The Americas Journal of Plant Science and Biotechnology ©2011 Global Science Books

Tetranucleotide and Dinucleotide Microsatellite Markers for Red Spruce (*Picea rubens*)

Christophe M. Herbinger[1*] • Kristian Gordos[1,2] • Heather Allen[1,3]

[1] Department of Biology, Dalhousie University Halifax, N.S. B3H 4J1 Canada
[2] Present address: Pacific Biological Station, DFO, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada
[3] Present address: Forensic and DNA Services Division, Maxxam Analytics, Units 2 & 4, 335 Laird Rd. Guelph ON N1H 6J3, Canada
Corresponding author: * **christophe.herbinger@dal.ca**

ABSTRACT

Three library enrichment protocols were used to develop tetra-and dinucleotide microsatellite markers for red spruce (*Picea rubens*). Two enrichment methods based on a single round of magnetic beads-based subtractive hybridization showed limited success. A third method with two successive rounds of subtractive hybridization had a high enrichment success but was found to increase the rate of redundancy and chimerical sequences among the clone sequences. Several single locus tetra- and dinucleotide markers were discovered and can be useful in mapping studies. All newly developed markers except one were characterized by a high frequency of null alleles. One new and one previously described microsatellite marker were identified as informative, reliable and characterized by a low null allele frequency, and were suitable for paternity/parentage and population genetic analyses in red spruce.

Keywords: black spruce, conifers, magnetic bead enrichment, null allele, white spruce

INTRODUCTION

Red spruce (*Picea rubens*) is an ecologically and economically important species of late successional forests, now found predominantly in the Canadian maritime provinces and the North Eastern United States, with some isolated montane remnants in southern Appalachian regions (Blum 1990; Mosseler *et al*. 2000). It is shade tolerant and adapted to growth in upland areas with cool, moist climate (Manley 1972; Gordon 1976; Manley and Ledig 1979). Except over its southern Appalachian range, red spruce is sympatric with black spruce (*Picea mariana*), an early successional species with a large distribution in boreal and temperate North America. The two species are morphologically and genetically similar and are known to hybridize naturally, although estimates of the range of natural hybridization and introgression are somewhat unclear (Gordon 1976; Manley and Ledig 1979; Perron *et al.* 1995; Bobola *et al.* 1996; Major *et al.* 2005). Red spruce also seems to be less genetically diverse than other boreal conifers, and based on allozyme, random amplified polymorphic (RADP) and sequence tagged site (STS) marker data, it has been proposed that red spruce is a derivative species of black spruce (Perron *et al.* 1995, 2000). Red spruce has declined substantially over its geographical range, likely attributable to silviculture practices such as clearcutting (Leak and Smith 1996) as well as climate and adverse environmental changes (Hamburg and Cogbill 1988). It is now found in small isolated patches over much of its original range and there is growing interest in its conservation and restoration (Gordon 1996; Mosseler *et al.* 2000). This is also a species of silviculture interest and a red spruce breeding program has been underway in Nova Scotia since 1977 (Fowler 1986).

Access to highly informative genetic markers would be very useful for population genetics, conservation genetics and breeding studies of red spruce. Nuclear microsatellite markers, also known as simple sequence repeat (SSR), are among the most powerful classes of current genetic markers (Agarwal 2008). They are widely used in many fields of research, such as genome mapping, marker-assisted breeding programs, paternity and parentage determination, and population diversity and differentiation over a very large range of species.

Microsatellite markers developed in one conifer species can sometimes be used successfully in related species (e.g. Hodgetts *et al.* 2001; Rajora *et al.* 2001). However, their cross species transferability is often quite limited (Rungis *et al.* 2004). Many microsatellite markers that appear initially promising based on successful amplification in a few individuals are often found later to be problematic to use on a large scale (e.g. Doerksen and Herbinger 2008). Thus, microsatellite markers often need to be developed afresh for a new species of interest. Efficient strategies for microsatellite isolation have been published (e.g. Zane *et al.* 2002), but this development phase is nonetheless challenging and it has proven to be complex and fairly onerous for plants (Squirrell *et al.* 2003). The development of microsatellite DNA markers in conifers has been especially challenging and slow. This has been inferred to be a consequence of their very large genome of which the bulk is assumed to be repetitive DNA; a high proportion of primer pairs produce multi-band patterns or poor amplification and cannot be used (Pfeiffer *et al.* 1997; Elsik *et al.* 2000; Elsik and Williams 2001a; Scotti *et al.* 2002a, 2002b). To circumvent this problem, various approaches have been attempted, including recovery of microsatellites from libraries enriched in low-copy (i.e. non-repetitive) DNA (Elsik and Williams 2001a; Scotti *et al.* 2002b); targeting trinucleotide microsatellites that are preferentially found in low-copy sequences (Scotti *et al.* 2002a); using libraries enriched for undermethylation and presumably containing less highly repetitive DNA (Zhou *et al.* 2002) or using Expressed Sequence Tags (EST) databases to recover single, low-copy, conserved sequences for microsatellite marker development (Rungis *et al.* 2004; Bérubé *et al*. 2007).

The present study intends to present a realistic picture of the challenges associated with microsatellite marker development in red spruce. We characterize several dinuc-

Received: 20 December, 2009. Accepted: 5 May, 2011.

Original Research Paper

leotide and tetranucleotide markers isolated from red spruce (*Picea rubens*) that were developed using three library enrichment protocols. We also provide information on the amplification success of these loci in four other spruce species, and we discuss the presence of null alleles. We focused on several types of tetranucleotides (AAAG, GATA, GTAT, GGAT and GACA) as well as one type of dinucleotide (AC) marker. Even though the AG dinucleotide is a more frequent microsatellite motif in conifer genomes, the AC motif was chosen because it had been shown to be over-represented in low-copy DNA in pine and spruce (Elsik and Williams 2001b; Scotti *et al.* 2002b). Although generally less common than dinucleotide markers, tetranucleotide markers are much less prone to stuttering (i.e. the presence of stutter or shadow bands next to the real amplification product). This problem is thought to result from replication slippage during Polymerase Chain Reaction (PCR) and can create difficulties for accurate determination of allele size (O'Reilly *et al.* 2000; Scotti *et al.* 2002a). Also, alleles at dinucleotide markers are typically separated by a minimum of 2 bp, less than alleles at tri- and tetranucleotide loci, increasing the difficulty of resolving some adjacent alleles (Scotti *et al*. 2002a). In conifers, tri- and tetranucleotide repeat motifs are also thought to be more common in low-copy DNA (expressed regions or pseudo-genes) than dinucleotides (Scotti *et al.* 2002a).

