Sequence-tagged-site (STS) markers of arbitrary genes: the amount and nature of variation revealed in Norway spruce

DANIEL J. PERRY*†, NATHALIE ISABEL‡ & JEAN BOUSQUET†

†*Centre de recherche en biologie forestière, Université Laval, Sainte-Foy, Québec, Canada G1K 7P4 and*
‡*Laurentian Forestry Centre, Natural Resources Canada, Sainte-Foy, Québec, Canada G1V 4C7*

We examined the amount and nature of variation revealed by cDNA-based sequence-tagged-site (STS) markers in Norway spruce (*Picea abies* (L.) Karst.) using 39 pairs of heterologous primers that were based upon arbitrary genes in black spruce (*Picea mariana* (Mill.) B.S.P.). A panel of 22 diverse Norway spruce genotypes was screened for variation that could be observed directly using standard agarose gel electrophoresis, without additional manipulation of amplification products. Examination of marker segregation among haploid megagametophytes revealed that nine markers behaved in a codominant manner, two markers had codominant length polymorphisms and null alleles, and four others had dominant length polymorphisms. DNA sequencing of codominant alleles at seven loci indicated that most insertions/deletions (indels) were in noncoding regions and that alleles often differed by the presence or absence of direct repeats that ranged in size from three to 23 bp. The nine markers that showed exclusively codominant polymorphisms in Norway spruce had an average observed heterozygosity of 0.30 and an average of 2.9 alleles in the panel of 22 trees. These levels of variation are similar to those previously found for similar sets of markers in other spruces, and appear to be at least as high as those revealed by polymorphic allozyme markers in Norway spruce. Polymorphisms at one STS locus suggested a higher affinity between Norway spruce and white spruce (*Picea glauca* (Moench) Voss) than between either of these spruces and black spruce. The STS markers described in this report should be useful in a variety of applications in Norway spruce, including population studies and genome mapping.

Keywords: allelic sequence variation, codominant PCR-based markers, conifer genes, multiplex PCR, *Picea abies*.

Introduction

Molecular markers are widely used in genetic studies of forest trees. Of currently available markers, those that are based on the polymerase chain reaction (PCR), such as random amplified polymorphic DNAs (RAPDs; Welsh & McClelland, 1990; Williams *et al.*, 1990) and simple sequence repeats (SSRs, also known as microsatellites; Tautz, 1989; Weber & May, 1989) are appealing because they require very small amounts of DNA and they may be essentially unlimited in number. Also, unlike allozyme markers, these DNA markers do not vary among tissues or developmental stages, and

they can provide direct access to the genomic regions that they mark.

Of PCR-based markers, RAPDs are the easiest to develop because they are technically simple and require no prior knowledge of DNA sequences. However, their dominant nature may lead to unreliable estimates of population genetic parameters when they are assessed in diploid materials (Isabel *et al.*, 1995, 1999; Szmidt *et al.*, 1996). Also, their lack of specificity can lead to genotyping errors caused by competition among PCR products in different genetic backgrounds (Hallden *et al.*, 1996) and they can be difficult to reproduce between different laboratories (Jones *et al.*, 1997). Such problems may be avoided by using codominant, specifically targeted markers. SSR markers generally fall into this latter category. However, assessment of allelic SSR variation can be rather technically demanding, often requiring high-resolution techniques such as

polyacrylamide gel electrophoresis followed by silver staining and, as for any class of specifically targeted PCR-based markers, the availability of suitable primer sequences may be limiting. Recently, we developed a set of primers to amplify sequence-tagged-site (STS) markers of arbitrarily selected genes in black spruce (*Picea mariana* (Mill.) B.S.P.) (Perry & Bousquet, 1998a). Several of these gene-specific markers have codominant polymorphisms in black spruce that may be detected by standard agarose gel electrophoresis, without further manipulation of amplification products. In cross-species trials using these same primers, amplification was often successful in other members of the Pinaceae family, and almost always successful in other spruce species (Perry & Bousquet, 1998b).

In an effort to increase the potential utility of these STS primers in a broader range of applications, we have now examined the variation that they reveal in Norway spruce. Norway spruce is an ecologically and economically important European conifer that has been the subject of much genetic research. A variety of questions have been addressed using allozymes (e.g. Lagercrantz & Ryman, 1990; Muona *et al.*, 1990; Krutovskii & Bergmann, 1995) and RAPDs (e.g. Lehner *et al.*, 1995; Bucci *et al.*, 1997; Scheepers *et al.*, 1997). Efforts have also been directed towards developing other types of molecular markers for this species, such as amplified-fragment-length-polymorphisms (AFLPs; Paglia & Morgante, 1998), SSRs (Pfeiffer *et al.*, 1997) and sequence-characterized-amplified-regions (SCARS; Scotti *et al.*, 1998). In the current report, we examine the amount and nature of variation that gene-specific STS markers reveal in this conifer. We also determine combinations of informative markers that are amenable to multiplexing, and characterize differences among many codominant STS alleles at the DNA sequence level.

