Theor Appl Genet (2004) 109: 1283–1294
DOI 10.1007/s00122-004-1742-5

ORIGINAL PAPER

Dainis Rungis · Yanik Bérubé · Jun Zhang · Steven Ralph · Carol E. Ritland · Brian E. Ellis · Carl Douglas · Jörg Bohlmann · Kermit Ritland

Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags

Received: 5 December 2003 / Accepted: 17 May 2004 / Published online: 3 September 2004

Abstract Traditionally, simple sequence repeat (SSR) markers have been developed from libraries of genomic DNA. However, the large, repetitive nature of conifer genomes makes development of robust, single-copy SSR markers from genomic DNA difficult. Expressed sequence tags (ESTs), or sequences of messenger RNA, offer the opportunity to exploit single, low-copy, conserved sequence motifs for SSR development. From a 20,275-unigene spruce EST set, we identified 44 candidate EST-SSR markers. Of these, 25 amplified and were polymorphic in white, Sitka, and black spruce; 20 amplified in all 23 spruce species tested; the remaining five amplified in all except one species. In addition, 101 previously described spruce SSRs (mostly developed from genomic DNA), were tested. Of these, 17 amplified across white, Sitka, and black spruce. The 25 EST-SSRs had approximately 9% less heterozygosity than the 17 genomic-derived SSRs (mean H=0.65 vs 0.72), but appeared to have less null alleles, as evidenced by much lower apparent inbreeding (mean F=0.046 vs 0.126). These robust SSRs are of particular use in comparative studies, and as the EST-SSRs are within the expressed portion of the genome, they are more likely to be associated with a particular gene of interest, improving their utility for quantitative trait loci mapping and allowing detection of selective sweeps at specific genes.

Communicated by O. Savolainen

D. Rungis · Y. Bérubé · J. Zhang · S. Ralph · C. E. Ritland · J. Bohlmann · K. Ritland (✉)
Department of Forest Sciences, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada
e-mail: kermit.ritland@ubc.ca
Fax: +1-604-8229102

B. E. Ellis · C. Douglas · J. Bohlmann
Department of Botany, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

B. E. Ellis · J. Bohlmann
Biotechnology Laboratory, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

Introduction

Molecular genetic markers are variable regions of DNA that provide valuable genetic tools in genetic linkage mapping, association studies, phylogeographic studies, and for the estimation of several population genetic parameters, such as diversity, gene flow, and inbreeding (Bruford and Wayne 1993). To date, the molecular markers most widely applied to tree species have been isozymes, random fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphisms (AFLPs), and simple sequence repeats (SSRs). Each marker technique has attributes that offer different advantages (Ritland and Ritland 2000). Isozymes are well studied and established, but are not numerous or highly polymorphic. RFLPs utilize probes derived from either genomic or coding DNA (cDNA) and are codominant markers, but require a large amount of high quality DNA. RAPDs and AFLPs do not require any sequence knowledge of the genome, and so are easy to apply to uncharacterized genomes. However, they are usually dominant and are often difficult to transfer between different mapping populations or species. Additionally, RAPDs are notoriously difficult to transfer across laboratories (Jones et al. 1997).

SSR markers exhibit codominance and are usually highly polymorphic, and thus, seem to be the ideal marker (Ritland and Ritland 2000). However, their development requires a significant investment, and their cross-species transferability is normally quite limited due to either disappearance of the repeat region, or to degeneration of the primer binding sites. Traditionally, the first stage of SSR marker development is to probe a genomic library with a particular SSR oligonucleotide and sequence positive clones. However, the success rate of identifying robust SSR markers from genomic DNA is typically low due to the high proportion of primers that do not amplify genetically interpretable PCR fragments (Squirrell et al. 2003).

In conifers, SSR discovery from genomic libraries (van de Ven and McNicol 1996; Pfeiffer et al. 1997; Rajora et

al. 2000; Hodgetts et al. 2001; Scotti et al. 2002a, b) has been a particularly difficult process, with very low success rate, probably because of the large, repetitive nature of their genomes (Pfeiffer et al. 1997; Bérubé et al. 2003). Despite these problems, SSRs remain the marker system of choice for a number of conifer mapping projects (Paglia and Morgante 1998).

Expressed sequence tags (ESTs) are sequenced portions of messenger RNA and offer an alternative route for SSR marker discovery, particularly for the repetitive genomes found in conifers. The advent of large-scale databases with tens of thousands of ESTs provides resources for the novel, "in silico" identification of genetic markers. In marker development, EST databases have largely been used for identification of single nucleotide polymorphisms (SNPs) (Rafalski 2002). However, SSRs are found in both the untranslated regions of ESTs and occasionally within coding regions (Cardle et al. 2000).

One advantage of these "EST-SSRs" is that they are directly associated with a coding gene, and so may be useful for association with phenotypic traits. Also, because EST sequences are evolutionary conserved, cross-species PCR amplification of EST-SSRs are expected to be more successful compared to SSRs developed from genomic DNA (Arnold et al. 2002; Saha et al. 2003); however, their levels of variability may not be as great due to selective constraints. Finally, with their relatively high levels of variability, EST-SSRs seem especially appropriate for the detection of selective sweeps (Vigouroux et al. 2002).

Here, we utilize an EST database—developed as part of the Genome British Columbia Forestry Genomics project—to identify and characterize SSR markers for spruce. This database provides a valuable and unique resource for the development of new SSR markers within spruce and also for comparative analysis of genome structure and organization. We report 25 new EST–SSR markers of primary use with white, Sitka, and black spruce. We also evaluate 101 previously reported spruce SSRs (derived from genomic DNA libraries), evaluate the use of all SSRs across 23 spruce species, and arrive at a total set of 42 robust microsatellites markers for spruce.

Materials and methods

Library construction and DNA sequencing

Nine directional cDNA libraries were constructed from a range of tissues (xylem, phloem, bark, foliage, and roots) at different developmental stages of seedlings and mature trees, as well as from trees or seedlings exposed to chemical elicitors (methyl jasmonate), or mechanical wounding. Tissues were obtained from three different spruce species: white spruce (*Picea glauca*) cultivar PG29, Sitka spruce (*Picea sitchensis*) cultivar Gb2-229, and the interior spruce (*P. glauca* × *Picea engelmannii*) cultivar Fal-1028. cDNA libraries were constructed (5′ *Eco*RI, 3′ *Xho*I) using the pBluescript II XR cDNA Library Construction Kit, following manufacturer's instructions with modifications (Stratagene). Select cDNA libraries were normalized according to the Soares method (Soares et al. 1994). A complete technical description of library construction methods will be reported elsewhere.

Library-stock plasmid DNA was transformed into electrocompetent DH10B T1-phage-resistant *Escherichia coli* cells (Invitrogen) and robotically arrayed into 384-well plates from which glycerol stocks were prepared. Plasmid DNA was extracted from overnight 96-well cultures and BigDye Terminator (ABI) cycle sequenced on an ABI Prism 3700 DNA Analyzer, using conventional procedures and the −21 M13 forward primer (5′-TGTAAAACGACGGCCAGT-3′) to obtain predominantly 3′ end sequences. DNA sequence chromatograms were processed using the PHRED software (Ewing and Green 1998; Ewing et al. 1998). Sequences were quality trimmed according to the high-quality contiguous region determined by PHRED and then vector trimmed using CROSS_MATCH software (http://www.phrap.org). Sequences with less than 70 quality bases after trimming were discarded.