MATERIALS AND METHODS

Library enrichment protocols to increase microsatellite recovery

Method I: Single enrichment from untreated DNA

DNA was extracted from 80 mg of -20ºC frozen red spruce needles using DNeasy Plant Mini kit spin columns (Qiagen) following manufacturer's specifications, with the following modifications: proteinase K was added to the lysis buffer (5% final concentration) and the lysis time was extended from 10 min to 2 hrs. These modifications resulted in the doubling of our DNA yields, in comparison to the yields without the modifications.

The library enrichment protocol follows the Hamilton *et al.* (1999) procedure with some modifications as described in McPherson *et al.* (2001) and Herbinger *et al.* (2006). Six mixtures of Streptavidin-coated magnetic beads (Dynabeads, Dynal) were created in the presence of the biotinylated (5′) oligonucleotide probes $(AC)_9$, $(AAAG)_7$, $(GATA)_7$, $(GTAT)_7$, $(GGAT)_5$ and $(GACA)_4$. The spruce genomic DNA of a single individual tree was digested with *Hinc*II restriction enzyme (New England Biolabs), dephosphorylated with CIP (calf intestinal phosphatase) and ligated to the SNX linker complex in the presence of the *XmnI* restriction enzyme (New England Biolabs). The recombinant fragments were amplified using the SNX forward linker as the primer, taking care to prevent over-amplification of the SNX linker-ligated genomic DNA (Zane *et al.* 2002), then denatured and incubated with the six different magnetic bead-probe complexes at 52°C for 20 min to allow for hybridization to the specific oligonucleotides.

To remove the non-targeted DNA, the mixtures were washed three times at 52°C with decreasing concentrations of sodium chloride sodium citrate (SSC): 2X, 1X, 0.5X SSC each with 0.5 ηg/μl SNX Forward linker and with 0.1% SDS. Following the washes, the potentially enriched linker ligated DNA fragments were released by incubation in 60 μL of TE buffer at 96°C for 15 min. The recovered DNA was PCR amplified, again taking care to prevent over-amplification, purified using QIAquick PCR purification kits (Qiagen) and ligated to the pDrive Cloning Vector (Qiagen). Invitrogen's ONE Shot TOP10 competent cells were transformed using the recombinant vector-insert DNA and were plated and grown in the presence of ampicillin.

Plasmid DNA was isolated from each of 864 random clones [192 for each of the $(GATA)_7$, $(GGAT)_5$ and $(GTAT)_7$ libraries and 96 for each of the $(AC)_9$, $(AAAG)_7$ and $(GACA)_4$ libraries] using the QIAprep 96 Turbo Miniprep kit (Qiagen), and cycle sequenced on a CEQ 8000 capillary sequencer (Beckman Coulter) using the DTCS Quick Start Mix (Beckman Coulter) and M13 primers.

Method II: Single enrichment from McrBC-treated DNA

The *McrBC* digestion of genomic DNA was carried out to enrich for hypomethylated, low-copy DNA and was performed as described in Zhou *et al.* (2002) with minor modifications. The genomic DNA was obtained from -20°C frozen red spruce needles using DNeasy Plant Maxi kit spin columns (Qiagen) following manufacturer's specifications and modifications described above. 40 μg of genomic DNA was digested overnight with 400 U of *McrBC* (New England Biolabs). The digested DNA was run on a 1% agarose gel in 1X Tris-Borate-EDTA (TBE). The digested fragments that were larger than 10 kb were excised and extracted using a QIAEX II gel extraction kit (Qiagen). The library enrichment protocol was the same as described above with the exception that the *McrBC* digested DNA was enriched for only the following repeats: $(GATA)_7$, $(GTAT)_7$ and $(GGAT)_5$. Plasmid DNA was isolated from 288 random clones [96 for each of the $(GATA)_7$, $(GGAT)_5$ and $(GTAT)_7$ libraries] and sequenced as described above.

Method III: Double enrichment from McrBC-treated DNA

The DNA that was used for the library double enrichment protocol was the same DNA as that used for the single enrichment of *McrBC* treated DNA. This protocol followed the two previously described protocols until the post-magnetic bead recovery step. For the double enrichment protocol, the following repeat elements were used: $(AC)_9$, $(AAAG)_7$, $(CATC)_4$ and $(GACA)_4$.

The major modification in this method was that instead of inserting the amplified and enriched DNA into a cloning vector, the magnetic bead hybridization procedure was performed for the second time on the DNA recovered from the first selective hybridization step. At the end of the second hybridizations step, three parallel PCR were carried out with 5, 10 and 15 cycles, for each of the repeats. The products of the three PCR amplifications were visualized using ethidium-bromide stained 1.5% agarose gel, and the detected reaction product generated using the lowest number of cycles was used in subsequent cloning as described in Method I. Preliminary screening indicated that smaller inserts generally did not contain both microsatellite repeats and sufficient flanking sequences. Colonies were thus screened for large inserts and plasmid DNA was only isolated from colonies that showed an insert of 500 bp or more. Plasmid DNA isolation and sequencing was performed as described in Method I.