Methods

PCR and electrophoretic conditions

Primer sequences were previously selected based upon arbitrarily chosen black spruce cDNA sequences (Perry & Bousquet, 1998a). Thirty-nine pairs of primers that had successfully directed amplification of genomic targets in black spruce were considered in this study. A standard set of reaction conditions was adopted, with PCRs containing 0.12 μM of each primer, 0.2 mM of each dNTP, 0.025 units/μL *Taq* DNA polymerase (Pharmacia Biotech or Boehringer Mannheim) and 1× of the supplied reaction buffer (included 1.5 mM MgCl$_2$). About 5–50 ng of template DNA was used in a reaction volume of 15 μL. Reaction mixtures were placed in a preheated (94°C) DNA Thermal Cycler (Perkin-Elmer) and PCR was carried out for 40 cycles (94°C, 1 min; 55°C, 2 min; 72°C, 3 min) followed by 10 min at 72°C. The ramp time to annealing and extension temperatures was 4 s/degree. Amplification products were electrophoresed through thin ($\approx$3 mm) gels (2% agarose in TAE, except for *Sb01* and *Sb71* for which 1.2% agarose in TAE was used). The bromophenol blue tracking dye was allowed to migrate to the end of a 22-cm gel (about 6.5 h at 130 V) and the gels were stained with ethidium bromide.

Screening of provenance trees

Twenty-two Norway spruce trees were selected, each from a different provenance represented in test plantations that were established by the Canadian Forest Service near Quebec City (Corriveau *et al.*, 1988). These provenances had origins scattered throughout several central European countries (Fig. 1). DNA was isolated from needles of each selected tree using a DNeasy Plant Mini Kit (Qiagen) and subjected to PCR using each of the 39 pairs of primers. PCR products obtained from this diploid panel were then examined for polymorphisms that were revealed by electrophoresis, without further manipulation.

Segregation of polymorphic markers

All markers found to be polymorphic in the provenance tree panel were examined in a panel of 90 haploid

Fig. 1 Origins of 22 Norway spruce provenances from which 22 trees that comprised the screening panel were selected.

DNAs, obtained from six megagametophytes from each of 15 trees that were selected from the diploid panel. These 15 trees included individuals that represented every electrophoretic banding pattern observed on the diploid panel for each polymorphic marker, with the exception of *Sb01*. *Sb01* was also examined in six megagametophytes from each of the seven remaining members of the diploid panel.

DNA of haploid tissues was isolated from individual megagametophytes as follows. The seed coat and embryo were removed and the megagametophyte was ground in 200 μL of extraction buffer (2% CTAB, 100 mM Tris pH 8.0, 20 mM EDTA pH 8.0, 1.4 M NaCl and 1% PVP mw 40 000), incubated at 65° for 1 h, extracted first with phenol:chloroform:isoamyl alcohol (25:24:1), and then with chloroform:isoamyl alcohol (24:1), followed by precipitation with ethanol. The pellet was washed with 70% ethanol, dried, and resuspended in 100 μL of autoclaved nanopure H_2O. One μL of DNA solution prepared in this manner was used per 15 μL PCR.

Production of synthetic heterozygotes

For codominant markers, heterozygote genotypes having allele combinations that were not found in the panel of provenance trees were simulated. Approximately equal quantities of allelic products that were amplified from megagametophyte DNAs in separate PCRs were combined and then subjected to five additional thermal cycles as used in amplification. This process of melting and reannealing was to allow the formation of a mixture of heteroduplex and homoduplex products as expected of actual heterozygotes.

Sequence comparisons of codominant alleles

PCR products were amplified from single megagametophytes with reaction conditions as above, except in larger volumes (50 μL), and purified using a QIAquick PCR purification kit (Qiagen). Sequencing was carried out using a Perkin-Elmer-ABI model 373 automated DNA sequencer, from the same forward primer as used for template amplification. Allelic sequences were aligned manually.

Multiplex PCR

Combinations of markers that exhibited strictly codominant polymorphisms were chosen on the basis that they did not have overlapping primary, or heteroduplex product sizes. No attempt was made to assess primer sequences for potential incompatibilities. PCR was

carried out as described above, but using from two to four pairs of primers in a single reaction.

Results

Markers showing codominant polymorphisms

Of the 39 pairs of primers tested against the 22-member diploid panel, 11 revealed length polymorphisms that were inferred to be codominant following examination of allelic segregation in the haploid megagametophyte panel. However, for two of these loci (*Sb06* and *Sb17*), null-amplification alleles were also uncovered in the haploid panel. Although *Sb17* appeared to have several codominant alleles, we did not investigate this marker further. We judged its potential utility in population studies to be very limited because the codominant alleles differed only slightly in size and genotype scoring would be further complicated by the occurrence of null alleles. Although *Sb06* also had null-amplification alleles, we did proceed with an investigation of its positive-amplification alleles because this was one of the loci at which polymorphisms had been characterized previously in black spruce (Perry & Bousquet, 1998a). Sequencing revealed that the two less frequent alleles (*Sb06–546a* and *Sb06–546b*) were equal in size, each differing from the common allele (*Sb06–539*) by a different insertion of 7 bp in the 3′-untranslated region (UTR). *Sb06–546a* and *Sb06–546b* were easily distinguished in heterozygotes (either natural or synthetic) by their characteristic heteroduplex products (Fig. 2b).

If only the nine loci having exclusively codominant polymorphisms were considered, the average observed heterozygosity (H_o) was 0.30 with an average of 2.9 alleles (Table 1). Size differences among alleles at these loci ranged from 3 bp (*Sb08*, *Sb29* and *Sb51*) to an estimated 500 bp (*Sb01*) (Fig. 2). In cases where the size differences among alleles were small, the formation of more slowly migrating heteroduplex products often aided in the identification and classification of heterozygous genotypes present in the diploid panel (e.g. Fig. 2e). One may also take advantage of distinctive heteroduplex patterns to verify the correct assignment of homozygote genotypes by constructing synthetic heterozygotes.

