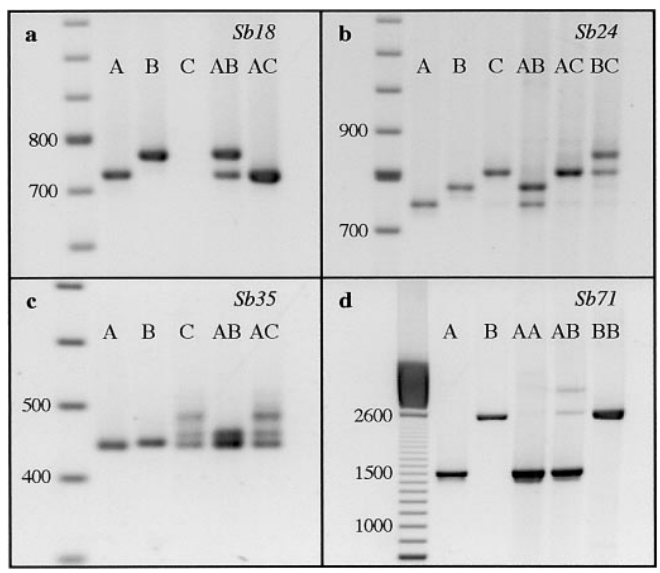

Fig. 5 a–d White spruce alleles and heterozygous genotypes observed at four black spruce–derived STS loci at which dominant polymorphisms were evident in a panel of 15 white spruce trees. Products from homozygous genotypes are also included for *Sb71* (panel d). Individual alleles were PCR–amplified from the DNA of single megagametophytes while diploid products were amplified from the DNA of needles. Polymorphisms were revealed by electrophoresis without further manipulation of PCR products. Negative images of ethidium bromide-stained agarose gels are shown. The assignment of alphabetic designations to alleles was arbitrary. The size markers in the left–hand lanes are fragments of a 100-bp ladder (Pharmacia)

achieved using our black spruce-derived primers. This could be a reflection of differing primer placement strategies. For each of our markers, we made an effort to locate the reverse primer in the 3′-UTR of the cDNA (Perry and Bousquet 1998). While this approach may have achieved our goals of ensuring the inclusion of noncoding, presumably more variable, DNA in the product and increasing specificity when primers were based upon one member of a gene family, it is also expected to adversely affect the taxonomic range across which primers can be transferred. This is because 3′-UTRs are likely to have diverged much more rapidly than protein-coding regions. Tsumura et al. (1997) did not provide details about the positioning of their amplification primers nor did they indicate whether their cDNA library was constructed using methods that would promote the inclusion of 3′-UTRs, but it is likely that in the absence of a specific effort to place reverse primers in the 3′-UTR, at least some would fall within the protein-coding region. Under such circumstances, primer sites would likely tend to be conserved across a wider range of taxa. However, we also note that Tsumura et al. (1997) did not report having verified the homology of any of their amplification products. In each of five cases where we had tentatively scored amplification as positive in species outside of the Pinaceae, homology of the amplification product was later disproved by DNA sequencing. This was true even in a case where a *Taxus canadensis* product was the same size as the product targeted in black spruce. It is possible, particularly in some of the more taxonomically distant trials, that Tsumura et al. (1997) may have inadvertently included some false positives.

How does cross-species transferability of cDNA-based STS markers compare to that of nuclear microsatellite (also known as simple-sequence-repeat, SSR) markers? At present, published data concerning the cross-species transferability of microsatellite markers in plants are only available from non-coniferous species. In one study (Whitton et al. 1997), just two of 13 sets of primers developed for *Helianthus* sp. amplified homologous microsatellite loci outside of the tribe Heliantheae and no single locus amplified in all species tested within that tribe. In investigations of microsatellite

markers developed for *Quercus myrsinifolia* (nine loci) (Isagi and Suhandono 1997) and *Quercus petraea* (17 loci) (Steinkellner et al. 1997), amplification success within the source subgenus was nearly complete while it was reduced by about one half in other *Quercus* subgenera and generally (but not always) declined markedly in other Fagaceae genera. In another study (White and Powell 1997) of 11 microsatellite markers developed for *Swietenia humilis*, four did not function in other *Swietenia* species, while three were successful in all six additional Meliaceae genera tested. Our ability to draw conclusions by comparing our results with these cross-species trials in angiosperms is clearly limited. Intuitively, one might expect cDNA-based primer sequences to be better conserved among taxa than microsatellite primers; however, we reiterate that our practice of placing one primer in the 3′-UTR most likely reduced the conservation across taxa. A proper evaluation must await the availability of results from cross-species trials of microsatellite markers performed within the Pinaceae.

Successful amplification of homologous targets is only one aspect of cross-species primer utility. With heterologous microsatellite markers, a reduced level of polymorphism compared to the primer source species is often observed. This trend has been reported for a variety animals (e.g., Fitzsimmons et al. 1995; Primmer et al. 1996; Ellegren et al. 1997) and also in plants (Steinkellner et al. 1997; van Treuren et al. 1997). In general, the greater the evolutionary distance, the greater the reduction in polymorphism. It has been argued that this reduction in polymorphism is due to the ascertainment bias inherent in microsatellite-marker development procedures (Ellegren et al. 1995). The sequences selected are not a random sample of simple-sequence-repeats from the source species; rather they are biased in favor of longer repeat regions that are more likely to be polymorphic. In a non-source species, the same set of loci would be more random with respect to length, making them shorter on average and less likely to be variable. Development of cDNA-based STS markers does not promote such an ascertainment bias, and therefore a reduction of polymorphism in non-source species is not expected. Indeed, the proportion of polymorphic loci that we observed in the panel of 15 white spruce was about the same as we had found for black spruce (Perry and Bousquet 1998) and the average observed heterozygosity at these polymorphic loci may actually be somewhat higher. Also, the composition of the subsets of polymorphic loci in the two spruces differed; some loci were polymorphic in black spruce but not in white spruce and vice versa. As for the other taxa that we examined, two individuals per species was clearly insufficient to allow a rigorous assessment of marker polymorphism; however, we did observe putative length polymorphisms in several instances, indicating that many of these markers may be polymorphic in diverse taxa.