Characterization of microsatellite markers

Primers for PCR amplification were designed using Primer3 (Rozen and Skaletski 2000) for all unique microsatellite sequences containing appropriate flanking regions, and were tested on a panel of 8 individuals. Buds and needle tissue was collected from one seed orchard (Debert, NS) containing the seed parents involved in the red spruce tree breeding program administered by the Nova Scotia Tree Improvement Working Group (Fowler 1986). In addition, tissue samples were also collected from 13-16 progeny trees in each of 4 full sib family crosses. About 50 mg of frozen tissue was reduced to a fine powder in a Retch mill (Retch Gmbh & Co.) and total DNA was extracted with a Qiagen plant DNeasy kit. Each reaction was carried out in a final volume of 10 μL at a concentration of 1X buffer (10 mmol KCl, 10 mmol $(NH_4)SO_4$, 20 mmol Tris-HCL, 2 mmol $MgSO_4$, 0.1% Triton X-100 (pH 8.8 at 25°C)) with 10-100 ηg of template DNA, ~ 0.5 ρmol forward and reverse primers, 0.2 mmol dNTP each and 0.5 U *Taq* DNA polymerase (New England Biolabs). PCR consisted of an initial denaturation at 95°C for 2 min followed by 30 cycles with denaturation at 95°C for 20 s, annealing for 20 s, extension at 72°C for 30 s and a final extension at 72°C for 3 min. The forward primers were HEX labeled (Integrated DNA Technologies) and visualized using an FMBIO II (Hitachi Software Engineering).

For all markers loci with detected product amplification, the total number of alleles observed was tabulated in 32 to 296 individual trees. Most non-amplifying (blank) individuals were assayed at least twice. For each locus, a lower bound to the null allele frequency was estimated by assuming that all individuals that gave no amplification signal were carrying 2 null copies and that there were no other null alleles present in the tested set of individuals. In

other words, if *x* individuals were tested of which *y* individuals were blanks, a lower bound to the null allele frequency was *2y/2x* or *y/x*. An upper bound to the null allele frequency was estimated by assuming that in addition to the non-amplifying individuals carrying 2 null copies, every single-band individual actually carried a null allele in addition to an amplified allele. If *x* individuals were tested of which *y* individuals were blanks and *z* individuals showed only one band (one allele), an upper bound to the null allele frequency was *(2y + z)/2x*. Strict Mendelian inheritance of parental alleles was verified in four full sib families consisting of both parents and 13 to 16 progeny. When at least one of the parents was homozygous, possible Mendelian inheritance of the observed parental allele and inferred null alleles was verified as well. In addition, two microsatellite loci originally developed for white spruce (*Picea glauca*), Pgl 12 (Rajora *et al.* 2001) and UAPGCA24 (Hodgetts *et al.* 2001), were assessed as well, as they appeared promising to use with red spruce.

The presence of amplification products was also verified in 4 to 12 black spruce (*Picea mariana*), and 1 to 3 individuals each for white spruce (*Picea glauca*), Norway spruce (*Picea abies*) and Sitka spruce (*Picea sitchensis*). Finally, in three cases where the original vector insert had sufficient sequence information, new primers were designed to amplify both the variable region and several hundred bases of the flanking regions. Multiple individuals within and across species were then sequenced to determine the rate of substitution in the flanking region and new primers were then redesigned to be positioned in less polymorphic regions.

RESULTS

Library enrichment success

Three microsatellite library enrichment methods were used to isolate microsatellite markers from *Picea rubens* with varying levels of success as determined by actual sequencing of clones from 13 libraries (**Table 1**). Method I, a slightly modified version of Hamilton *et al.* (1999) involving a single round of genomic DNA enrichment for different repeat motifs resulted in very low levels of enrichment. In the case of the various tetranucleotide motifs, only 2.9% (19/647) of the sequences actually contained microsatellite repeats, and out of these only 0.8% (5/647) had suitable flanking sequences for primer development (**Table 1**). Enriching for AAAG and GACA motifs was completely unsuccessful; no microsatellite containing sequence was found among 185 sequenced clones. Enriching for GATA, GTAT and GGAT motifs was marginally more successful. Nineteen clones containing at least 3 tandem repeats were found among 462 clones sequenced, but ultimately, only 5 exhibited enough unique flanking sequence for primer development. Success with the single attempted dinucleotide repeat motif (AC_9) was higher with 24% (20/83) of the sequences containing at least 6 AC tandem repeats, but only 4.8% (4/83) contained sufficient unique flanking sequence to allow primer development.

Method II used *McrBC*-digested DNA, presumably enriched for undermethylated DNA (Zhou *et al.* 2002). It involved a single hybridization step and was tried with the GATA, GTAT and GGAT repeat motifs, as they exhibited slightly higher success among the tetranucleotide motifs tried in the first treatment (Method I). The overall level of success was still extremely low (**Table 1**). Twenty clones (out of the 280 clones sequenced, or 8.9%) contained 3 or more tetranucleotide microsatellite repeats, compared to 4.1% (19/462) seen with Method I. Ultimately, only three clones among the 280 sequenced clones (1.1%) were suitable for primer development, a similarly low success rate as Method I (1.1%, 5/462).

In contrast, Method III, with a double enrichment step, resulted in a large number of sequences containing microsatellite repeat elements (**Table 1**). Double enrichment with the three attempted tetranucleotide motifs resulted in 95.6% (264/276) of the clones sequenced containing at least three tetranucleotide repeats. Among the motifs screened, two (AAAG and GACA) had been attempted with method I and did not yield a single microsatellite containing sequence. Ultimately, a much smaller portion (13.7%, 38/276) of the sequences were suitable for further primer development, but this still represented a substantial relative increase, especially considering that no GACA containing sequences were found with Method I and that GGAT (same motif as CATC) showed limited enrichment success, 7.6% in Method I and 3.3% in Method II, but had produced no useful sequences for primer development. In the case of the AC dinucleotide motif, Method III resulted in similarly high enrichment success with 92.7% (51/55) of the clones sequenced containing at least 6 AC repeats, and 36.4% (20/55) suitable for primer development.

Table 1 Rates of success of the three enrichment methods for various microsatellite repeat motifs.