In order to understand better the nature of allelic polymorphism, the DNA sequences of all observed allelic products were determined for each of the nine loci that had exclusively codominant polymorphisms, except for *Sb01* and *Sb71* (see below). Twenty Norway spruce DNA sequences generated here have been deposited in the GenBank database (accession nos AF127432–AF127451). Almost all of the polymorphisms investigated at the DNA level were found to
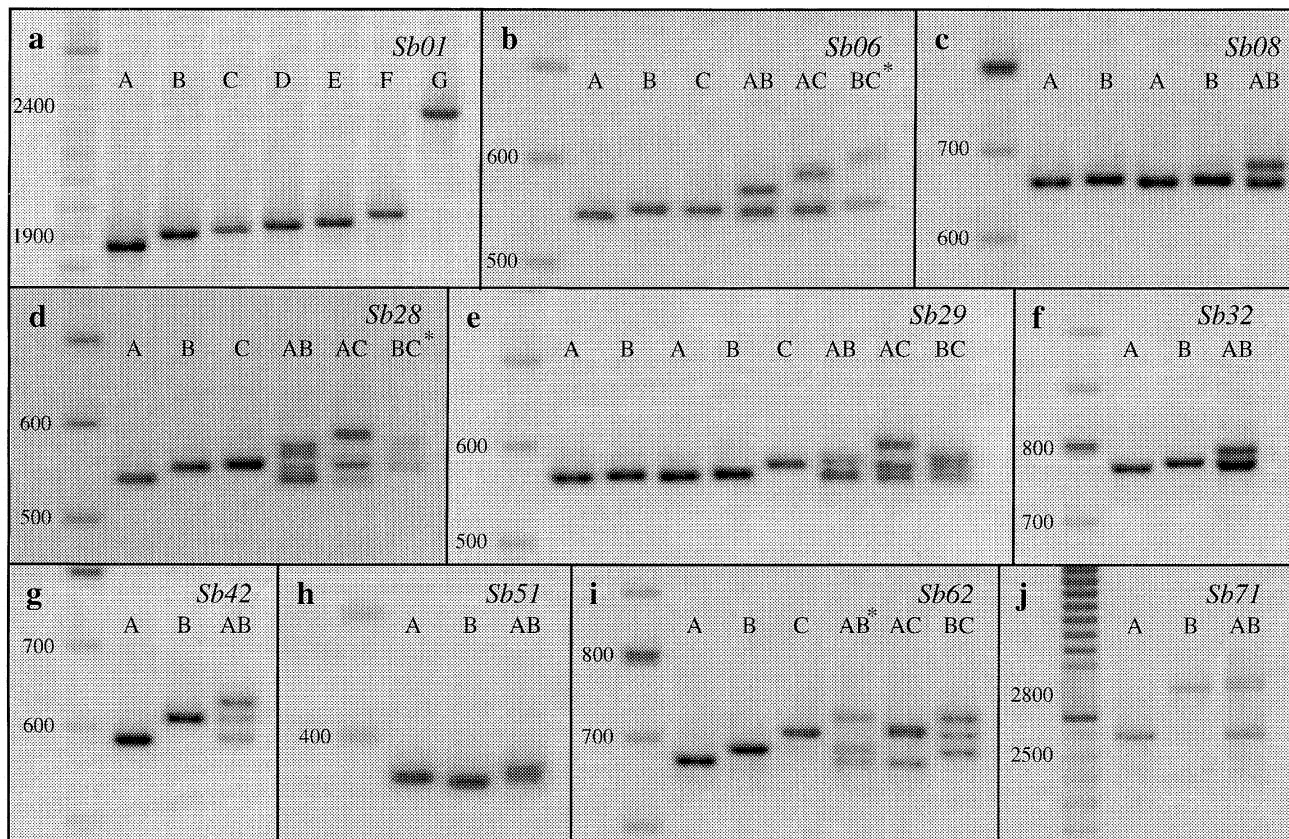
Fig. 2 Codominant cDNA-based sequence-tagged-site (STS) markers in Norway spruce. Polymorphisms were observed on ethidium bromide-stained agarose gels without additional manipulation of amplification products. Negative images are shown. Products of individual alleles were PCR-amplified from the DNA of single megagametophytes whereas heterozygote products were amplified from the DNA of needles or, in cases where the genotype was not found on a 22-tree screening panel, synthetic heterozygote products are shown (marked by asterisks, see Methods). Size markers (left-hand lanes) are fragments of a 100-bp ladder (Pharmacia). (a) A through G are alleles *Sb01–1860, Sb01–1900, Sb01–1910, Sb01–1920, Sb01–1930, Sb01–1970* and *Sb01–2360.* (b) A, B and C are alleles *Sb06–539, Sb06–546a* and *Sb06–546b,* respectively. Note that null alleles were also observed at this locus. (c) A and B are alleles *Sb08–647* and *Sb08–650,* respectively. (d) A, B and C are alleles *Sb28–535, Sb28–547* and *Sb28–551,* respectively. (e) A, B and C are alleles *Sb29–561, Sb29–564* and *Sb29–574,* respectively.(f) A and B are alleles *Sb32–771* and *Sb32–778,* respectively. (g) A and B are alleles *Sb42–584* and *Sb42–607,* respectively. (h) A and B are alleles *Sb51–358* and *Sb51–355,* respectively. (i) A, B and C are alleles *Sb62–672, Sb62–684* and *Sb62–706,* respectively. (j) A and B are alleles *Sb71–2550* and *Sb71–2800,* respectively.

result from one or more insertions/deletions (indels) in introns (*Sb08, Sb32* and *Sb62*), or the 3′-UTR (*Sb28, Sb42* and *Sb51*). Polymorphisms at *Sb29* were noteworthy because they occurred within the protein-coding region, as was also the case for *Sb29* polymorphisms in black spruce (Perry & Bousquet, 1998a). Relative to the common allele (*Sb29–564*), alleles *Sb29–561* and *Sb29–574* would encode proteins having a deletion of one amino acid and an insertion of three amino acids, respectively (this latter allele also had a 1-bp insertion in its 3′-UTR).