EST database and SSR search

The EST database used for this search consisted of 34,846 EST sequences, which were quality clipped using PHRED and our own in-house software "EST Clean." This step also removed poly-A tail sequence from the ESTs. The clipped sequences were aligned to generate 20,275 unigenes, using the CAP-3 software package (Huang and Madan 1999). We developed an EST–SSR discovery software package (BuildSSR, available at http://www.genetics.forestry.ubc.ca/ritland/programs.html) to search for SSRs within this unigene set. This included database organization, repeat-finding software, and tools for SSR-distribution analysis. This SSR-discovery pipeline identifies SSRs in the unigene set, constructs a summary table, and then, builds a FASTA-format database that includes the repeat type, size, and position. A minimum perfect-repeat number of nine dinucleotide repeats, six trinucleotide repeats, and four tetranucleotide repeats was used for the search. EST containing SSRs were then annotated using BLAST software.

Primer design and PCR conditions

Primers spanning 44 EST-SSRs were designed. These repeats were detected in sequences from the interior, white, and Sitka spruce libraries. Primers were designed using the Primer 3 software (Rozen and Skaletsky 2000). Regions 50 bp from each end of the repeat were excluded from primer site consideration, and all primers were designed to have similar annealing temperatures to allow for uniform PCR cycling conditions. The forward primer was tailed with an M13 sequence (Oetting et al. 1995) to facilitate visualization of PCR products on a LiCor 4200 (LiCor, Lincoln, Neb., USA).

The 101 previously described spruce SSR primers derived from the genomic DNA approach were also synthesized (Pfeiffer et al. 1997; Rajora et al. 2000; Scotti et al. 2000; Hodgetts et al. 2001; Scotti et al. 2002a, b; C. Newton, personal communication), with the forward primer tailed with an M13 sequence as above.

PCR was performed with 25 ng genomic DNA, 0.2 μM of each forward and reverse primer, 0.05 μM M13 IRD labeled primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, and 1 U of AmpliTaq DNA polymerase (Roche) in a 20-μl volume. PCR cycling conditions consisted of an initial denaturation step of 95°C for 2 min; 30 cycles of 95°C for 20 s, 53°C for 20 s, and 72°C for 30 s; followed by a final extension step of 72°C for 3 min.

Plant material

Fresh needle tissue from the current year's growth was collected from 20 mature trees in wild populations of both white and Sitka spruce. The white spruce population was sampled in the region surrounding the town of Fort Nelson, situated in the northeast corner of the province of British Columbia, Canada. Trees of this population were sampled 1–2 km apart. The Sitka spruce population was located on Kodiak Island, Alaska, which marks the northern migrating tip of the species' range. Trees of this population were sampled 30–50 m apart. DNA from 20 black spruce individuals was obtained from samples collected in Manitoba and Saskatchewan, Canada. Buds were collected from a single tree from each of 23 spruce species (for list see Table 3), which were growing as a collection at the British Columbia Ministry of Forests Kalamalka research station, Vernon, B.C. Only one genetic individual was available for each species in this collection. DNA was isolated from the bud and needle tissue following the CTAB method described by Doyle and Doyle (1990).

SSR testing and assay

The 44 SSR developed from the EST database were tested on the above described population collections of Sitka, white, and black spruce. In addition, the 101 previously described SSR primer pairs were tested on a panel of two white, two Sitka, one black, and one red spruce individual. The SSR primer pairs that amplified products from these species were then tested on the collection of 23 spruce species.

In the testing and assay, presence or absence of microsatellite PCR products was scored on 2% agarose gels. When products were found, they were tested for polymorphism on 6% (Long Ranger) polyacrylamide gels, using a LiCor 4200 automated sequencer. Microsatellite products were detected by M13 tailed primer (Oetting et al. 1995).

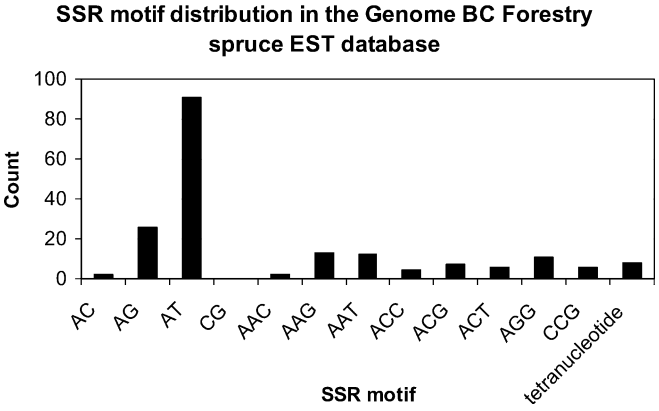

Fig. 1 Distribution of simple sequence repeat motifs in the Genome British Columbia Forestry spruce EST database

Analyses

Observed heterozygosity (H_o), expected heterozygosity (H_e) and the inbreeding coefficient (F: $F=1-H_o/H_e$) were estimated for each SSR locus within each of the three spruce species (Sitka, white, black). Standard errors of F were determined by bootstrapping individuals within populations, using a Fortran 95 program written by K. Ritland.

Genetic distances between individuals in the 23 species set were estimated as the mean squared difference of allele sizes (Goldstein et al. 1995)—after sizes were normalized by dividing by the variance of allele size (specific for each locus)—using a Fortran 95 program written by K. Ritland. A total of 100 bootstrap datasets were constructed by resampling loci. For each replicate, the computer program NEIGHBOR (in PHYLIP, Phylogeny Inference Package, version 3.57c, Felsenstein 1995) was used to construct an unrooted tree, using the neighbor-joining (NJ) method (Saitou and Nei 1987). The 100 trees were then evaluated by CONSENSE (in PHYLIP) to find an overall consensus tree, with confidence numbers attached to each branch.

The metric of mean squared allele size difference outperforms heterozygosity at differences over larger time periods [≥1,000 generations, particularly when standardized (Neff 2004)]. Hence, instead of standard measures such as Nei's genetic distance or the proportion of bands not shared, we used a mean squared allele size difference, standardized by mutation rate (as the variance of allele size is proportional to the mutation rate, c.f. Goldstein et al. 1995).

Results

From the Genome British Columbia (BC) spruce EST unigene database, 188 unique SSR sequences were found within 183 contigs. A total of 119 dinucleotide, 61 trinucleotide, and eight tetranucleotide repeats were found (Fig. 1). The most common class of repeat was AT (91 of 188 SSRs). Of the SSR sequences found in the EST database, 31 were at the extreme 3′ end of the ESTs

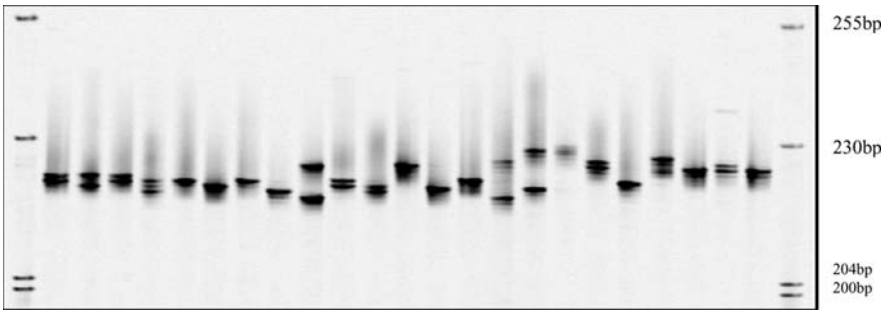

Fig. 2 Amplification of locus WS0092.A19 across the 23 spruce species

(adjoining the poly-A tail) and 22 were at the 5′ end of the sequences; therefore, primers could not be designed for these sequences. The distribution of the SSR repeat types in relation to the coding sequence was non-random. Of the 31 SSRs at the 3′ end of the ESTs, 30 were AT repeats (the remaining SSR was an ATT repeat). Of the 22 repeat types at the 5′ end of the ESTs, 19 were AG repeats (the remaining SSRs were one each of AT, GAC and AGA). Only two AC repeats were identified within the EST database. G+C content within the SSR containing ESTs was 40.2%, which is comparable to the G+C content in and *Arabidopsis* SSR containing ESTs (43.8%) (Morgante et al. 2002).