Method	Repeat element	Number of clones with:			
		no repeat element	3 to 6 repeat elements	more than 6 repeat elements	appropriate flanking sequence[a]
I	AC	63	12	8	4
	AAAG	94	0	0	0
	GATA	181	3	1	2
	GTAT	177	5	3	3
	GGAT	85	0	7	0
	GACA	91	0	0	0
II	GATA	79	11	4	2
	GTAT	85	4	3	1
	GGAT	91	3	0	0
III	AC	4	7	44	20
	AAAG	2	8	43	0
	CATC	7	14	95	23
	GACA	3	17	87	15

[a] Clone containing at least 6 repeat elements and enough flanking sequence information (generally 30-50 bp) to locate primers

For Method III, there was a marked decrease for all repeat motifs in the number of suitable sequences for primer development as compared to the number of microsatellite containing sequences. This was due to increased levels of redundancy and a reduction of fragment size. When compared to Methods I and II, redundancy among the various clones sequenced increased 10-fold (27% as opposed to 2.9%) and the number of clones with insufficient flanking sequences for primer development also increased by 33%. The double enrichment process also resulted in a number of sequences that shared either the 3′ or the 5′ flanking region of the microsatellite-bearing fragment.

Marker development and characterization

Several potentially useful polymorphic dinucleotide and tetranucleotide markers for red spruce were discovered and characterized (**Tables 2** and **3**). The majority of the dinucleotide markers were derived from the sequences resulting from library enrichment Method III. Interestingly, no such success was observed with the tetranucleotide markers: every primer pair developed from the 38 tetranucleotide-containing sequences resulting from library enrichment method III (**Table 1**) failed to generate sufficient amplification product to permit detection. The five tetranucleotide markers reported in **Tables 2** and **3** (RS 2, 3, 12, 22 and 24) were all developed from the 8 sequences derived from Methods I and II.

RS 27 and RS 54 primer pairs produced multilocus bands without readily interpretable segregation patterns. Conversely, RS 7 and RS 12 appeared to be fixed in all spruce species we tried. Most other loci were moderately variable. RS 2 only generated enough products to permit detection in red spruce. The other loci resulted in amplification products that were very similar in all species tested. For each locus, every allele size seen in the alternate spruce species tested here was also seen in red spruce.

Loci AC 136 and UAPGCA24 (Hodgetts *et al.* 2001)

Table 2 Primer sequences, core repeat motifs and annealing temperatures for 12 new and 2 already described microsatellite markers for red spruce.

Locus ID	Primer sequence	Repeat motif	Annealing temp. (°C)
RS 2	F-AGT TAG CCT GTT TGC CTT CC	(GTAT)	65-60 TD[b]
	R-CTG TGC AAA CGC TAT TCT GG	perfect	
RS 3	F-GAG ATA TGG TCA CAA ATG TTT TAC C	(GTAT)	67-62 TD
	R-TGA AGG GTC CAA AGA ATA TAT GTA G	perfect	
RS 7	F-GAA GCT TGT TTC AAT TCG TGT TGT G	(GATA) (AC)	56
	R-AAA CAA ACC TCT TAT TGC GAT CCT G	di-tetra compound	
RS 12	F-TCA TGT TCT TGC TCT TGA TCG	(GATA)	56
	R-CAA GCT CTG AGA GTT CCG TTG	variable imperfect	
RS 22[a]	F-CCT CAA AAG GGC ATC TTA ACC	(GATA)	67-62 TD
a&b	R-CTT TGA AGA AGT TTT GTG GGA TG	perfect	
RS 24	F-TTT GAA CAT GGG ACC TCT GG	(GTAT)	67-62 TD
	R-GGC ATG TTC ACC TAC ATT TG	perfect	
RS 27	F-GGA CAT TCC AAT GGA GTA ACA G	(AC)	56
	R-GGG GTA CCC TAT GAA CCA AC	perfect	
RS 54	F-CTC ATT ATC ACT TCA TCA TCT TGT G	(TC) (AC)	60
	R-CCT CCC AGA TAA GAG AGA GAA AG	di-di compound	
RS 26	F-CAT CCT GTT TGG GGC AAT AC	(CA) (TA)	67-62 TD
	R-GCA GCG AAT GGG TAT CAG AC	di-di compound	
RS 52	F-GGA TAT GGA AAC ATG GCA ATG	(AC)	67-62 TD
	R-AGT CTG TCC ATG ATG ATA AAG AAG TTC	perfect	
AC 54	F-TTG AAA TGA AGA AAT GGA GG	(AC)	56
	R-TCT CTT CCC GTT CTC GTC	perfect	
AC 136	F-GTT CTG CTT CGT GTT TGT GTT C	(AC)	56
	R-GAG TTT CGC CAG TTC TGA GG	perfect	
Pgl 12	see Rajora *et al.* 2001		60-54 TD
UAPGCA24	see Hodgetts *et al.* 2001		67-62 TD

[a] RS 22 a&b. This primer set amplifies two separate loci in red spruce, hereafter designed RS 22a and RS 22b
[b] TD. Variable annealing temperature, beginning with the first °T value specified and declining by 1 degree each cycle to the last °T value specified, and then staying constant prescribed cycle

Table 3 Number of alleles, null allele frequency, Mendelian inheritance observed or inferred in red spruce, and number of alleles observed in four other spruce species for 12 new and 2 already described microsatellite markers.

	Red spruce				Black spruce	White spruce	Norway Spruce	Sitka spruce
Locus ID	# observed alleles / # Individuals assayed	Null allele frequency		Mendelian inheritance, including inferred null alleles	# observed alleles / # Individuals assayed			
		Lower bound	Upper bound					
RS 2[a]	3 / 93	0.69	0.83	verified in 4 families	0 / 12	0 / 3	0 / 3	0 / 3
RS 3[b]	3 / 256	0.41	0.62	verified in 4 families	3 / 12	2 / 3	1 / 3	2 / 3
RS 7[c]	1 / 32	N/A	N/A	N/A	1 / 12	1 / 3	1 / 3	1 / 3
RS 12[c]	1 / 32	N/A	N/A	N/A	1 / 12	1 / 3	1 / 3	1 / 3
RS 22[d]	3a / 40	0a	0.28a	verified in 4 families	? / 12	? / 3	? / 3	? / 3
a&b	4b / 40	0.19b	0.41b	(a&b)				
RS 24	2 / 40	0.75	0.95	verified in 4 families	3 / 12	0 / 3	1/ 3	3 / 3
RS 27[e]	24? / 256	0.06	0.23	could not be verified	17? / 12	9? / 3	6? / 3	3? / 3
RS 54[f]	7 / 88	0.18	0.46	verified in 4 families	? / 12	? / 3	? / 3	? / 3
RS 26	4 / 66	0.16	0.45	verified in 4 families	3 / 5	1 / 2	3 / 2	2 / 1
RS 52[g]	3 / 39	0	0.41	verified in 4 families	2 / 5	1 / 2	3 / 2	1 / 1
AC 54[e]	many / 39	?	?	could not be verified	many / 5	? / 2	? / 2	? / 1
AC 136	8 / 256	0	0.06	verified in 4 families	2 / 4	1 / 2	1 / 2	1 / 1
Pgl 12	5 / 99	0.2	0.49	verified in 4 families				
UAPGCA24	9 / 256	0.01	0.09	verified in 4 families				