In many cases, indel polymorphisms involved the presence or absence of a direct repeat; however, the sizes of these repeats were generally small. *Sb08–650* had a 4-bp repeat that *Sb08–647* lacked; relative to *Sb28–535*, *Sb28–547* had a direct repeat of 12 bp, and *Sb28–551* had this same 12-bp repeat plus an additional 4-bp repeat; *Sb29–561* had two elements of a 'TCC' repeat, where *Sb29–564* and *Sb29–574* had three; and *Sb42–607* had a repeat of 23 bp that was not present in *Sb42–584*. Two markers had large direct duplications (*Sb06,* 70 bp; *Sb42,* 106 bp) that were as previously observed in black spruce (Perry & Bousquet, 1998a). However, the 70 bp repeat in *Sb06* of Norway spruce was not found to vary in copy number as it did in black spruce. The 106 bp repeat in *Sb42* has not been found to vary in copy number in Norway spruce (this study), nor in black spruce (Perry & Bousquet, 1998a).

Table 1 Codominant length polymorphisms revealed by nine STS markers in a panel of 22 Norway spruce trees. Polymorphisms were directly observed on agarose gels without additional manipulation of amplification products

Locus	Putative identification†	Allele frequencies‡	Observed heterozygosity
Sb01	Aquaporin	0.07, 0.57, 0.02, 0.11 0.11, 0.09, 0.02	0.64
Sb08	Unknown	0.95, 0.05	0.09
Sb28	Nucleosome assembly protein	0.07, 0.91, 0.02	0.14
Sb29	ATAF1	0.09, 0.80, 0.11	0.36
Sb32	Mitotic cyclin	0.27, 0.73	0.45
Sb42	Ribosomal protein L17	0.73, 0.27	0.45
Sb51	Ribosomal protein L3	0.07, 0.93	0.14
Sb62	Ribosomal protein L15	0.09, 0.02, 0.89	0.23
Sb71	TAT-binding protein homologue	0.91, 0.09	0.18

†Putative identification was based on similarity to gene products identified in searches of public databases (Perry & Bousquet, 1998a or, for *Sb28*, more recent unpublished results).
‡Alleles are listed in order of increasing size (see Fig. 2).

Alleles of *Sb01* or *Sb71* were not sequenced owing to their large sizes (about 1900–2800 bp) that would have made it necessary to employ several additional sequencing primers. With seven alleles and an H_o of 0.64, *Sb01* was the most variable locus. Because the amplification primer locations for this locus would exclude UTR sequences (Perry & Bousquet, 1998a), the observed polymorphisms probably occurred within the single large intron that is predicted to be included in *Sb01* products, based upon the locations of introns in similar genes of other plants (Guerrero & Crossland, 1993; Kaldenhoff *et al.*, 1993). If there is just one intron included, its size range is estimated to be from about 1640–2140 bp among the alleles observed in Norway spruce. The other large marker having exclusively codominant polymorphisms, *Sb71*, was the largest marker for which we have detected polymorphism. Additional *Sb71* alleles, having sizes similar to the common allele, were also apparent in a few putatively heterozygous individuals on the diploid panel; however, only the large allelic variant was reliably scored and it alone is reported here.

Markers showing dominant length polymorphisms

Four pairs of primers revealed length polymorphisms that were found to behave in a dominant fashion when examined in the megagametophyte panel. Of these, polymorphisms of *Sb35* were the most complex. Nine different multibanded alleles segregated in 11 of the 15 trees that contributed to the megagametophyte panel; shown in Fig. 3(a) are four of these alleles that segre-

gated among megagametophytes of two individuals. Although it is possible that some *Sb35* allele combinations might behave in a codominant manner, the overall complexity of this marker would make scoring problematic in diploid tissues.

The remaining dominant length polymorphisms were more straightforward (Fig. 3), each having one multibanded allele that was dominant over the common, single-banded allele. We did not investigate these loci further, but we can suggest at least two mechanisms by which a multibanded allele might arise. It could reflect a sequence duplication that caused repetition of a primer site, or it could reflect length polymorphism within a haplotype between duplicate copies of the entire region. Based upon segregation among megagametophytes of the 15 trees that contributed to the megagametophyte panel, the H_o values were 0.07, 0.33 and 0.33 for *Sb18*, *Sb68* and *Sb72*, respectively.