Of the 145 primer pairs, 41 detected a single locus, and one previously developed SSR detected two loci across these four species (Table 1). This set of 42 primer pairs included 25 EST-SSRs and 17 previously developed SSRs.

H_eH_o heterozygosity, and F are shown in Table 2 (in some cases there were insufficient individuals to obtain adequate estimates of F). As is normal with microsatellites, heterozygosity varied widely among loci. The average heterozygosity was highest in white spruce (0.78), lower in black spruce (0.72), and lowest in Sitka spruce (0.55).

The EST-SSRs showed significantly less variation than the genomic-derived SSRs; H_e values were 6.25% less in white spruce, 15% less in black spruce, and 9% less in Sitka spruce. Likewise, the numbers of alleles at ESR-SSR loci were comparably lower in all three species. Interestingly, F values were significantly lower at the EST-SSR compared to the genomic-derived loci (0.02 vs 0.13 in Sitka, 0.03 vs 0.10 in white, and 0.09 vs 0.14 in black).

Of the 43 loci amplified by the 42 primer pairs determined to be informative across white, Sitka, black, and red spruce, 33 were identified in all 23 spruce species (Fig. 2; Table 3). The minimum number of species in which a particular locus was present was 17. Twenty-five EST–SSRs primer pairs developed from the Genome BC spruce EST database were included in this set of markers. Of these, 20 amplified single locus markers from across all 23 spruce species tested, while five amplified single-locus markers from 22 of the 23 spruce species tested.

Figure 3 gives the NJ tree of microsatellite genetic distances among the 23 spruce species, and Fig. 4 gives the consensus tree of microsatellite genetic distances among 23 spruce species. The relatively deep rooting of each species is due to the variability and high evolutionary rate at SSR loci. While some clustering of related species is evident, bootstrap confidence levels are not high.

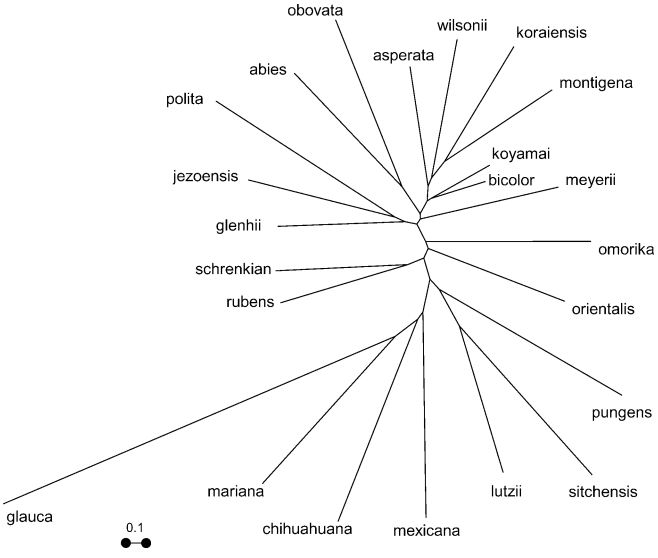

Fig. 3 Neighbor-joining tree of microsatellite genetic distances among the 23 spruce species

Discussion

The EST–SSR markers are adjacent to coding genes, and the function of these genes can be often identified via sequence similarity to annotated genes in other plant species. Thus, they are useful in quantitative trait locus mapping and particularly “genomic scans” (Vigouroux et al. 2002). Their association with coding genes makes EST-SSRs more likely to be single copy, which is particularly useful for species with large genomes such as spruce. Furthermore, as coding regions tend to be more conserved, this potentially increases the transferability of these EST-SSRs across spruce species. While the EST-derived SSR markers in this study were somewhat less variable than the genomic SSR markers, the F values were also significantly lower, suggesting a lower frequency of troublesome null alleles in EST-SSRs.

Table 1 Primer pair sequences and repeat motifs. Forward primers were 5′ tailed with the M13 sequence 5′-CACGACGTTGTAAAAC-GAC- 3′ to facilitate visualization on LiCor sequencers