[a] RS 2 only amplifies in red spruce
[b] RS 3. This primer set also amplifies a second much larger locus
[c] RS 7 and RS 12 are fixed in all species
[d] RS 22 a&b. This primer set amplifies two separate loci in red spruce. Loci are difficult to assay in other species
[e] RS 27 and AC 54. Multilocus patterns in all species. Segregation pattern could not be deduced
[f] RS 54. Limited stuttering in red spruce but much more pronounced in alternate spruce species
[g] RS 52. Low variability with 1 common allele observed in most samples

were the most informative markers for red spruce, both because of higher allelic richness and because of inferred low null allele frequency (**Table 3**). Mendelian inheritance of parental alleles at these two markers was confirmed in four families (e.g. AC 136 in **Fig. 1**, right panel). Nine other easily scored loci (RS 2, RS 3, RS 22a, RS 22b, RS 24, RS 54, RS 26, RS 52, Pgl 12) displayed an inferred high frequency of null alleles. The presence of null alleles was particularly marked for the tetranucleotide markers (**Table 3**). These 9 loci did not generally exhibit strict Mendelian inheritance of parental alleles in the tested four full sib families. However, Mendelian inheritance could be inferred to take place when including assumed parental null allele(s). An example is provided for one of the four families for RS 3 (**Fig. 1**, left panel). The male parent in this family was homozygous while the female parent was heterozygous (**Fig. 1**). All 14 offspring displayed one or the other of the female alleles, but only 8 offspring were heterozygous and displayed the male allele as well. The other 6 offspring (# 2, 3, 9, 12, 13, 14) were homozygous and only displayed a female allele (**Fig. 1**). This pattern of inheritance can easily be reconciled with Mendelian inheritance if we assume that the male parent is actually heterozygous with one amplifying allele and one non-amplifying (null) allele. In summary, the presence and inheritance of null allele(s) was inferred as described above for every easily scored locus except AC

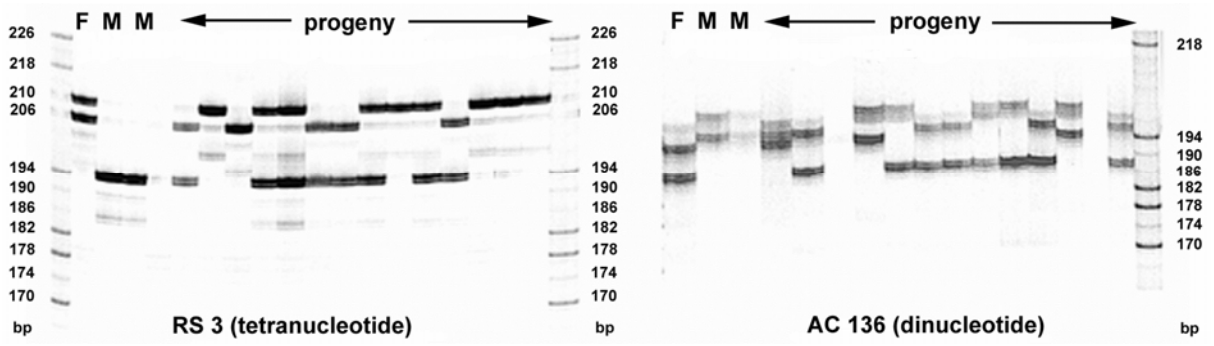

Fig. 1 Observed genotypes at locus RS 3 and AC 136 for the male and female parents and 13-14 progeny in one of the four tested families.

136 and UAPGCA24 in at least one of the four tested families.

In the case of clones RS 2, RS 3 and RS 24, the original vector insert had sufficient sequence information to allow the redesign of new primers to amplify both the variable tandem array region and several hundred bases of the flanking sequence. This permitted the sequencing of the microsatellite and its adjoining flanking regions in multiple red spruce individuals. In all three cases, a high number of closely spaced polymorphic sites were observed in the flanking regions. These polymorphisms were mostly due to nucleotide substitutions and were located every 15 bp in average. The flanking regions of RS 3 and RS 24 were also sequenced for several individuals of the other spruce species. Interestingly, the same sites were polymorphic in all species. Based on the sequence information, it was possible to redesign primers in slightly less polymorphic regions for RS 2 and RS 24. This resulted in a moderate decrease in the number of inferred null alleles at these two loci. For example, the original set of primers for RS 2 and RS 24 permitted the detection of 3 and 2 amplification products, respectively among the 8 parents of the four full sib families, while the re-designed primers allowed the detection of 6 and 5 amplification products respectively. Nonetheless, null alleles were still quite frequent, considering that 8 to 16 amplification products should be detectable, 8 products if all 8 parents were true homozygotes and 16 if all 8 parents were heterozygous.