Other markers

Of the remaining 24 pairs of primers screened against the diploid panel, 22 consistently yielded products that had the same or similar sizes as those amplified from black spruce. All but two of these (*Sb56* and *Sb67*) appeared invariant under current electrophoretic conditions. *Sb56* and *Sb67* produced double-banded patterns in diploid amplifications of some trees, but no length variation was detected among corresponding haploid products amplified from megagametophytes. The extra bands obtained from some diploid templates may have been heteroduplex products formed between alleles that
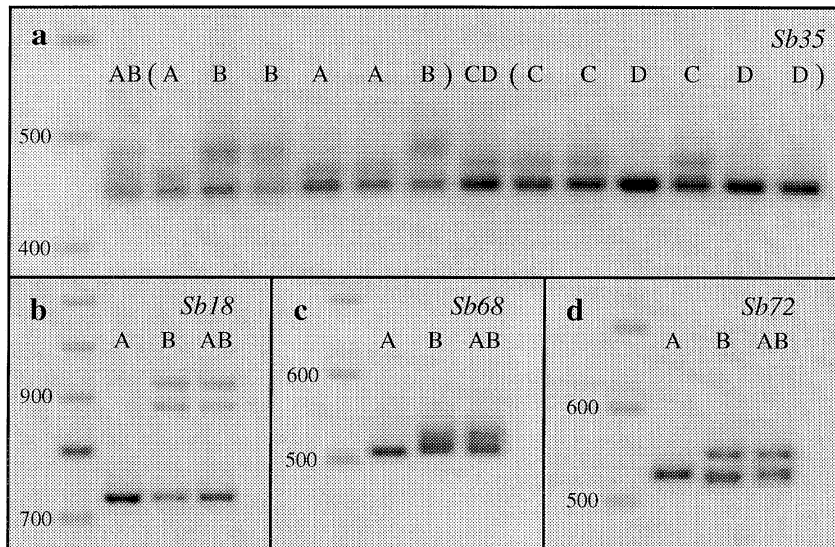
Fig. 3 Dominant cDNA-based sequence-tagged-site (STS) markers in Norway spruce. Polymorphisms were observed on ethidium bromide-stained agarose gels without additional manipulation of amplification products. Negative images are shown. Individual alleles were PCR-amplified from DNA of single megagametophytes whereas heterozygote products were amplified from DNA of needles. Assignment of alphabetic designations to alleles was arbitrary. Size markers in left-hand lanes are fragments of a 100-bp ladder (Pharmacia). For *Sb35* (panel a), allelic segregation is shown among six megagametophytes from each of two different heterozygotes (parentheses indicate family membership). Several additional alleles that were detected at this locus are not shown.

were identical, or almost identical in size, but differed in sequence.

The final two pairs of primers gave unsatisfactory results. Attempts to amplify *Sb52* yielded only an invariant product (about 200 bp), but this was much smaller than the minimum size (756 bp) that was predicted based upon the original black spruce cDNA sequence. Similarly, null-amplification alleles were common at *Sb53* where amplification of the expected product was sporadic among panel members, with no evident variation in length. Again, amplification products included additional smaller fragments.

Multiplexing of codominant STS markers

In general, attempts to multiplex codominant STS markers were successful. Using only three multiplex combinations, it was possible to amplify all nine of the markers for which only codominant polymorphisms had been observed (Fig. 4). *Sb01* and *Sb71* were amplified together because it was preferable to resolve them both on 1.2% agarose gels. The remaining combinations (*Sb29/Sb62/Sb51* and *Sb08/Sb28/Sb32/Sb42*) were chosen because they had compatible product sizes and were all routinely resolved on 2% agarose gels. Although no effort was made to avoid possible primer interactions, no such problems were apparent. However, band intensities were not always even among loci.

In our experience, a second photograph with a longer exposure was necessary for reliable scoring of *Sb28* heterozygotes, which tended to be fainter than homozygotes (Fig. 4b).

Discussion

In Norway spruce, cDNA-based STS markers often reveal allelic variation that may be observed directly on agarose gels, without any additional manipulation of PCR products. Although the primers used in this study were initially developed for black spruce, the transferability across species was almost complete, with all but two of the 39 pairs of primers tested being capable of directing the amplification of expected products from the diploid DNAs of all 22 Norway spruce trees included on our screening panel. Of markers that were polymorphic in these trees, nine showed only codominant polymorphisms and should be well suited to population studies. We have also demonstrated that these nine markers can be multiplexed, a practice that should increase their convenience while lowering costs. The multiplex combinations that we have suggested provide an average of three loci per reaction, but other combinations may also be possible.

The number of markers showing only codominant polymorphisms in Norway spruce was slightly lower than the 12 previously found in both black spruce and
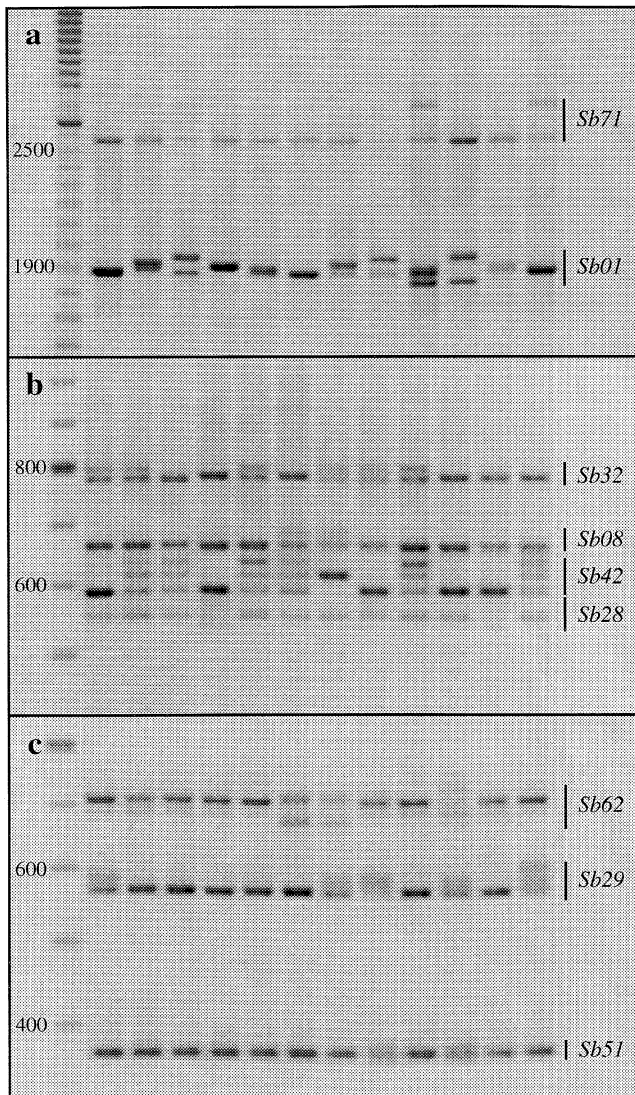
Fig. 4 Multiplexed codominant cDNA-based STS markers in Norway spruce. Panels a, b and c show products of three different multiplex PCR combinations (loci are as indicated) amplified from the first 12 members of a 22-tree screening panel. Polymorphisms were observed on ethidium bromide-stained agarose gels without additional manipulation of amplification products. Negative images are shown.

