Theor Appl Genet (1999) 99: 37–49 

J. Germano · A. S. Klein

# Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana* and *P. rubens*

Received: 16 December 1998 / Accepted: 5 January 1999

**Abstract** *Picea rubens* (red spruce) and *P. mariana* (black spruce) are closely related species which are difficult to differentiate morphologically. They are sympatric with *P. glauca* (white spruce) in the northern portion of their ranges. In order to identify potential interspecific polymorphisms, the chloroplast *trn*K intron and *rpl*33-*psa*J-*trn*P region were sequenced, and the nuclear-encoded ITS region of the rDNA repeat was partially sequenced. Thirteen chloroplast and 12 nuclear candidate interspecific single nucleotide polymorphisms (SNPs) were identified. The species-specificity of several SNPs was determined by surveying DNAs amplified from trees representing range-wide provenance tests; these included 46 red spruce from 11 provenances, 84 black spruce from 30 provenances and 90 white spruce from 22 provenances. Two SNPs (1 chloroplast and 1 nuclear), which distinguish black spruce from red and white spruce, were consistent among 96–100% of the trees surveyed. Five SNPs (4 chloroplast and 1 nuclear), which distinguish white spruce from red and black spruce, were consistent among 100% of surveyed trees. These species-specific SNPs were used to identify anonymous spruce samples in a blind test, and their utility for small amounts of tissue, as little as single needles, was demonstrated. Scoring these SNPs is much less labor intensive than previous molecular methods for taxa differentiation (restriction fragment length polymorphisms or random amplified polymorphic DNAs), therefore they can be applied to large population studies.

**Key words** *Picea glauca* · *Picea mariana* · *Picea rubens* · Single nucleotide polymorphisms (SNPs) · Spruce

This is scientific contribution 2002 of the New Hampshire Agricultural Experiment Station, Durham, N.H.

Communicated by G. Wenzel

J. Germano · A. S. Klein (✉)
Department of Biochemistry and Molecular Biology, University of New Hampshire, 46 College Rd, Durham, NH 03824-2617, USA
Fax: + 1 603 862 4013
E-mail: anita.klein@unh.edu

## Introduction

Red spruce (*Picea rubens* Sarg.), black spruce [*P. mariana* (Mill.) B.S.P.] and white spruce [*P. glauca* (Moench) Voss] are sympatric species of northeastern North America. Red spruce's current range is from the Appalachians in North Carolina to the Maritimes including New England, New York, southern Quebec and restricted areas of Ontario (Morgenstern and Farrar 1964; Little 1971). The ranges of black and white spruce extend west to Alaska and north to the tree-line, overlapping with red spruce primarily in northern New England and eastern Canada (Morgenstern and Farrar 1964; Little 1971; Fowler et al. 1988). Although their ranges overelap, each of these species has a distinctive ecological niche (Morgenstern and Farrar 1964; Gordon 1976).

Morphological classification of the closely related red spruce and black spruce is difficult and controversial (Morgenstern and Farrar 1964; Manley 1971; Gordon 1976; Fowler et al. 1988); many morphological characters are too variable within one species to reliably distinguish it from the other (Gordon 1976; Fowler et al. 1988). Some traits are phenotopically plastic, and other characteristics are only discernable during limited time periods (Morgenstern and Farrar 1964; Gordon 1976; Donoghue and Sanderson 1992). Gordon (1976) conducted factor analysis of 24 morphological characters to distinguish red and black spruce. White spruce is more distantly related to red and black spruce (Gordon 1976; Sigurgeirsson and

Szmidt 1993) and can be easily identified with morphological characters.

Attempts have been made to distinguish red and black spruce with molecular markers including isozymes (Eckert 1989), random amplified polymorphic DNAs (RAPDs) (Perron et al. 1995) and restriction fragment length polymorphisms (RFLPs) (Bobola et al. 1992a, b, 1996). Bobola et al. used Southern blot analysis to identify polymorphisms between red and black spruce including five RFLPs of the nuclear ribosomal DNA (rDNA) repeat (1992a, b), two RFLPs in the chloroplast genome and three RFLPs in the mitochondrial genome (1996). None of these markers were 100% species-specific, however a three-character index, using the organelle markers together with one of the nuclear markers, reliably distinguished red and black spruce and their hybrids (Bobola et al. 1996). Although these markers are reliable, conventional RFLPs are tedious and costly both in their characterization and application.

Perron et al. (1995) identified seven RAPDs distinguishing red and black spruce in individual trees from six black spruce provenances and three red spruce provenances outside the sympatric zone. Four of these markers were 100% species-specific (i.e. present in all surveyed trees of one species and absent in all surveyed trees of the other species). A disadvantage of RAPDs is that this type of polymerase chain reaction (PCR™) is very sensitive to varying reaction conditions, temperature profiles and DNA quality (Ellsworth et al. 1993; Muralidharan and Wakeland 1993; Micheli et al. 1994; Vos et al. 1995). Hence, fingerprints may be difficult to reproduce, particularly in different laboratories.