DISCUSSION

Library enrichment

1. Enrichment success with conventional approaches

Method I, which involved a single magnetic bead hybridization step showed very weak library enrichment success in the present study. Yet, this protocol has been repeatedly and successfully used by our laboratory to isolate tetranucleotide microsatellites from marine and freshwater fish and shellfish species (e.g. McPherson *et al.* 2001; O'Reilly *et al.* 2002; Herbinger *et al.* 2006). The limited enrichment results observed here with red spruce were likely due to strong competitive hybridization taking place during the magnetic bead enrichment step. This non-specific or competitive hybridization was possibly a consequence of the large, complex and highly repetitive nature of the red spruce genome. Pfeiffer *et al.* (1997) had noted early on that the presence of microsatellite repeats in highly repetitive DNA regions in conifers could impact the success of library enrichment protocols. It appears that the probes attached to the magnetic beads only found a few targeted sequences, and other fragments with similar but not identical sequences instead hybridized to the probes. Close examination of many of the recovered sequences that contained no true microsatellite repeats did indeed reveal sequences with similarities to the search motifs used. Similarly, Garner (2002) observed a significant relationship between genome size and the proportion of non-amplifying microsatellite primer sets. This was attributed to a dilution of the proportion of available target DNA as well as an increasing probability of non-specific binding of primers in species with large genomes. Increasing the stringency of the hybridization was attempted but did not appreciably increase the microsatellite recovery rate (CM Herbinger, K Gordos, H Allen, unpublished data). In order to increase the efficiency of microsatellite recovery, we pre-treated the spruce genomic DNA with *McrBC* (Method II). *McrBC* digestion does not eliminate all highly repetitive DNA because it does not cleave every type of methylated site, but it should enrich for undermethylated, non duplicated DNA (Zhou *et al.* 2002). A marginal increase in microsatellite recovery efficiency was observed (**Table 1**) but ultimately, only 1.1% of the clone sequences were suitable for subsequent primer development.

2. Enrichment success and problems with double enrichment protocol

Adding a second round of hybridization (Method III) was then attempted to try to increase microsatellite enrichment success as a simple alternative to having to screen and sequence a very large number of clones to isolate a sufficient number of dinucleotide and tetranucleotide microsatellites. Adding this step allowed a considerable increase in specific enrichment success, as seen by the number of clone sequences that actually contained microsatellite repeats (**Table 1**). It seems probable that the second enrichment step eliminated most of the non-specific DNA fragments that were not screened out by the first round of enrichment. Based on these results, the double enrichment method would appear to be a promising modification of standard protocols for isolating microsatellite markers from red spruce and possibly from other conifer species. However, complications were encountered. A sharp increase in the number of clones with similar or identical sequences to other clones isolated using method III was observed. This was possibly due to the fact that the population of inserts underwent a severe bottleneck during the two consecutive hybridization steps involved in this particular procedure. Duplication among the clone sequences is a common observation with most library enrichment process and the potential for very efficient enrichment protocols to lead to higher level of redundancy had been noted (Squirrell *et al.* 2003).

Another potential problem might be related to the PCR amplification process. As noted, Method III resulted in a number of clones that exhibited the same sequence on either the 3′ or 5′ side of the microsatellite repeat array. This could be an indication that these sequences were chimera created during the library enrichment process. Such artifacts of PCR amplification are thought to result when prematurely terminated amplification products of two similar templates can anneal to each other, forming a new recombinant sequence during the PCR (Cronn *et al.* 2002). This could be especially problematic for microsatellite enrichment protocols because of the sequence similarity of the targeted fragments, namely the repeat array itself. Putatively chimerical sequences where one of the flanking regions but not the other was identical among different clones have been ob-

served during library enrichment (Squirrell *et al.* 2003; Roratto *et al.* 2008). In their attempt to enrich for trinucleotide microsatellite markers for Norway spruce, Scotti *et al*. (2002a) also noted that a large number of primer pairs did not produce any detectable product via PCR, possibly because the cloned sequences were chimeras of different fragments. In the present study, steps were taken to reduce as much as possible the number of PCR cycles for all methods. However, due to the additional PCR amplification performed after the second round of magnetic bead hybridization, Method III was probably more prone to such PCR artifacts than enrichment methods based on a single hybridization step. This may be the reason why most of the primer sets (including all the tetranucleotide motifs) derived from Method III yielded no amplification products in the present study. Alternatively, it could be that one of the two regions flanking the repeat array was embedded in a family of highly repetitive elements, resulting in the failure to detect PCR amplification products. Such a multilocus microsatellite arrangement has been observed in the genome of Lepidoptera (Van't Hof *et al.* 2007).

Marker development and characterization

Several potentially useful markers were developed in this study, including, to the best of our knowledge, the first tetranucleotide microsatellites reported for any spruce species. These markers can be informative in a number of areas such as for example, genetic mapping, individual tree identification and paternity tracing in tree breeding programs or estimation of diversity and genetic distance among populations. In fact, several of these loci have already helped discover one misidentified tree in a Nova Scotia clone orchard. This orchard contains multiple copies (multiple grafted trees or clones) of 20 individuals that have been used as source of pollen for pollen polymix preparation. The identity of all the clones was verified with several loci developed here and one clone was recognized as being misidentified. As a result, the group of pollen parents that have been used in the red spruce tree breeding program since its inception was shown to actually comprise 21 individual trees and not 20 as originally planned (Fowler 1986). This had important consequences for the subsequent study of male reproductive success and pedigree error in a polycross mating system derived from this pollen source (Doerksen and Herbinger 2008).

1. Patterns of segregation, variability and null alleles

Interestingly, only 2 markers (RS 27 and AC 54) showed multi-band patterns without clear segregation (**Table 3**). The common occurrence of such non-interpretable patterns has widely been reported to be a major source of problems when developing microsatellite markers in conifers. It has been attributed to the large and highly repetitive nature of their genomes (Pfeiffer *et al.* 1997; Elsik *et al.* 2000; Elsik and Williams 2001a; Scotti *et al.* 2002a, 2002b). The markers developed in the present study were based on AC and various tetranucleotide repeat motifs that have been reported to be over-represented in the low-copy DNA in pine and spruce (Elsik and Williams 2001b; Scotti *et al.* 2002a, 2002b). In addition, most markers were derived from libraries enriched for undermethylated DNA and presumably less duplicated (Zhou *et al.* 2002). It would appear that these strategies were effective in limiting the occurrence of markers with problematic multi-band patterns.