Locus	Forward primer	Reverse primer	Repeat motif	Accession no.	Best-matched protein (blastx)
WS0011.P12[a]	cgataagatggctcctcaaa	ggaggctgaaaagtggttaca	$(AGGA)_{32}$	CN480892	AAF86307—EF-hand Ca2+—binding protein CCD1 (*Triticum aestivum*)
WS0015.I04[a]	caccctttaaccaagcaagc	ggtctacatgtttatcaccaacga	$(AT)_{29}$	CN480893	NP_565298—putative chloroplast nucleoid DNA binding protein (*Arabidopsis thaliana*)
WS0016.O09[a]	ctttgggggctagcaagttt	attcgggcttcatagcacaa	$(AT)_9$	CN480894	NP 197764—expressed protein (*A. thaliana*)
WS0019.M09[a]	tttcaaatcggagtgcattg	ggagatcgtggtaacccaaa	$(AT)_{20}$	CN480895	NP_700730—hypothetical protein (*Plasmodium falciparum* 3D7)
WS0019.F22[a]	aagcgtttctcattttcttgg	gggcccagaactaacaatga	$(AT)_{13}$	CN480896	Unknown
WS0022.B15[a]	tttgtaggtgctgcagagatg	tggctttttattccagcaaga	$(AG)_{12}$	CN480899	AAM61048—unknown (*A. thaliana*)
WS0023.B03[a]	agcagctggggtcaaagtt	aaagaaagcatgcatatgactcag	$(AT)_{10}$	CN480900	NP_179617—putative glutaredoxin (*A. thaliana*)
WS0023.B12[a]	gatgagtggaattgggagaga	aaagtcaatttttcatggcttca	$(TA)_{22}$	CN480901	AAL99613—mitochondrial aldehyde dehydrogenase RF2B (*Zea mays*)
WS0032.M17[a]	gcttgacacctgaaaattacattag	aaggcaagagggatcgtaaa	$(ATT)_6$	CN480906	CAD36515—putative beta-glycosidase (*Oryza sativa* [japonica cultivar-group])
WS0033.A18[a]	ggctgctctcttatccgtttt	tggctctcatccagaaaagaa	$(TA)_{26}$	CN480907	NP_565298—putative chloroplast nucleoid DNA binding protein
WS0035.A01[a]	gggcgaaatatgtcgatttt	tcatccctgcattgtctcg	$(AT)_{11}$	CN480908	NP_197051—glycosyltransferase family 8 (*A. thaliana*)
WS0046.M11[a]	cactagggcattgggaagaa	atgagaggctggggtatgaa	$(AAG)_6$	CN480891	A45612—H+—transporting ATP synthase protein 6 homolog—*Leishmania tarentolae* mitochondrion
WS0053.K16[a]	acatatcatggttgcgatgc	ccacagcccctaaaatgtga	$(AT)_{13}$	CN480898	CAD66637—phytocyanin protein, PUP2 (*A. thaliana*)
WS0061.C21[a]	tttttagcctcatggacgtt	ggttaaacggacgctgaaag	$(CTTT)_5$	CN480886	AAL55635—hexokinase-related protein 1 (*Solanum tuberosum*)
WS0061.K02[a]	tcaagaatcagctccgcttt	ggcgcagatacgttgacat	$(AT)_9$	CN480887	AAF35901—expansin 2 (*Zinnia elegans*)
WS0071.J15[a]	tttttaaccatgggaattgg	ggatcgaagggatgtcaaga	$(AT)_{22}$	CN480902	AAF61443—root border cell-specific protein (*Pisum sativum*)
WS0073.H08[a]	tgctctcttattcgggcttc	aagaacaaggcttcccaatg	$(AT)_{14}$	CN480903	AAD28506—remorin 1 (*Lycopersicon esculentum*)
WS0079.H08[a]	gggatgcctgggtaaataaaa	ttttgcatttgctttgatatgtg	$(GCAG)_6$	CN480904	Unknown
WS0082.E23[a]	caggtcaaatccttccttcc	gaagaaaatgctggctttcg	$(TA)_{11}$	CN480909	AAM28914—TIR/P-loop/LRR (*Pinus taeda*)
WS0082.O23[a]	agtgacagttgtcttagcacatca	aaggtttccgatcgcatcta	$(TA)_{15}$	CN480910	T09251—embryonic abundant protein EMB24 (*P. glauca*)
WS0092.A19[a]	tgtggttttctgcttggaaa	cccattttgactttgaaataagc	$(AC)_9$	CN480888	AAA56991—formerly called HAT24; synaptobrevin-related protein (*A. thaliana*)
WS0092.M15[a]	gatgttgcaggcattcagag	gcaccagcatcgattgacta	$(TCC)_6$	CN480889	NP_192787—oxidoreductase, 2OG-Fe(II) oxygenase family (*A. thaliana*)
WS0092.H13[a]	ccacgatgtcgttgaaagaa	tttcagtcttcctgcattcg	$(GCT)_8$	CN480890	Unknown
WS00111.K13[a]	gactgaagatgccgatatgc	ggccatatcatctcaaaataaagaa	$(AT)_9$	CN480897	BAB86071—putative beta-glucosidase (*O. sativa* [japonica cultivar-group])
WS00716.F13[a]	tcaagtaatggacaaacgataca	tttccaatagaatggtggattt	$(GA)_{10}$	CN480905	NP_195313—expressed protein (*A. thaliana*)
PAAC17[b]	gaaacaaaaattattacgcg	atgccctcctaatgaatg	$(AC)_{36}$	AJ131107	NA[i]
PAAC19[b]	atgggctcaaggatgaatg	aactccaaacgattgatttcc	$(CT)_{23}CAA(TG)_{12}$	AJ131108	NA
PAAC23[b]	tgtggccccacttactaatatcag	cgggcattggtttacaagagttgc	$(GT)_{14}$	AJ131109	NA
PGL14[c]	aaaaatgatttatatcttcttattgtct	gngtcataaacgcccatcaatag	$(AG)_{20}$	NA	NA
UAPgAG105[d]	caactaccttgagccaatca	gtccggcattattgatcatt	$(AG)_{11}$	NA	NA
UAPgAG150A[d]	accaatgcttttaccaaacg	ttgattgcaagtgatggttg	$(AG)_{19}$	NA	NA

Table 1 (continued)

Locus	Forward primer	Reverse primer	Repeat motif	Accession no.	Best-matched protein (blastx)
UAPgAG150B[d]	as above	as above	as above	NA	NA
SPAGC1[e]	ttcaccttagccgagaacc	cactggagatcttcgttctga	$(TC)_5TT(TC)_{10}$	NA	NA
SPAGG3[e]	ctccaacattcccatgtagc	agcatgttgtcccatatagacc	$(GA)_{24}$	NA	NA
SPL3AG1A4[e]	catactcaatgcacctagatatgc	aagcaaatgaaagctccttgt	$(GA)_{21}$	NA	NA
SPL3AG1H4[e]	ggaaaggaggaggacaagag	taaggatcgagtctctcactcc	$(GA)_{20}$	NA	NA
EAC6A06[f]	aattaaggggtaatgtgccac	aatgatgttaaagcaatatgtcttg	$(AC)_{20}$	AJ292706	NA
EAC6B03[f]	gaaggttataatattcagtgaagg	taatgcttatcaatgaggttg	$(AC)_{25}$	AJ292712	NA
EAC7C11[f]	aactctataaaataacgcacctcg	ccaaaacaaaggaaggatgtt	$(AC)_{19}$	AJ292730	NA
EAC7H07[f]	ggttcaaacctcccacctac	accaactaagccacaagtgc	$(CA)_{23}(CAT)_{10}$	AJ292739	NA
EATC3C05[g]	ttagtggacgttcatcatcatc	tcacaatcacttttttagtcgc	$TAT(CAT)_{10}(AT)_{20}$	AJ296736	NA
2[h]	tttggactctttttaatgagattg	acagacaatgtgacaatatagtg	$(TC)_{25}$	NA	NA
44[h]	ttacacttcagagagagagaga	ggcccacatcaacccttacc	$(AG)_n$	NA	NA

[a]Genome British Columbia expressed sequence tag-simple sequence repeats (EST-SSRs)
[b]Scotti et al. (2000)
[c]Rajora et al. (2000)
[d]Hodgetts et al. (2001)
[e]Pfeiffer et al. (1997)
[f]Scotti et al. (2002a)
[g]Scotti et al. (2002b)
[h]Craig Newton (personal communication)
[i]*NA* Not available

SSR locations in spruce ESTs

The SSRs exhibited differential distribution within the expressed sequences. AT repeats were preferentially found at the 3′ end of the EST sequences, while AG repeats were preferentially found at the 5′ end of sequences. While Scotti et al. (2000) found six AC repeat regions from a Norway spruce cDNA library clustered at the 3′ end of the expressed sequences, we found only two AC repeats within our 3′ EST collection. This may reflect a difference in SSR composition between Norway spruce and the North American spruces used for our cDNA libraries. Alternatively, by specifically targeting AC repeats, Scotti et al. (2000) may have identified the rare AC repeats found in expressed portions of the spruce genome.

SSR motif types in spruce ESTs

The most common SSR motif found in our EST database was AT, accounting for 91 of the 188 repeats identified. By contrast, in *Arabidopsis* ESTs, AAG is the most common class of SSR, and AT repeats are less prevalent (Cardle et al. 2000). AT repeats, however, are the second most common SSR type (after poly A repeats) in *Arabidopsis* and other plant genomic DNA (Cardle et al. 2000). The prevalence of AT repeats in spruce ESTs is also supported by SSR searches of *Picea* sequences in the EMBL database, where of seven SSRs developed, four were AT repeats (Besnard et al. 2003). This prevalence of AT repeats in spruce ESTs may be a hitherto unnoted feature, as other studies where SSRs have been isolated from spruce coding sequences have utilized specific repeat probes (not AT) (e.g., Scotti et al. 2000). Alternatively, as AT repeats were found to be preferentially clustered at the 3′ end of the ESTs, this preponderance of AT repeats may be a consequence of the 3′ sequencing of this EST database. We are currently in the process of obtaining full-length EST sequences, and a survey of these may reveal a different distribution of SSR repeats.

SSR polymorphism

The amplification of the SSRs in Sitka, white, and black spruce populations revealed high levels of polymorphism, as indicated by the high average number of alleles and the high H_e and H_o, both typical of SSR markers. This suggests that most of these SSR markers will be useful in parentage and clonal assessments because of their high potential for discrimination. They will also be useful in constructing genetic linkage maps, as these markers will likely be segregating in a range of crosses.