white spruce (*Picea glauca* (Moench) Voss) (Perry & Bousquet, 1998a,b) using this same set of 39 primer pairs. However, in terms of heterozygosity, the amount of variation revealed by the nine markers in Norway spruce ($H_o = 0.30$) was about midway between that revealed by the similar sets of 12 markers in black spruce ($H_o = 0.26$) and white spruce ($H_o = 0.37$). The average number of alleles per locus for Norway spruce (2.9) also fell between the corresponding values observed for black spruce (2.8) and white spruce (3.2).

The variability of these nine polymorphic STS markers in Norway spruce is at least as high as that previously reported for allozymes of that species. For example, considering 18 polymorphic allozyme markers, Müller-Starck (1995) reported an average observed heterozygosity of 0.226 and 2.52 alleles per locus for 20 high-elevation populations of Norway spruce in Switzerland. Similarly, based on eight polymorphic allozyme loci, Bergmann & Ruetz (1991) reported an average observed heterozygosity of 0.23 and 2.21 alleles per locus for three Bavarian Norway spruce stands. Therefore, these STS markers may be about as informative as allozymes, perhaps slightly more so. However, STS markers do have practical advantages over allozymes; they do not vary among tissue types or developmental stages, and the actual number of STS markers possible is virtually unlimited. Furthermore, whereas allozyme markers require different staining protocols for each additional enzyme system assayed, these STS markers all follow one standard protocol.

In general, cDNA-based STS markers may represent a wider variety of genes than allozymes, which are restricted to only those genes that encode known enzymes for which specific histochemical staining protocols have been derived. Gene functions need not be known for STS markers but, based upon comparative searches of public databases, many of the cDNAs sequenced in black spruce have been putatively identified as encoding enzymes or components of various cellular structures (Perry & Bousquet, 1998a). Of the nine STS markers showing only codominant polymorphisms in our panel of Norway spruce, eight have been assigned putative identifications (Table 1).

Most of these nine markers have also revealed length polymorphism in black spruce (Perry & Bousquet, 1998a) or white spruce (Perry & Bousquet, 1998b) or both; however, two markers (*Sb28* and *Sb51*) were invariant in these other spruces. Four markers (*Sb01*, *Sb08*, *Sb29* and *Sb62*) have revealed codominant length polymorphisms in all three of the spruces that have been examined in detail, suggesting that some loci may be more prone to mutations that affect length than are others.

The allelic DNA sequences that we have determined here for codominant Norway spruce STS markers, and those previously obtained for STS markers in black spruce (Perry & Bousquet, 1998a) indicate that many of the observed length polymorphisms are caused by the presence or absence of direct repeats. However, although some large repeats (up to 106 bp) are present in these Norway spruce sequences, differences among alleles generally involve only small repeats, with the longest, 23 bp, found in *Sb42–607*. Nearly all of the length polymorphisms investigated occur in noncoding
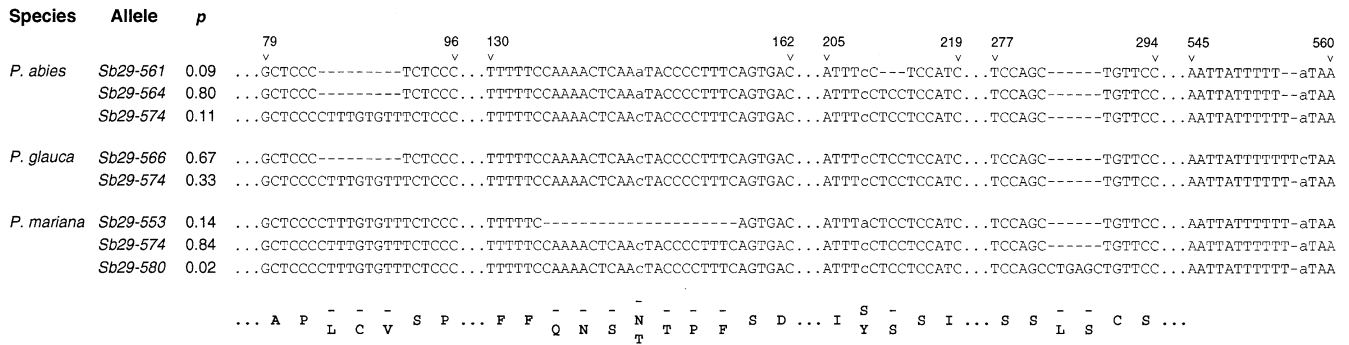
Fig. 5 Partial alignment of nucleotide sequences of *Sb29* alleles of Norway spruce (*Picea abies*), white spruce (*P. glauca*) and black spruce (*P. mariana*) showing the distribution of indels among the sequences and allele frequencies (*p*) in *n*-tree panels, each originating from multiple populations (*n* = 22, 15 and 22 for *P. abies, P. glauca* and *P. mariana*, respectively). The amino acid translation is shown under the alignment. The stop codon is located at nucleotide positions 496–498. GenBank accession nos (ordered as sequences appear in the figure) are AF127441, AF127440, AF127442, AF127599, AF127600, AF051748, AF051749 and AF051750.

regions. Therefore, these STS markers may be less likely than allozyme variants to be visible to selection.