The tetranucleotide markers reported were, as anticipated, easier to assay than dinucleotide markers because of limited stuttering and larger size differences between adjacent alleles. However they all suffered from two major problems: very limited variability and a high frequency of null alleles. Two tetranucleotide markers were fixed and 5 others only displayed 2 to 4 alleles in a large group of seed parents from the red spruce tree breeding program (**Table 3**). In contrast, in an analysis involving 6 dinucleotide markers and the same group of trees, an average of 15 alleles per locus was observed (Doerksen and Herbinger 2008). Trinucleotide loci also exhibited lower polymorphism compared to dinucleotide markers in Norway spruce (Scotti *et al.* 2002a) and in Sitka, white and black spruce (Rungis *et al.* 2004). Additionally, the five variable tetranucleotide loci developed here showed very high occurrence of non-amplifying (null) alleles, and this was confirmed by examining Mendelian inheritance in four families. In 3 cases where sequencing the primer regions was possible, the predominance of non-amplifying alleles was shown to result from high polymorphisms in the primer annealing sites. Repositioning primers in less variable regions improved the situation somewhat, but the null allele frequency remained very high. Most of the dinucleotide markers developed here displayed moderately high variability, but also exhibited a high frequency of null alleles, except for AC 136 and for UAPgCA24. It is becoming increasingly clear that null alleles are especially common in several taxa, including, in particular, mollusks (Carlsson 2008; Herbinger pers. obs.). This also seems to be the case with conifers. Non-amplifying alleles are best detected when following Mendelian inheritance in known pedigrees and can also be indirectly inferred when estimating inbreeding levels in relatively large samples. Despite the fact that few studies of microsatellite variation have been carried out in spruce, null alleles have been observed in both homologous and heterologous (i.e. developed on a different species) microsatellite loci (Rajora *et al.* 2001; Rungis *et al.* 2004; de Sousa *et al.* 2005; this study).

2. Attrition rate in red spruce microsatellite discovery

Most microsatellite primer publications are generally under the form of primer notes and do not involve the analysis of large data sets. Many loci may appear promising based on a few tested individuals but problems such as null alleles only become apparent in large-scale analyses. The proportion of published primer sets that proved to reliably amplify microsatellite loci in red spruce can be informative in this respect. Fourteen microsatellite loci were recently screened in 348 individuals for their utility in resolving paternity in Red spruce (Doerksen and Herbinger 2008). Most loci were selected based on a very large panel of markers evaluated by Rungis *et al.* (2004). They characterized 44 new microsatellite loci developed from EST sequences and screened all 101 previously published spruce markers. Forty two primers pairs including 25 new EST-based and 17 previously described microsatellites were identified as satisfactory, based on their ability to reliably amplify a single locus from white, Sitka, black, and red spruces (Rungis *et al.* 2004). The most promising loci (12) were selected among these 42 markers to be tested for paternity analysis with red spruce (Doerksen and Herbinger 2008). Finally, two additional loci (AC 136 and UAPgCA24) were also selected to be tested, based on the result of the present study. Considered jointly, the 14 loci under evaluation came from a pool of 138 published primer pairs. Ultimately, only 6 loci among the 14 tested were identified as being informative, reliably assayed, and exhibiting a low frequency of null alleles (Doerksen and Herbinger 2008). SPAG.C1 and SPAG.G3 had been developed from a Norway spruce genomic library (Pfeiffer *et al.* 1997) and UAPgCA24 from a white spruce genomic library (Hodgetts *et al.* 2001). WS007.J15 and WS0092.A19 were isolated from white/Sitka/interior spruce EST libraries (Rungis *et al.* 2004) and AC 136 was developed in the present study from a red spruce genomic library. This represents 6 satisfactory loci out of 138 published primers pair, which themselves came a much larger set of loci that were assayed in each of the original studies (Pfeiffer *et al.* 1997; Scotti *et al.* 2000; Hodgetts *et al.* 2001; Rajora *et al.* 2001; Scotti *et al.* 2002a, 2002b; Rungis *et al.* 2004; this study). The rate of attrition from sequences containing microsatellite repeats with appropriate flanking regions to reliable informative single locus markers appears

to be extremely high in this genus.

CONCLUSION

In summary, several loci were developed and can be useful for mapping studies. One new and one previously described marker were characterized by high variability, reliable PCR amplification and low frequency of null alleles. These 2 loci can be used successfully for paternity/parentage and population genetic analyses of red spruce. Further efforts at developing reliable microsatellite loci for the various spruce species appear warranted. These efforts should probably be directed at dinucleotide markers as they are more common and more polymorphic than trinucleotide and tetranucleotide loci. Searches for genomic sequence-derived microsatellite loci will likely continue and will require highly efficient enrichment techniques. The double enrichment method described here appears useful, but many primer pairs will have to be assayed to eliminate chimerical sequences and other artifacts. Searching the increasingly large EST databases generated by functional genomics projects (Rungis *et al*. 2004; Bérubé *et al*. 2007) looks like a promising way to find EST-derived microsatellites, which are presumably more transferable across species than genomic sequence-based markers. Identifying loci with appropriately low null allele frequency will be necessary and will require testing for proper Mendelian inheritance across a substantial number of families, or alternatively, assaying proper segregation of maternal alleles in megagametophytes (endosperm).

ACKNOWLEDGEMENTS

The authors thank Pat O'Reilly for much appreciated help with this work, which was supported by an AIF (ACOA) grant to C.M.H.