Table 2 Polymorphism in Sitka, white, and black spruce as described by the number of alleles, the expected (H_e) and observed (H_o) heterozygosities, the estimated inbreeding coefficient (F), and its estimated standard error (SE)

Locus	*Picea sitchensis*						*P. glauca*						*P. mariana*					
	n	No. of alleles	Allele size range (bp)	H_e	H_o	F (SE)	n	No. of alleles	Allele size range (bp)	H_e	$H_{eo/Subscript>}$	F (SE)	n	No. of alleles	Allele size range (bp)	H_e	H_o	F (SE)
WS0011.P12	19	3	291–295	0.15	0.16	−0.07 (0.25)	20	4	283–295	0.67	0.80	−0.20 (0.15)	20	2	289–293	0.14	0.15	−0.08 (0.12)
WS0015.I04	20	8	179–229	0.73	0.90	−0.24 (0.12)*	20	17	181–253	0.92	0.94	−0.03 (0.05)	20	14	185–225	0.87	0.80	0.05 (0.09)
WS0016.O09	16	9	390–406	0.84	0.56	0.34 (0.15)*	19	11	396–424	0.81	0.53	0.36 (0.15)*	20	5	394–402	0.49	0.50	−0.06 (0.12)
WS0019.M09	5	3	236–312	0.62	0.00	–	18	23	208–272	0.97	0.83	0.15 (0.09)	20	20	210–276	0.96	0.80	0.15 (0.08)
WS0019.F22	19	6	352–366	0.70	0.58	0.18 (0.14)	20	22	358–428	0.97	0.90	0.07 (0.07)	18	20	370–424	0.96	0.94	−0.01 (0.05)
WS0022.B15	19	4	183–203	0.57	0.47	0.18 (0.17)	20	13	169–207	0.92	0.85	0.07 (0.08)	20	10	169–211	0.71	0.50	0.28 (0.19)
WS0023.B03	16	9	174–218	0.79	0.69	0.14 (0.10)	16	14	170–216	0.89	0.94	−0.06 (0.07)	16	7	174–194	0.74	0.81	−0.13 (0.13)
WS0023.B12	19	7	160–188	0.58	0.42	0.28 (0.16)	20	13	160–230	0.90	0.70	0.23 (0.12)	19	22	194–254	0.96	0.79	0.16 (0.09)
WS0032.M17	18	8	278–308	0.58	0.72	−0.24 (0.10)*	19	8	287–323	0.73	0.68	0.06 (0.13)	20	7	281–302	0.66	0.60	0.07 (0.12)
WS0033.A18	18	3	145–149	0.37	0.39	−0.06 (0.13)	19	5	143–151	0.78	0.74	0.05 (0.14)	19	4	153–161	0.63	0.42	0.31 (0.16)
WS0035.A01	18	2	148–150	0.36	0.44	−0.26 (0.10)*	20	5	144–152	0.77	0.70	0.10 (0.13)	20	7	144–158	0.83	0.40	0.51 (0.13)*
WS0046.M11	16	1	287–287	0.00	1.00	–	20	3	287–293	0.66	0.90	−0.37 (0.10)*	20	6	287–353	0.83	0.70	0.13 (0.12)
WS0053.K16	19	5	201–217	0.66	0.68	−0.04 (0.17)	20	4	193–225	0.50	0.35	0.31 (0.21)	20	8	201–229	0.78	0.65	0.14 (0.13)
WS0061.C21	16	3	259–279	0.28	0.31	−0.14 (0.20)	14	3	259–275	0.59	0.50	0.16 (0.22)	–	–	–	–	–	–
WS0061.K02	20	2	209–217	0.26	0.30	−0.15 (0.07)*	19	5	207–223	0.52	0.63	−0.22 (0.11)	20	2	209–213	0.05	0.05	−0.03 (0.12)
WS0071.J15	18	11	205–247	0.89	1.00	−0.13 (0.04)*	18	14	203–243	0.93	0.94	−0.02 (0.06)	20	15	197–241	0.91	0.60	0.33 (0.13)*
WS0073.H08	18	8	188–218	0.74	0.61	0.18 (0.14)	20	12	202–236	0.84	0.75	0.11 (0.10)	20	5	206–214	0.45	0.45	−0.02 (0.13)
WS0079.H08	15	2	252–256	0.23	0.13	0.42 (0.30)	17	4	248–260	0.50	0.47	0.05 (0.18)	19	6	236–260	0.78	0.79	−0.04 (0.12)
WS0082.E23	5	4	239–247	0.73	1.00	–	15	10	219–245	0.87	0.73	0.17 (0.13)	20	9	195–237	0.77	0.75	0.00 (0.14)
WS0082.O23	20	6	214–224	0.61	0.80	−0.31 (0.10)*	20	3	210–222	0.49	0.65	−0.34 (0.10)*	20	5	212–224	0.68	0.70	−0.05 (0.12)
WS0092.A19	20	5	215–223	0.69	0.65	0.06 (0.14)	20	4	213–221	0.67	0.60	0.11 (0.17)	20	4	215–221	0.67	0.80	−0.22 (0.13)
WS0092.M15	20	3	212–218	0.10	0.10	−0.01 (0.14)	20	1	212–212	–	–	–	20	2	212–215	0.43	0.20	0.52 (0.23)*
WS0092.H13	17	3	220–226	0.53	0.65	−0.24 (0.23)	20	2	223–226	0.36	0.35	0.02 (0.23)	17	2	220–223	0.11	0.12	−0.06 (0.22)
WS00111.K13	18	3	215–225	0.37	0.28	0.25 (0.24)	16	20	215–265	0.96	0.94	0.03 (0.06)	20	10	214–246	0.78	0.70	0.08 (0.12)
WS00716.F13	16	7	281–307	0.79	0.81	−0.03 (0.14)	17	13	279–315	0.87	0.94	−0.08 (0.07)	20	14	281–319	0.90	0.75	0.15 (0.13)
2	20	9	172–216	0.77	0.65	0.16 (0.10)	17	21	176–226	0.97	0.71	0.28 (0.13)*	19	16	176–230	0.93	0.79	0.13 (0.11)
44	20	6	107–129	0.58	0.65	−0.12 (0.14)	20	8	105–145	0.71	0.40	0.44 (0.12)*	20	17	109–159	0.94	0.65	0.29 (0.11)*
EAC6A06	19	8	95–141	0.71	0.37	0.49 (0.16)*	15	15	83–135	0.94	0.67	0.30 (0.12)*	19	11	99–123	0.87	0.79	0.07 (0.11)
EAC6B03	–	–	–	–	–	–	20	9	87–173	0.61	0.45	0.26 (0.20)	20	14	93–131	0.93	0.70	0.23 (0.10)
EAC7C11	19	11	105–137	0.90	1.00	−0.11 (0.02)*	17	17	101–149	0.89	0.88	0.01 (0.09)	18	11	105–143	0.84	0.61	0.25 (0.12)*
EAC7H07	10	5	125–139	0.78	0.70	0.11 (0.18)	20	12	121–149	0.87	0.75	0.14 (0.10)	18	7	99–113	0.72	0.44	0.37 (0.14)
EATC3C05	13	6	249–267	0.80	0.31	0.63 (0.19)*	19	9	231–261	0.80	0.68	0.15 (0.12)	18	20	237–361	0.96	0.72	0.22 (0.12)
PAAC17	19	5	132–148	0.33	0.16	0.53	20	10	136–168	0.81	0.55	0.33 (0.13)*	19	8	136–152	0.80	0.68	0.12 (0.15)
PAAC19	18	6	155–173	0.72	0.39	0.47 (0.15)*	20	13	155–199	0.86	0.68	0.21 (0.10)*	19	8	155–195	0.68	0.42	0.37 (0.16)*
PAAC23	20	3	266–276	0.42	0.45	−0.09 (0.10)	20	7	264–280	0.50	0.60	−0.20 (0.06)*	17	3	276–280	0.27	0.18	0.33 (0.14)
PGL14	16	10	132–164	0.84	0.75	0.11 (0.13)	19	17	130–170	0.94	0.95	−0.01 (0.06)	19	14	135–171	0.83	0.89	−0.11 (0.08)
SPAGC1	19	7	104–150	0.63	0.84	−0.35 (0.06)*	19	10	104–156	0.70	0.84	−0.20 (0.06)*	20	13	104–136	0.86	0.80	0.05 (0.09)
SPAGG3	19	8	105–137	0.73	0.74	−0.01 (0.14)	15	11	117–141	0.89	0.87	0.02 (0.11)	20	14	115–143	0.87	0.80	0.05 (0.09)
SPL3AG1A4	15	2	85–105	0.52	1.00	0.02 (0.10)	14	16	77–129	0.96	0.86	0.11 (0.10)	–	–	–	–	–	–
SPL3AG1H4	9	5	127–161	0.71	0.22	–	17	14	131–161	0.93	0.76	0.19 (0.12)	20	5	111–127	0.56	0.55	−0.01 (0.11)
UAPgAG105	20	3	153–163	0.14	0.15	−0.04 (0.15)	20	9	153–187	0.83	0.80	0.03 (0.10)	20	6	157–177	0.77	0.85	−0.13 (0.09)
UAPgAG150A	17	2	153–157	0.30	0.12	0.61 (0.30)*	18	8	141–161	0.84	0.33	0.61 (0.15)*	19	11	139–163	0.90	0.53	0.40 (0.13)*
UAPgAG150B	19	2	126–130	0.05	0.05	0.00 (0.20)	15	2	126–130	0.40	0.53	−0.33 (0.13)*	20	4	124–132	0.59	0.65	−0.14 (0.16)
Arithmetic mean	17.1	5.3		0.55	0.53	0.06 (0.03)	18.4	10.3		0.78	0.71	0.05 (0.02)*	19.3	9.5		0.72	0.61	0.11 (0.02)*
EST-SSRs	17.00	5.00		0.53	0.55	0.02 (0.04)	18.7	9.32		0.75	0.72	0.03 (0.03)	19.5	8.6		0.67	0.58	0.09 (0.03)
Genomic SSR	17.18	5.76		0.59	0.50	0.13 (0.07)	18.1	11.56		0.80	0.68	0.10 (0.03)	19.1	10.7		0.78	0.65	0.15 (0.04)

