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

Christophe M. Herbinger1* • Kristian Gordos1,2 • Heather Allen1,3

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 hybrize 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 (RAPD) 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 under-methylation 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-
leotid 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 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 bands next to the real amplification product). They are much less prone to stuttering than dinucleotide markers, tetranucleotide markers and are much less prone to stuttering. This problem is thought to result from replication slippage or shadow bands next to the real amplification product.

We chose 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 bands next to the real amplification product). This problem is thought to result from replication slippage or shadow bands next to the real amplification product.

We chose 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 bands next to the real amplification product). This problem is thought to result from replication slippage or shadow bands next to the real amplification product.

We chose 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 bands next to the real amplification product). This problem is thought to result from replication slippage or shadow bands next to the real amplification product.

We chose 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 bands next to the real amplification product). This problem is thought to result from replication slippage or shadow bands next to the real amplification product.

We chose 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 bands next to the real amplification product). This problem is thought to result from replication slippage or shadow bands next to the real amplification product.

MATERIALS AND METHODS

Library enrichment protocols to increase microsatellite recovery

Method I: Single enrichment from untreated DNA

DNA was extracted from 80 mg of -20°C 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. This problem is thought to result from replication slippage or shadow bands next to the real amplification product. 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)n, (AAAG)n, (GATA)n, (GTAT)n, (GGAT)n and (GACA)n. The probe DNA was digested with HinII 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-amplication 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. The targeted DNA, the mixture was washed three times at 52°C with decreasing concentrations of sodium chloride sodium citrate (SSC): 2X, 1X, 0.5X SSC each with 0.5 ng/μL SNX Forward linker and 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-amplication, purified using QiAquick PCR purification kit (Qiagen), and cycle sequenced on a CEQ 8000 capillary sequencer (Beckman Coulter) using the DTC5 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 repeat: (GATA)n, (GTAT)n and (GGAT)n. Plasmid DNA was isolated from 288 random clones [96 for each of the (GATA)n, (GTAT)n and (GACA)n 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)n, (AAAG)n, (CATC)n and (GACA)n. 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 Skalicka 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 μg of frozen tissue was reduced to a fine powder in a Retch mill (Retch GmbH & Co.) and total DNA was obtained 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 (NH4)2SO4, 20 mmol Tris-HCl, 2 mmol MgSO4, 0.1% Triton X-100 (pH 8.8 at 25°C)) with 10-100 μg of template DNA, ~0.5 μg 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 alleles frequency was \( y + z/2x \). 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 to be polymorphic in 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\(_n\)) 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 MerBC-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 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.

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)
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 were inferred to have inherited from the female parent. The presence and inheritance of null allele(s) was inferred as described above for every easily scored locus except AC

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.

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

^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, hereafter designated RS 22a and RS 22b
^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