Sb29 is an interesting exception because length differences observed among three alleles in Norway spruce are caused by indels in the protein-coding region. In black spruce, indels among three *Sb29* alleles also occur in the protein-coding region (Perry & Bousquet, 1998a), but differ in size and location from those found in Norway spruce (Fig. 5). We have now also obtained the sequences of two *Sb29* alleles previously detected in white spruce (Perry & Bousquet, 1998b); they differ by the same nine bp (three amino acid) indel that also varies between Norway spruce alleles (Fig. 5). Whether any of these *Sb29* polymorphisms affects protein function is not known, nor has the role of the *Sb29* gene product been determined. We previously found (Perry & Bousquet, 1998a) that the encoded protein shares sequence similarity with ATAF1 of *Arabidopsis thaliana* (encoded by GenBank accession no. X74755) for which a function has not been reported.

It is noteworthy that the same nine-bp stretch is absent from the common *Sb29* allele in both Norway spruce and white spruce, but is present in all black spruce alleles. Application of the principle of parsimony would indicate that the mutation leading to this polymorphism was probably a deletion that predated the divergence between Norway spruce and white spruce, suggesting a rather long allele coalescence time. This shared polymorphism may also indicate that Norway spruce and white spruce have a more recent common ancestry than either species has with black spruce, a phylogenetic view that is not supported by relative numbers of substitutions estimated from chloroplast restriction fragment variation (Sigurgeirsson & Szmidt, 1993). However, this view is supported by morphological data (Wright, 1955) and also by sequence data that we have recently obtained for another nuclear STS marker, *Sb16* (Fig. 6). Additional data will be necessary if the systematic relationships of these three taxa are to be resolved conclusively. As more sequence data from additional gene loci become available for these and other species, there may be greater opportunities to investigate the evolution of allelic variation and its utility in phylogenetic analysis. One locus of particular interest is *Sb01*, which encodes an aquaporin. *Sb01* is highly variable in all three spruces investigated to date (this report; Perry & Bousquet, 1998a,b), but genomic sequence data are currently not available for this locus.

Species	Sample	70	132	294	401	431	441	499	639	703
P. abies	A	G	C	A	A	GT	G	C	T	G
	B	G	C	A	A	GT	G	C	T	G
P. glauca	A	G	C	A	A	GT	G	T	T	A
	B	G	C	A	A	GT	G	C	T	A
P. mariana	A	A	T	T	T	--	T	T	A	A
	B	A	T	T	T	--	T	T	A	A

Fig. 6 Partial alignment of *Sb16* nucleotide sequences, encoding a putative ribosomal protein (Perry & Bousquet, 1998a), from Norway spruce (*Picea abies*), white spruce (*P. glauca*) and black spruce (*P. mariana*). Each sequence was obtained from the DNA of a single megagametophyte, with samples (A and B) within a species originating from different provenances. Only sites that were polymorphic among species are shown. Additional sites that were only variable within a single species are not included. The substitution at position 70 is located within the protein-coding region, whereas the remaining polymorphisms are found within two introns. No length variation was observed at this locus in Norway spruce. GenBank accession nos are AF127593 through AF127598.

In Norway spruce, our interest has focused primarily on the nine markers that show only codominant polymorphisms, owing to their anticipated convenience in population studies. However, two other markers that had both codominant and null alleles, and four more for which dominant length polymorphisms were found, may be useful in other applications such as genome mapping. However, considering the levels of observed heterozygosity present, it is unlikely that many of the polymorphisms that we have described would segregate simultaneously in any single Norway spruce tree or mapping population. When developing similar cDNA-based STS markers in loblolly pine (*Pinus taeda*), Harry *et al*. (1998) found that restriction digestion of PCR products (PCR-RFLP) was necessary to reveal codominant polymorphisms within two mapping populations that were the progeny of two controlled crosses. Higher resolution techniques such as PCR-RFLP or single-strand conformation polymorphism (SSCP) analysis may also be necessary with our markers to reveal sufficient variation in single Norway spruce trees or mapping populations.

Genetic maps of Norway spruce have already been constructed using RAPDs (Binelli & Bucci, 1994; Bucci *et al*., 1997) or a combination of marker types including AFLPs, selective amplification of microsatellite polymorphic loci (SAMPLs) and SSRs (Paglia *et al*., 1998). Because individual cDNA-based STS markers may be conserved across other species within the Pinaceae (Perry & Bousquet, 1998b), they may be potentially useful as anchor loci if incorporated into such maps. Hence, they could aid in the integration of individual-tree maps of Norway spruce and also in comparative mapping among more diverse taxa. Such mapping efforts may provide insights into genome structure and evolution in spruces and other members of the Pinaceae.