Significance level: *P<0.05

Table 3 Allele sizes resulting from the amplification of 37 simple sequence repeat loci across 23 spruce species

Locus	WS00111.K13	WS0015.I04	WS0016.O09	WS0019.M09	WS0019.F22	WS0022.B15	WS0023.B03	WS0023.B12	WS0033.A18	WS0035.A01	WS0053.K16	WS0061.K02	WS0071.J15	WS0073.H08	WS0082.E23	WS0082.O23	WS0092.A19	WS0092.M15
P. sitchensis	236	242	418	227	382	220	228	228	182	184	243	228	228	230	254	248	240	229
	234	198	416	227	274	210	198	224	182	184	231	228	224	228	216	248	238	217
P. rubens	234	250	412	227	416	216	230	262	182	182	233	230	226	224	266	248	240	229
	234	242	412	227	406	216	198	242	182	182	233	228	222	224	248	232	236	223
P. lutzii	256	250	424	267	386	218	234	258	188	188	259	228	256	236	218	248	240	229
	236	198	418	255	374	202	224	240	180	186	233	226	248	230	218	248	238	217
P. mariana	240	224	412	287	412	192	228	260	172	182	229	228	252	224	228	252	238	229
	236	218	412	259	382	188	216	260	172	182	229	228	222	224	220	248	234	223
P. pungens	240	212	412	265	408	220	226	220	172	182	265	226	268	224	224	248	238	229
	234	212	412	259	408	214	210	210	172	182	229	226	248	224	190	234	238	217
P. glauca	276	240	422	303	436	218	220	210	182	188	225	228	232	236	224	248	236	229
	236	200	420	279	416	206	210	210	182	184	215	226	228	234	224	248	236	217
P. chihuahuana	234	230	414	279	416	210	210	248	192	182	233	226	234	226	228	242	238	232
	234	230	414	263	390	196	208	238	182	180	225	226	232	226	228	222	238	217
P. mexicana	234	204	416	281	394	214	210	282	168	178	225	228	244	228	308	248	234	232
	234	200	416	267	394	210	208	274	168	178	215	226	240	228	224	232	234	229
P. abies	240	224	418	259	380	210	216	248	188	180	243	226	236	230	224	248	244	229
	234	220	416	245	380	222	214	244	182	180	225	226	236	228	224	232	232	217
P. omorika	262	198	430	257	412	221	234	234	182	182	243	228	236	242	224	250	240	229
	260	198	422	257	404	208	216	232	182	182	225	228	224	236	216	244	236	226
P. obovata	264	230	418	267	382	222	196	226	186	182	225	232	264	232	224	248	236	229
	256	212	414	253	380	210	196	218	182	180	215	224	244	228	220	248	234	229
P. koraiensis	240	224	414	279	404	222	214	238	180	188	225	228	252	226	220	254	246	229
	238	194	392	271	400	218	204	226	180	186	215	228	238	204	212	248	246	217
P. montigena	242	194	414	253	408	218	204	264	180	184	263	228	242	226	218	248	236	229
	242	194	414	253	408	200	204	246	180	184	259	228	240	204	212	248	236	226
P. orientalis	244	250	414	225	426	198	222	224	194	180	235	226	222	226	–	248	240	229
	244	220	414	211	404	198	206	216	194	180	225	226	222	226	–	242	240	217
P. asperata	260	242	420	233	388	212	224	250	180	186	–	228	246	226	218	254	246	229
	256	242	400	227	384	212	224	218	180	186	–	228	240	204	212	248	232	217
P. meyerii	248	202	416	227	408	222	222	244	184	186	251	228	250	228	244	252	252	229
	238	202	414	227	380	218	196	232	180	180	225	228	238	226	218	234	236	217
P. wilsonii	246	226	414	235	402	216	204	256	198	186	225	228	246	226	224	248	252	229
	242	202	392	235	402	208	204	228	198	186	215	228	232	204	220	248	250	217
P. jezoensis	242	202	420	259	418	218	200	272	180	186	263	228	242	258	218	248	248	229
	240	194	400	235	398	210	194	242	180	186	225	228	240	238	212	240	246	226
P. polita	236	224	420	259	380	212	228	218	178	180	225	230	236	258	220	250	240	229
	236	212	420	235	374	210	198	218	178	172	225	230	228	256	212	246	238	226
P. schrenkiana	238	248	416	227	438	218	198	232	178	178	229	228	238	226	228	248	248	229
	238	242	416	227	410	208	198	222	178	174	225	228	234	204	220	238	244	226
P. koyamai	260	250	414	235	406	212	216	224	180	184	227	228	240	226	264	248	244	229
	242	234	414	229	406	212	214	222	180	182	227	228	234	226	228	244	244	217
P. glenhii	244	248	430	253	378	216	214	224	178	182	225	228	242	228	266	248	246	229
	240	238	422	247	374	208	204	224	178	182	225	228	236	228	234	248	244	217
P. bicolor	242	250	414	235	406	212	214	224	180	182	227	228	240	226	220	248	244	229
	242	238	414	235	406	212	214	222	180	182	227	228	230	226	212	242	244	217