Single nucleotide polymorphisms (SNPs) provide an alternative form of molecular markers for the discrimination of spruce species. Unlike conventional RFLPs and RAPDs, SNPs are direct markers; the exact nature and location of the allelic variations are known. Another advantage of SNPs is that large numbers of samples can be screened for a marker using a variety of inexpensive, high-throughput techniques (reviewed by Landegren et al. 1998). These methods may include restriction digestion (Landegren et al. 1998), allele-specific PCR (ASPCR) (Okayama et al. 1989; Sommer et al. 1989; Wu et al. 1989), or single-strand conformation polymorphism (SSCP) (Sheffield et al. 1993). Since each of these screening methods is PCR-based, only a small amount of template DNA is required, allowing samples to be identified from DNA extracted from just a few needles.

SNPs are likely to occur at higher frequencies in variable, less conserved genes. Few gymnosperm genes have been sequenced, therefore the potential variability of candidate genes in spruce must be gauged according to (1) their variability within angiosperm genera and/or (2) the predicted degree of variability based on functional constraints of the gene.

The approximately 10530-bp *mat*K gene, encoding an RNA maturase, is located within the approximately 2500-bp chloroplast *trn*K (UUU) intron (Sugita et al. 1985). *Mat*K exhibits a relatively low percentage of amino acid similarity (59% between tobacco and rice) compared to other chloroplast genes (Olmsted and Palmer 1994). It has been used to resolve phylogenetic relationships within families (Johnson and Soltis 1994, 1995; Steele and Vilgalys 1994; Li et al. 1999), and it displays some variation at the intrageneric level (Johnson and Soltis 1995; Li et al. 1997).

In gymnosperms, *mat*K has an average of 3.4 times more nucleotide differences per site than *rbc*L (Johnson and Soltis 1995). *Pinus* is the sister genus to *Picea* (Chase et al. 1993; Chaw et al. 1997). Hilu and Liang (1997) reported 1.1% *mat*K nucleotide variation between *Pinus contorta* (lodgepole pine) and *P. thunbergii* (black pine).

Noncoding regions such as introns or intergenic spacers (IGS) of the chloroplast are expected to be more variable than coding regions (Taberlet et al. 1991; Gielly and Taberlet 1994; Demesure et al. 1995; Perez de la Rosa et al. 1995). Chloroplast gene order is relatively conserved throghout land plants (Olmstead and Palmer 1994). Gene organization of *Picea* is similar to that of *Pinus* and *Pseudotsuga* (White et al. 1993). In addition, the entire chloroplast genome of black pine has been sequenced (Wakasugi et al. 1994). The conservation of chloroplast gene order, and the knowledge of a complete conifer chloroplast sequence present an opportunity to design new PCR primers for the amplification of potentially variable noncoding regions in spruce.

In red and black spruce, the chloroplast is paternally inherited, and the mitochondria are maternally inherited (Bobola et al. 1996). Nuclear interspecific markers would complement chloroplast markers since nuclear genes represent both the maternal and paternal lineage (Soltis et al. 1992). The nuclear ribosomal DNA repeat in plants contains the 18S, 5.8S and 26S genes separated by internal transcribed spacers 1 and 2 (ITS1 and ITS2) (reviewed by Hillis and Dixon 1991; Hamby and Zimmer 1992).

The ITS region (encompassing ITS1, 5.8S and ITS2) evolves at a much faster rate than the 18S and 26S genes (Hamby and Zimmer 1992; Baldwin et al. 1995). Baldwin et al. (1995) concluded that the ITS region is phylogenetically useful in angiosperms at intrafamilial levels, resolving relationships between genera, between species and even, to some extent, within species. Liston et al. (1999) used the 3′ ITS sequence (5.8S, ITS2 and approx. 200 bp of ITS1) to infer phylogenetic relationships of 47 species of *Pinus*. ITS1 was more divergent than 5.8S and ITS2, therefore it was hypothesized that additional ITS sequences would be useful in resolving relationships between more closely related *Pinus* species (Liston et al. 1999).

ITS1 is unusually large in some gymnosperms, making the ITS region up to 4.5 times as long as in angiosperms. For example, the ITS regions, including 5.8S and ITS2, in red and black spruce (Bobola et al. 1996) and *Picea abies* (Liston et al. 1996) are approximately 3.1 kb. The potential variability and large size of the ITS region make it a candidate gene in which to find polymorphisms between spruce species.

We have identified SNPs in nuclear and organelle genes which distinguish between white, red and black spruce. The species-specificity of each SNP was verified by screening samples representing range-wide provenance tests. The species-specific SNPs were used to identify anonymous spruce samples in a blind test, and their utility for small amounts of tissue was demonstrated.

## Materials and methods

### Plant materials

Some of the tissue samples used in this study were from independent collections. Black spruce sample 63 was provided by Dr. Peter Garrett of the Northeastern Forest Experiment Station of the US Forest Service. White spruce sample 494 was provided by Dr. Gerald Rehfeldt of the Intermountain Research Station, US Forest Service. White spruce sample 64 was collected from a tree on the University of New Hampshire campus. DNA was extracted from foliage using a standard CTAB method (Doyle and Doyle 1987).