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

<table>
<thead>
<tr>
<th>Locus ID</th>
<th>Primer sequence</th>
<th>Repeat motif</th>
<th>Annealing temp. (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS 2</td>
<td>F-AGT TAG CCT GTG TGC CTT CC</td>
<td>(GTAT)</td>
<td>65-60 TD</td>
</tr>
<tr>
<td>RS 3</td>
<td>F-GAG ATA TGG TCA CAA ATG TTT TAC C</td>
<td>(GTAT)</td>
<td>67-62 TD</td>
</tr>
<tr>
<td>RS 7</td>
<td>F-GAA GCT TGT TCT ATG TGT TGT G</td>
<td>(GATA) (AC)</td>
<td>56</td>
</tr>
<tr>
<td>RS 12</td>
<td>F-TCA TGT TCT TCT TGT CCA</td>
<td>(GATA)</td>
<td>56</td>
</tr>
<tr>
<td>RS 22</td>
<td>F-CCT CAA AAG GCC ATC TTA ACC</td>
<td>(GATA)</td>
<td>67-62 TD</td>
</tr>
<tr>
<td>a&amp;b</td>
<td>F-CTT TGA AGA AGT TTT GTG GGA TG</td>
<td>perfect</td>
<td>67-62 TD</td>
</tr>
<tr>
<td>RS 24</td>
<td>F-TTT GAA CAT GGG ACC TCT GG</td>
<td>(GTAT)</td>
<td>67-62 TD</td>
</tr>
<tr>
<td>RS 27</td>
<td>F-GGA CAT TCC AAC GTA AAG AAG TTA ACC</td>
<td>perfect</td>
<td>56</td>
</tr>
<tr>
<td>RS 54</td>
<td>F-CTC ATT ATC ATC TCA TCA TCT TGT G</td>
<td>(TC) (AC)</td>
<td>60</td>
</tr>
<tr>
<td>RS 52</td>
<td>F-GGA TAT GGA AAC ATG GCA ATG</td>
<td>perfect</td>
<td>67-62 TD</td>
</tr>
<tr>
<td>AC 54</td>
<td>F-TGG AAA TGA AGA AAT GGG GGA</td>
<td>(AC)</td>
<td>56</td>
</tr>
<tr>
<td>AC 136</td>
<td>F-GTT CTG CTT GGT TGT TTC</td>
<td>perfect</td>
<td>56</td>
</tr>
<tr>
<td>Pgl 12</td>
<td>see Rajora et al. 2001</td>
<td>65-60 TD</td>
<td></td>
</tr>
<tr>
<td>UAPGCA24</td>
<td>see Hodgetts et al. 2001</td>
<td>67-62 TD</td>
<td></td>
</tr>
</tbody>
</table>

The Americas Journal of Plant Science and Biotechnology 5 (Special Issue 2), 105-111 ©2011 Global Science Books
recovered sequences that contained no true microsatellite hybridized to the probes. Close examination of many of the fragments with similar but not identical sequences instead beads only found a few targeted sequences, and other conifers could impact the success of library enrichment process and the potential for very efficient enrichment process and the potential for very efficient 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 protocols 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 chimeras 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-

**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

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.

![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.](image-url)
served during library enrichment (Squirrel et al. 2003; Roratto et al. 2008). In their attempt to enrich for trinucleo-
tide 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 hy-
ridization, 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 spe-
cies. These markers can be informative in a number of areas such as for example, genetic mapping, individual tree iden-
tification and paternity tracing in tree breeding programs or estimation of diversity and genetic distance among popula-
tions. In fact, several of these loci have already helped dis-
cover 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 iden-
ty of all the clones was verified with several loci deve-
loped here and one clone was recognized as being misiden-
tified. As a result, the group of pollen parents that have been used in the red spruce tree breeding program since its incep-
tion was shown to actually comprise 21 individual trees and not 20 as originally planned (Fowler 1986). This had impor-
tant consequences for the subsequent study of male repro-
ductive success and pedigree error in a polycross mating system derived from this pollen source (Doerksen and Her-
biner 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 mark-
ers developed in the present study were identified as being informative, reliably assayed, and exhibiting a low frequency of null alleles (Doerksen and Herbinger 2008). 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 on red spruce. Twenty-five loci were identified as being informative, reliably assayed, and exhibiting a low frequency of null alleles (Doerksen and Herbinger 2008). SPAGC1 and SPAGG3 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 WS092 A19 were isolated from white/Sitka/interior spruce libraries (Rungis et al. 2004). Finally, 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). SPAGC1 and SPAGG3 had been developed from a Norway spruce genomic library (Pfeiffer et al. 1997). SPAGC1 and SPAGG3 had been developed from a Norway spruce genomic library (Pfeiffer et al. 1997). SPAGC1 and SPAGG3 had been developed from a Norway spruce genomic library (Pfeiffer et al. 1997).
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 tetraneucleotide loci. Searches for genomic sequence-derived microsatellite loci will likely continue and will require highly efficient technologies and their applications in plant sciences. 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, assessing 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


111