Locus	WS0092.H13	PAAC 17	PAAC 19	PAAC 23	PGL 14	UAPgAG105	UAPgAG150A	UAPgAG150B	SPAG C1	SPAG G3	SPL 3AG 1A4	SPL 3AG 1H4	EAC 6A06	EAC 6B03	EAC 7C11	EAC 7H07	EATC 3C05	2	44
P. sitchensis	240	153	189	294	188	179	174	145	159	152	122	167	120	142	155	145	275	237	138
	237	153	183	284	172	173	172	145	135	124	120	135	120	134	129	131	275	205	130

Table 3 (continued)

Locus	WS0092.H13	PAAC 17	PAAC 19	PAAC 23	PGL 14	UAPgAG105	UAPgAG150A	UAPgAG150B	SPAG C1	SPAG G3	SPL 3AG 1A4	SPL 3AG 1H4	EAC 6A06	EAC 6B03	EAC 7C11	EAC 7H07	EATC 3C05	2	44
P. rubens	240	163	187	294	172	179	164	143	131	146	–	137	120	142	131	133	260	237	130
	240	157	179	284	170	175	164	143	123	130	–	135	120	134	131	123	260	221	130
P. lutzii	240	163	185	294	168	179	178	145	157	144	138	149	136	124	133	153	275	227	132
	237	153	177	284	166	175	176	145	149	134	102	135	134	124	125	153	275	227	128
P. mariana	240	167	187	298	170	181	168	145	141	154	–	155	140	154	153	129	263	219	144
	240	163	175	286	158	179	168	143	123	146	–	135	140	116	131	123	260	195	142
P. pungens	240	163	209	292	188	177	164	149	145	124	–	153	120	110	149	147	275	229	142
	240	155	201	280	156	173	162	143	141	124	–	135	120	110	143	143	254	223	128
P. glauca	240	179	199	294	188	205	–	145	169	150	124	179	146	168	147	145	254	255	128
	240	179	191	284	164	185	–	145	149	146	118	169	110	126	139	139	251	223	128
P. chihuahuana	240	157	185	294	162	175	172	149	145	140	–	151	146	132	157	131	242	209	122
	240	155	185	284	162	175	172	145	131	138	–	135	110	132	143	127	242	209	122
P. mexicana	240	165	185	294	158	181	–	145	93	150	108	153	126	130	151	147	–	219	130
	240	165	173	282	152	175	–	145	89	146	102	167	126	120	119	143	–	219	130
P. abies	240	157	175	300	188	175	170	149	135	150	106	151	136	114	129	165	305	209	148
	240	157	167	290	188	169	164	143	121	150	102	137	118	112	129	119	287	189	128
P. omorika	243	165	191	294	214	175	162	143	141	152	–	135	124	122	–	141	278	219	146
	240	157	183	290	214	175	162	143	141	148	–	135	122	114	–	139	275	215	128
P. obovata	243	167	193	298	188	179	164	145	121	162	118	137	120	110	131	121	290	199	152
	240	161	183	288	176	175	162	143	119	156	102	137	106	110	131	117	287	199	138
P. koraiensis	240	155	177	292	188	175	164	143	101	148	102	153	132	114	143	157	263	211	128
	240	155	169	284	156	169	160	143	97	146	102	145	132	112	143	147	263	211	128
P. montigena	240	155	177	296	182	177	162	145	107	140	–	139	124	120	149	135	269	229	126
	237	155	169	284	178	171	160	143	99	138	–	133	114	120	131	131	269	215	126
P. orientalis	240	155	185	292	188	175	160	145	121	152	94	135	122	110	145	157	257	215	126
	240	155	177	282	164	169	160	145	121	146	90	135	122	110	133	137	248	205	122
P. asperata	243	155	177	296	180	175	158	145	105	158	122	155	132	114	141	145	266	227	128
	243	155	177	284	172	175	158	143	101	158	122	135	128	112	135	133	251	209	128
P. meyerii	240	155	179	316	194	175	162	143	125	152	126	147	120	116	141	151	263	205	150
	240	155	179	296	188	175	158	143	103	146	112	135	106	112	141	141	251	205	126
P. wilsonii	240	155	177	296	188	175	158	145	121	158	126	137	120	118	151	139	263	237	130
	240	155	177	284	178	175	158	145	105	146	112	137	106	118	133	131	251	221	130
P. jezoensis	243	155	189	284	188	175	160	145	109	150	142	153	114	112	137	179	263	213	154
	243	155	189	290	188	175	158	143	101	146	124	139	106	112	133	149	251	213	128
P. polita	243	155	199	294	188	179	168	149	119	146	114	141	114	112	167	145	272	205	142
	243	143	181	290	166	179	168	145	109	142	114	135	102	112	137	145	257	191	130
P. schrenkiana	243	167	185	294	188	177	–	149	133	142	144	139	114	–	147	153	251	255	126
	240	155	185	282	188	177	–	145	129	142	122	139	114	–	141	149	251	223	126
P. koyamai	240	155	177	294	188	175	162	147	131	158	126	153	110	116	139	139	251	215	150
	240	155	177	282	188	175	162	143	101	146	108	137	110	116	139	133	251	215	130
P. glenhii	243	165	199	298	188	177	164	143	113	146	144	159	138	122	145	159	251	215	130
	243	165	177	288	188	175	162	143	105	128	122	137	114	122	143	151	251	215	130
P. bicolor	240	161	177	294	188	175	–	147	101	160	142	153	138	114	141	135	–	207	130
	240	155	177	282	188	175	–	143	101	152	112	137	114	114	141	135	–	207	130

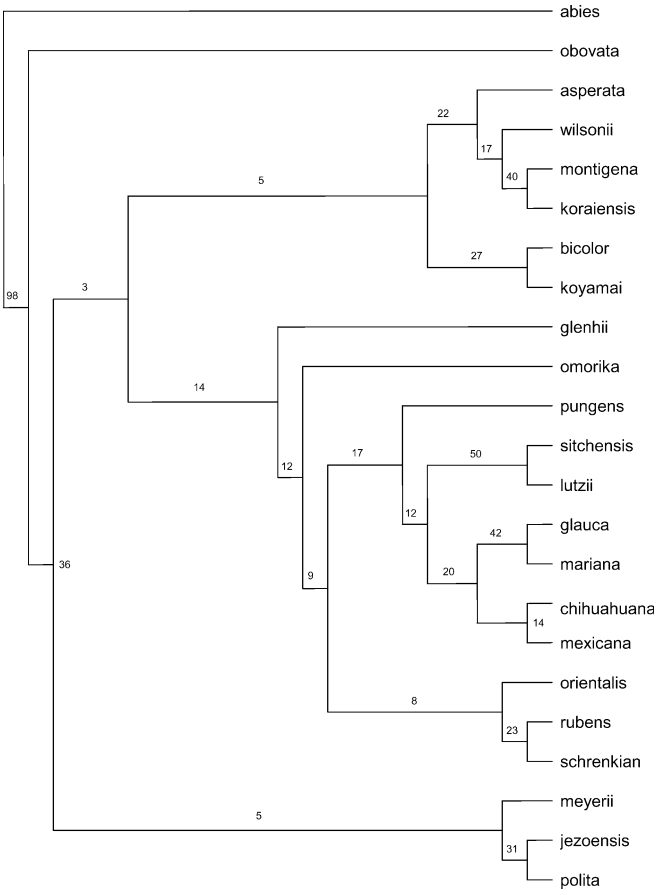

Fig. 4 Consensus tree of microsatellite genetic distances among 23 spruce species. The *numbers* at the forks indicate the number of times the group consisting of the species, which are to the right of that fork occurred among the trees, out of 100 trees

The *F* for specific SSRs allows identification of loci with putative null alleles, with those showing significantly higher *F* values indicating the presence of null alleles. Null alleles can bias estimates of genetic variation and genetic structure, and are not useful for genetic mapping. Also, loci with prominent stutter bands often exhibited higher *F* values, due to the difficulty in scoring of heterozygous genotypes for adjacent sized alleles. In contrast to the genomic DNA-derived markers, our EST–SSR markers gave more uniform *F* values the three spruce species. Two genomic-derived loci in particular showed consistent patterns across the three species: PAAC 19 (positive *F*) and UAPgAG150A (negative *F*).