We have demonstrated that cDNA-based STS markers that were originally developed in black spruce are a source of convenient, technically simple genetic markers in Norway spruce. It is our hope that these markers will find many practical applications and will contribute to a growing understanding of the genetics of Norway spruce and other conifers.

Acknowledgements

References

BERGMANN, F. AND RUETZ, W. 1991. Isozyme genetic variation and heterozygosity in random tree samples and selected orchard clones from the same Norway spruce populations. *For. Ecol. Manag*., **46**, 39–47.

BINELLI, G. AND BUCCI, G. 1994. A genetic linkage map of *Picea abies* Karst., based on RAPD markers, as a tool in population genetics. *Theor. Appl. Genet*., **88**, 283–288.

BUCCI, G., KUBISIAK, T. L., NANCE, W. L. AND MENOZZI, P. 1997. A population 'consensus', partial linkage map of *Picea abies* Karst. based on RAPD markers. *Theor. Appl. Genet*., **95**, 643–654.

CORRIVEAU, A., BEAULIEU, J. AND DAOUST, G. 1988. Phenotypic stability and productivity of central European Norway spruce provenances in Québec, Canada. *IUFRO Working Party S 2.02-11 meeting*, Tjörnarp, Sweden.

GUERRERO, F. D. AND 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.

HALLDEN, C., HANSEN, M., NILSSON, N. O., HJERDIN, A. AND SALL, T. 1996. Competition as a source of errors in RAPD analysis. *Theor. Appl. Genet*., **93**, 1185–1192.

HARRY, D. E., TEMESGEN, B. AND NEALE, D. B. 1998. Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. *Theor. Appl. Genet*., **97**, 327–336.

ISABEL, N., BEAULIEU, J. AND BOUSQUET, J. 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. U.S.A*., **92**, 6369–6373.

ISABEL, N., BEAULIEU, J., THÉRIAULT, P. AND BOUSQUET, J. 1999. Direct evidence for biased gene diversity estimates from dominant random amplified polymorphic DNA (RAPD) fingerprints. *Mol. Ecol*., **8**, 477–483.

JONES, C. J., EDWARDS, K. J., CASTAGLIONE, S., WINFIELD, M. O., SALA, F., VAN DE WIEL, C. ET AL. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed*., **3**, 381–390.

KALDENHOFF, R., KOLLING, A. AND 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.

KRUTOVSKII, K. V. AND BERGMANN, F. 1995. Introgressive hybridization and phylogenetic relationships between Norway, *Picea abies* (L.) Karst. and Siberian, *P. obovata* Ledeb., spruce species studied by isozyme loci. *Heredity*, **74**, 464–480.

LAGERCRANTZ, U. AND RYMAN, N. 1990. Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, **44**, 38–53.

LEHNER, A., CAMPBELL, M. A., WHEELER, N. C., POYKKO, T., GLOSSL, J., KREIKE, J. AND NEALE, D. B. 1995. Identification of a RAPD marker linked to the pendula gene in Norway spruce (*Picea abies* (L.) Karst. f. *pendula*). *Theor. Appl. Genet*., **91**, 1092–1094.

MÜLLER-STARCK, G. 1995. Genetic variation in high elevated populations of Norway spruce (*Picea abies* (L.) Karst.) in Switzerland. *Silvae Genet.*, **44**, 356–362.

MUONA, O., PAULE, L., SZMIDT, A. E. AND KÄRKKÄINEN, K. 1990. Mating system analysis in a central and northern European population of *Picea abies*. *Scand. J. For. Res.*, **5**, 97–102.

PAGLIA, G. AND MORGANTE, M. 1998. PCR-based multiplex DNA fingerprinting techniques for the analysis of conifer genomes. *Mol. Breed.*, **4**, 173–177.

PAGLIA, G. P., OLIVIERI, A. M. AND MORGANTE, M. 1998. Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.). *Mol. Gen. Genet.*, **258**, 466–478.

PERRY, D. J. AND BOUSQUET, J. 1998a. Sequence-tagged-site (STS) markers of arbitrary genes: development, characterization and analysis of linkage in black spruce. *Genetics*, **149**, 1089–1098.

PERRY, D. J. AND BOUSQUET, J. 1998b. Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived primers in other conifers. *Theor. Appl. Genet.*, **97**, 735–743.

PFEIFFER, A., OLIVIERI, A. M. AND MORGANTE, M. 1997. Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome*, **40**, 411–419.

SCHEEPERS, D., ELOY, M.-C. AND BRIQUET, M. 1997. Use of RAPD patterns for clone verification and in studying provenance relationships in Norway spruce (*Picea abies*). *Theor. Appl. Genet.*, **94**, 480–485.

SCOTTI, I., TROGGIO, M., SORANZO, N., VENDRAMIN, G. G. AND BUCCI, G. 1998. A new set of PCR-based, locus-specific markers for *Picea abies* (L.) Karst. *Mol. Ecol.*, **7**, 789–792.

SIGURGEIRSSON, A. AND SZMIDT, A. E. 1993. Phylogenetic and biogeographic implications of chloroplast DNA variation in *Picea*. *Nord. J. Bot.*, **13**, 233–246.

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

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

WEBER, J. L. AND MAY, P. E. 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 MCCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.*, **18**, 7213–7218.

WILLIAMS, J. G., KUBELIK, A. R., LIVAK, K. J., RAFALSKI, J. A. AND TINGEY, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.*, **18**, 6531–6435.

WRIGHT, J. W. 1955. Species crossability in spruce in relation to distribution and taxonomy. *Forest Sci.*, **1**, 319–349.