We often found considerable variation in size among products amplified from our multi-species panel (e.g., Figs. 1 and 2). This contrasts with the report by Tsumura et al. (1997) that, under high-stringency conditions, the sizes of PCR products amplified in their 14-species, five-family panel were not different from those obtained from the primer source species (*C. japonica*). Fairly consistent product sizes might be expected if only conserved protein-coding regions are included, but the lengths of introns and UTR sequences are presumably under less selective constraint. With one of the loci that we examined (*Sb01*), differences in product sizes among taxa were due not only to presumed variation in intron length, but also to the presence or absence of an intron. The lack of an intron in the jack pine *Sb01* product suggests that we may have amplified a portion of a processed pseudogene in that case. Processed pseudogenes have been reported in *Picea* (Kvarnheden et al. 1995, 1998) and retrotransposon reverse transcriptase gene sequences are present in conifers (Voytas et al. 1992). Indeed, we used reverse transcriptase-targeted primers as a positive amplification control for our DNA. However, the *P. banksiana Sb01* reading frame was uninterrupted (57 codons) and, in comparison to the aquaporin protein inferred from the black spruce cDNA, there were only two conservative amino-acid differences, both having isoleucine substituted for valine. Therefore, it is also possible that a functional, but intron lacking, gene was amplified.

In many cases the size differences among STS markers from different taxa may be due to the presence or absence of large direct repeats. Of the ten products for which homology was confirmed by sequencing, seven differed notably in size from the corresponding products in spruce. The size differences for five of these seven products were explained by the presence or absence of large repeats. *Sb49* was particularly interesting because our ability to classify all of the Pinaceae examined into two groups (*Picea*, *Larix* and *Pseudotsuga*, or *Abies*, *Tsuga* and *Pinus*) based upon the presence or absence of a large direct repeat may be indicative of evolutionary relationships. We are becoming increasingly aware of the ubiquity of large direct repeats in conifer DNA sequences (Perry and Furnier 1996; Perry and Bousquet 1998).

It is encouraging that many of the STS markers that we developed for black spruce are successfully amplified in diverse species. These primers may provide a ready source of markers to initiate population studies of other species, at least within the Pinaceae. Heterologous primers also have many other potential applications such as in comparative genome mapping.

Acknowledgments 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. J. Beaulieu (Natural Resources Canada) kindly

provided the white spruce seeds, F. Larochelle and A. Roy helped with tissue collections, and P. Perry provided much assistance in the laboratory.

References

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

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias? Nature Genet 11 : 360–362

Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC (1997) Microsatellite evolution — a reciprocal study of repeat lengths at homologous loci in cattle and sheep. Mol Biol Evol 14 : 854–860

Fitzsimmons NN, Moritz C, Moore SS (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol Biol Evol 12 : 432–440

Guerrero FD, Crossland L (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

Isagi Y, Suhandono S (1997) PCR primers amplifying microsatellite loci of *Quercus myrsinifolia* Blume and their conservation between oak species. Mol Ecol 6 : 897–899

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

Kvarnheden A, Tandre K, Engström P (1995) A *cdc2* homologue and closely related processed retropseudogenes from Norway spruce. Plant Mol Biol 27 : 391–403

Kvarnheden A, Albert VA, Engström P (1998) Molecular evolution of *cdc2* pseudogenes in spruce (*Picea*). Plant Mol Biol 36 : 767–774

Perry DJ, Bousquet J (1998) Sequence-tagged-site (STS) markers of arbitrary genes: development, characterization and analysis of linkage in black spruce. Genetics 149 : 1089–1098

Perry DJ, Furnier GR (1996) Pinus banksiana has at least seven expressed alcohol dehydrogenase genes in two linked groups. Proc Natl Acad Sci USA 93 : 13020–13023

Primmer CR, Moller AP, Ellegren H (1996) A wide-range survey of cross–species microsatellite amplification in birds. Mol Ecol 5 : 365–378

Sigurgeirsson A, Szmidt AE (1993) Phylogenetic and biogeographic implications of chloroplast DNA variation in *Picea*. Nord J Bot 13 : 233–246

Steinkellner H, Lexer C, Turetschek E, Glössl J (1997) Conservation of $(GA)_n$ microsatellite loci between *Quercus* species. Mol Ecol 6 : 1189–1194

Tsumura Y, Suyama Y, Yoshimura K, Shirato N, Mukai Y (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 Treuren R, Kuittinen H, Kärkkäinen K, Baena-Gonzalez E, Savolainen O (1997) Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. Mol Biol Evol 14 : 220–229

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) *Copia*-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89 : 7124–7128

White G, Powell W (1997) Cross-species amplification of SSR loci in the Meliaceae family. Mol Ecol 6 : 1195–1197

Whitton J, Rieseberg LH, Ungerer MC (1997) Microsatellite loci are not conserved across the Asteraceae. Mol Biol Evol 14 : 204–209