Cross-species amplification of SSR markers

Of the 43 loci identified as informative in white, Sitka, black, and red spruce, the majority (33/43) were able to amplify alleles across all 23 spruce species tested. The minimum number of species in which a particular locus was identified was 17 (locus SPL3AG1A4). This suggests that the regions flanking the SSRs are well conserved across the spruce species tested, and that if a particular locus can be amplified from white, Sitka, black, and red spruce, then it is likely that the locus will be widely transferable throughout other spruce species as well. Of the 44 SSR markers developed from the EST database, 25 were informative in white, Sitka, and black spruce. From these 25 loci, 20 were identified in all 23 spruce species tested, while the remaining five loci were detected in 22 of the 23 species.

While SSRs are instrumental in genetic mapping (e.g., Dib et al. 1996), studies of kinship (e.g., Queller et al. 1993), and population structure (e.g., Bowcock et al. 1994), they have received limited use as a tool for phylogenetic reconstruction of closely related species (reviewed by Schlötterer 2001). This is mainly due to allele size homoplasy resulting from an exceptionally high mutation rate. However, when a genetic distance measure that takes into account the mutational process is used (Goldstein et al. 1995; Neff 2004), SSRs, particularly those developed from ESTs, can be potentially very informative in resolving newly diverged specific complexes or groups with slower rates of evolution.

Interestingly, the tree topology obtained by microsatellite genetic distances among species (Figs. 3, 4) was similar to that obtained from phenetic and cladistic analyses of chloroplast DNA RFLPs (Sigurgeirsson and Szmidt 1993). Highlights of this similarity include *P. mexicana* and *P. glauca* clustering together in congruence with Sigurgeirsson and Szmidt's (1993) "*P. glauca* alliance" and the association of *P. asperta*, *P. koyamai*, and *P. koraiensis*. Results from the bootstrap routine, however, showed no support for the branches of the phylogenetic tree generated. This is most likely a product of sampling a single individual per tree species. Because portions of the phylogenetic tree obtained matched the results of Sigurgeirsson and Szmidt (1993) and was in agreement with generally accepted views of classification within *Picea*, we propose that the microsatellite markers tested in this study, if applied to multiple individuals of each species, will likely prove to be powerful tools for investigating phylogenetic relationships within *Picea*.

Comparison of EST-derived SSRs with other SSRs in spruce

In this study, we found that the use of an EST database to develop novel SSR markers led to a high rate of success when compared to other studies. In addition, the EST–SSR markers developed and presented here have been readily transferable across species. Of 44 EST-SSRs, 25 were widely transferable across spruce species (~57%), while only 17 of 101 previously developed SSR markers were as widely transferable (~17%). These SSR markers are in the process of being placed onto a genetic linkage map of white spruce. This will increase their usefulness for other purposes such as population studies because markers evenly spaced throughout the genome will be able to be chosen. Also, the large allele size difference between different loci will allow placement of loci into "bins" for multiplexing. Although the potential for coamplification

of loci has not been tested yet, even post-PCR pooling of these loci will save time and money by reducing the number of gels that have to be run.

Previous attempts at developing SSR markers from conifer genomic sequences have been hampered by a low success rate due to many primer pairs yielding complex banding patterns that cannot be genetically interpreted. This has been attributed to the large proportion of repetitive or low complexity sequence present in conifer genomes (Pfeiffer et al. 1997). The use of an EST database to identify SSR markers has resulted in the development of a higher proportion of useful and informative loci. This study identified a preponderance of AT repeats from spruce ESTs, in contrast to other plant genomes. When a full-length EST database is available for spruce, we will be able to determine if this SSR distribution is confirmed or if it is an artefact caused by the 3′-sequence data currently in our EST database.

Acknowledgements Genome Canada and the Province of British Columbia, through the Genome BC Forestry Genome Project, funded this research. We acknowledge the support of the Vancouver Genome Sciences Centre for EST sequencing and database development. We thank Dr. Sally Aitken (University of British Columbia) for the white spruce collections, Washington Gapare (University of British Columbia) for the Sitka spruce collections, and Dr. Om Rajora (Dalhousie University) for the black spruce collections. Dr. Barry Jaquish of the B.C. Ministry of Forests Kalamalka research station provided the spruce species collection.

References

Arnold C, Rossetto M, McNally J, Henry RJ (2002) The application of SSRs characterized for grape (*Vitis vinifera*) to conservation studies in Vitaceae. Am J Bot 89:22–28

Bérubé Y, Ritland CE, Ritland K (2003) Isolation, characterization, and cross-species utility of microsatellites in yellow cedar (*Chamaecyparis nootkatensis*). Genome 46:353–361

Besnard G, Achere V, Faivre Rampant P, Favre JM, Jeandroz S (2003) A set of cross-species amplifying microsatellite markers developed from DNA sequence databanks in *Picea* (Pinaceae). Mol Ecol Notes 3:380–383

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Bruford MW, Wayne RK (1993) Microsatellites and their application to population genetic studies. Curr Opin Genet Dev 3:939–943

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 23:13–15

Ewing B, Green P (1998) Base-calling of automated sequencer traces using PHRED II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. Genome Res 8:175–185

Felsenstein J (1995) PHYLIP (Phylogeny Inference Package), ver 3.57c, University of Washington

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

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

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868–877

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Neff BD (2004) Mean d^2 and divergence time: transformations and standardizations. J Hered 95:165–171

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

Paglia G, Morgante M (1998) PCR-based multiplex DNA fingerprinting techniques for the analysis of conifer genomes. Mol Breed 4:173–177

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

Queller DC, Strassmann JE, Hughes CR (1993) Microsatellites and kinship. Trends Ecol Evol 8:285–288

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rajora OP, Rahman MH, Dayanandan S, Mosseler A (2000) Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulness in other spruce species. Mol Gen Genet 264:871–882

Ritland C, Ritland K (2000) DNA fragment markers in plants. In: Baker AJ (ed) Molecular methods in ecology. Blackwell, Oxford, pp 208–234

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy OUK, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schlötterer C (2001) Genealogical inference of closely related species based on microsatellites. Genet Res 78:209–212

Scotti I, Magni F, Fink R, Powell W, Binelli G, Hedley PE (2000) Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. Genome 43:41–46

Scotti I, Magni F, Paglia G, Morgante M (2002a) Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers. Theor Appl Genet 106:40–50

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

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

Soares MB, Bonaldo MF, Jelene, P, Su L, Lawton L, Efstratiadis A (1994) Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci USA 91:9228–9232

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339–1348

Ven WTG van de, McNicol RJ (1996) Microsatellites as DNA markers in Sitka spruce. Theor Appl Genet 93:613–617

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication Proc Natl Acad Sci USA 99:9650–9655