REFERENCES

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* **27**, 617-631

Blum BM (1990) *Picea rubens* Sarg. – Red spruce. In: Burns RM, Honkala BH (Eds) *Silvics of North America (Vol 1) Conifers*, USDA Forest Service Agricultural Handbook, pp 250-259

Bérubé Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K (2007) Characterization of EST-SSRs in loblolly pine and spruce. *Tree Genetics and Genomes* **3**, 251-259

Bobola MS, Eckert RT, Klein AS, Stapelfeldt K, Smith DE, Guenette D (1996) Using nuclear and organelle DNA markers to discriminate among *Picea rubens*, *Picea mariana* and their hybrids. *Canadian Journal of Forest Research* **26**, 433-434

Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. *Journal of Heredity* **99**, 616-623

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* **104**, 482-489

De Sousa SN, Finkeldey R, Gailing O (2005) Experimental verification of microsatellite null alleles in Norway spruce (*Picea abies* [L.] Karst.): Implications for population genetic studies. *Plant Molecular Biology Reporter* **23**, 113-119

Doerksen TK, Herbinger CM (2008) Male reproductive success and pedigree error in red spruce open-pollinated and polycross mating systems. *Canadian Journal of Forest Research* **38**, 1742-1749

Elsik CG, Minihan VT, Hall SE, Scarpa AM, Williams CG (2000) Low-copy microsatellite markers for *Pinus taeda* L. *Genome* **43**, 550-555

Elsik CG, Williams CG (2001a) Low-copy microsatellite recovery from a conifer genome *Theoretical and Applied Genetics* **103**, 1189-1195

Elsik CG, Williams CG (2001b) Families of clustered microsatellite in a conifer genome. *Molecular Genetic and Genomics* **265**, 535-542

Fowler DP (1986) Strategies for the genetic improvement of important tree species in the Maritimes. Information report M-X-156. Canadian Forestry Services. Fredericton, New Brunswick, 30 pp

Garner TWJ (2002) Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome* **45**, 212-215

Gordon AG (1976) The taxonomy and genetics of *Picea rubens* and its relationship to *Picea mariana. Canadian Journal of Botany* **54**, 781-813

Gordon AG (1996) The sweep of the boreal forest in time and space, from forest formation to genes, and implication for management. *Forest Chronicle* **72**, 19-30

Hamburg SP, Cogbill CV (1988) Historical decline of red spruce populations and climate warming. *Nature* **331**, 428-431

Hamilton MB, Pincus EL, Fiore DI, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* **27**, 500-507

Herbinger CM, Smith CA, Langy S (2006) Development and characterization of novel tetra- and dinucleotide microsatellite markers for the French Polynesia black-lipped pearl oyster, *Pinctada margaritifera. Molecular Ecology Notes* **6**, 107-109

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. *Theoretical and Applied Genetics* **102**, 1252-1258

Leak WB, Smith ML (1996) Sixty years of management and natural disturbance in a New England forested landscape. *Forest Ecology and Management* **81**, 63-73

Major JE, Mosseler A, Johnsen K, Rajora OP, Barsi DC, Kim K-H, Park J-M, Campbell M (2005) Reproductive barriers and hybridity in two spruces, *Picea rubens* and *Picea mariana*, sympatric in eastern North America. *Canadian Journal of Botany* **83**, 163-175

Manley SAM (1972) The occurrence of hybrid swarms of red and black spruce in central New Brunswick. *Canadian Journal of Forest Research* **2**, 381-391

Manley SAM, Ledig FT (1979) Photosynthesis in black and red spruce and their hybrid derivative: ecological isolation and hybrid adaptive inferiority. *Canadian Journal of Botany* **57**, 305-314

McPherson AA, O'Reilly PT, McParland TL, Jones MW, Bentzen P (2001) Isolation of nine novel tetranucleotide microsatellites in Atlantic herring (*Clupea harengus*). *Molecular Ecology Notes* **1**, 31-33

Mosseler A, Major JE, Simpson JD, Daigle B, Lange K, Park Y-S, Johnsen KH, Rajora OP (2000) Indicators of population viability in red spruce, *Picea rubens*. I. Reproductive traits and fecundity. *Canadian Journal of Botany* **78**, 928-940

O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2000) Isolation of twenty low stutter di- and tetranucleotide microsatellites for population analyses of walleye pollock and other gadoids. *Journal of Fish Biology* **56**, 1074-1086

O'Reilly PT, McPherson AA, Kenchington E, Taggart CT, Jones MW, Bentzen P (2002) Isolation and characterization of tetranucleotide microsatellites from Atlantic haddock (*Melanogrammus aeglefinus*). *Marine Biotechnology* **4**, 418-422

Perron M, Gordon AG, Bousquet J (1995) Species-specific RADP fingerprints for the closely related *Picea mariana* and *P. rubens*. *Theoretical and Applied Genetics* **91**, 142-149

Perron M, Perry DJ, Andalo C, Bousquet J (2000) Evidence from sequence-tagged-site markers of a recent progenitor-derivative species pair in conifers. *Proceedings of the National Academy of Sciences USA* **97**, 11331-11336

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

Rajora OP, Rahman MH, Dayanandan S, Mosseler A (2001) Isolation, characterization, inheritance and linkage of microsatellite markers in white spruce (*Picea glauca*) and their usefulness in other spruce species. *Molecular and General Genetics* **264**, 871-882

Roratto PA, Buchmann D, Santos S, Bartholemei-Santos ML (2008) PCR-mediated recombination in development of microsatellite markers: Mechanism and implications. *Genetics and Molecular Biology* **31**, 58-63

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, NJ, USA, pp 365-386

Rungis D, Bérubé Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* **109**, 1283-1294

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* **4**, 41-46

Scotti I, Magni F, Paglia GP, Morgante M (2002a) Trinucleotide microsatellites in Norway spruce (*Picea abies*): Their features and the development of molecular markers. *Theoretical and Applied Genetics* **106**, 40-50

Scotti I, Paglia GP, Magni F, Morgante M (2002b) Efficient development of dinucleotide microsatellite markers in Norway spruce (*Picea abies* Karst.) through dot-blot selection. *Theoretical and Applied Genetics* **104**, 1035-1041

Squirrell J, Hollingworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* **12**, 1339-1348

Van't Hof AE, Brakefield PM, Saccheri IJ, Zwaan BJ (2007) Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. *Heredity* **98**, 320-328

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: A review. *Molecular Ecology* **11**, 1-16

Zhou Y, Bui T, Auckland LD, Williams CG (2002) Undermethylated DNA as a source of microsatellites from a conifer genome. *Genome* **45**, 91-99