Anonymous white, black and red spruce samples were provided by Dr. Robert Eckert, Department of Natural Resources, University of New Hampshire. A scaled down version of the CTAB method (Doyle and Doyle 1987), employing a 1.5-ml micro-centrifuge tube and mini-pestle to grind and material, was used to extract DNA from one to seven needles (9–16 mg).

DNAs from samples representing provenance tests were utilized for population studies. Red spruce DNAs represent a range-wide provenance test located in Coleman State Forest, Stewartstown, New Hampshire (see Bobola et al. 1996). Black spruce DNAs represent a provenance test corresponding to the eastern half of the black spruce range. This site is maintained by the USDA Forest Service (Northeastern Forest Experiment Station) in the Massabesic Experimental Forest, Alfred, Maine (see Bobola et al. 1996). White spruce DNAs representing a range-wide provenance test at Grand Rapids, Minnesota was provided by Dr. Glenn Furnier, Departments of Forest Resources and Plant Biology, University of Minnesota (see Furnier et al. 1991; Furnier and Stine 1995).

### PCR methods

*Taq* DNA Polymerase (Promega, Madison, Wis.) was used in all reactions; *Taq* Extender™ PCR Additive (Stratagene, La Jolla, Calif.) was added to reactions to amplify fragments longer than 2 kb. Reactions contained 1 × Magnesium Free Reaction Buffer B (Promega) or 1 × *Taq* Extender Reaction Buffer (Stratagene), 0.2 m*M* each dNTP (Promega), 0.4 μ*M* each primer (Fig. 1, Table 1) and 5–10 ng/μl whole genomic DNA. All amplifications were carried out in an MJ Research PTC-100 Programmable Thermal Controller. The denaturation temperature was 94°C, and the extension temperature was 72°C. Each profile had an initial denaturation step of 3 min and a final extension step of 10 min. The *trn*K intron, the *rpl*33-*psa*J-*trn*P region and the ITS region (Fig. 1) were amplified as indicated in Table 2. Amplification of the ITS region was preceded by a hot start (D'Aquila et al. 1991). Primers rpl33 and trnP (Fig. 1b, Table 1) were designed from a black pine chloroplast sequence (GenBank D17510; Wakasugi et al. 1994). Additional amplification and sequencing primers (Fig. 1, Table 1) were designed using the PrimerSelect algorithm, part of the Lasergene software package (DNASTAR, version 3.72, Madison, Wis.).

PCR products were purified via electrophoresis through low-melting-point agarose (Gibco BRL, Gaithersburg, Md.). Bands were excised and liquified by adding 5 U Agarase (Sigma®, St. Louis, Mo.) per 100 μl gel and incubating at 37°C for 1 h.

### Cloning

Gel-purified ITS PCR product (approx. 3.1-kb fragment) was digested with *Sau*3AI (New England Biolabs (NEB), Beverly, Mass. or Stratagene) according to the manufacturers' specifications. The resulting fragments, ranging from 200 bp to 900 bp, were gel-purified as described above, precipitated and resuspended in water. pGEM®-3Z vector (Promega) was digested with *Bam*HI (Promega), dephosphorylated with Calf Intestinal Alkaline Phosphatase (Promega) and ligated with ITS fragments using T4DNA ligase (Promega). The recombinant plasmids were used to transform Epicurian Coli® XL1-Blue MRF' supercompetent cells (Stratagene) according to the manufacturer's directions. Transformants were grown up on LB medium containing 50 μg/ml ampicillin. Plasmids were isolated using Wizard® Minipreps kit (Promega).

### DNA sequencing

The ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq® DNA Polymerase, FS (Perkin Elmer, Emeryville, Calif.) was used. Plasmids were sequenced in separate reactions (20 μl) containing 3.2 pmol pUC/M13 forward or reverse primers (Promega), 250–500 ng DNA and 8 μl Ready Reaction Mix. PCR-amplified fragments were sequenced using 10 pmol primer (Fig. 1, Table 1) and 30–90 ng template. Cycle Sequencing was carried out, and extension products were purified using Ethanol Precipitation Protocol 1 as described by the manufacturer. Extension products were separated and analyzed on an ABI PRISM 373 Automated Sequencer (UNH Sequencing Facility). Sequences were generated with ABI DNA Sequencing Software version 2.1.1, Base caller ABI50, and edited by eye using ABI SeqEd Software version 1.0.3.

Both strands of DNA were sequenced for *mat*K and the *rpl*33-*psa*J-*trn*P region. One strand of DNA was sequenced for the ITS region and for the ITSI cloned fragment.

### Sequence analysis

Contiguous sequences were assembled and aligned using SeqMan II and MegAlign (DNASTAR) algorithms. Percentage divergence between sequences was calculated with MegAlign (DNASTAR). Restriction maps of the sequences were generated using MapDraw (DNASTAR).

### Population studies

In order to assess the distribution of SNPs across populations, we used several molecular screening methods. Restriction degests using *Dra*I, *Ssp*I, *Bst*UI, *Bsp*1286I (NEB) and *Msp*I (or isoschizomer *Hpa*II,

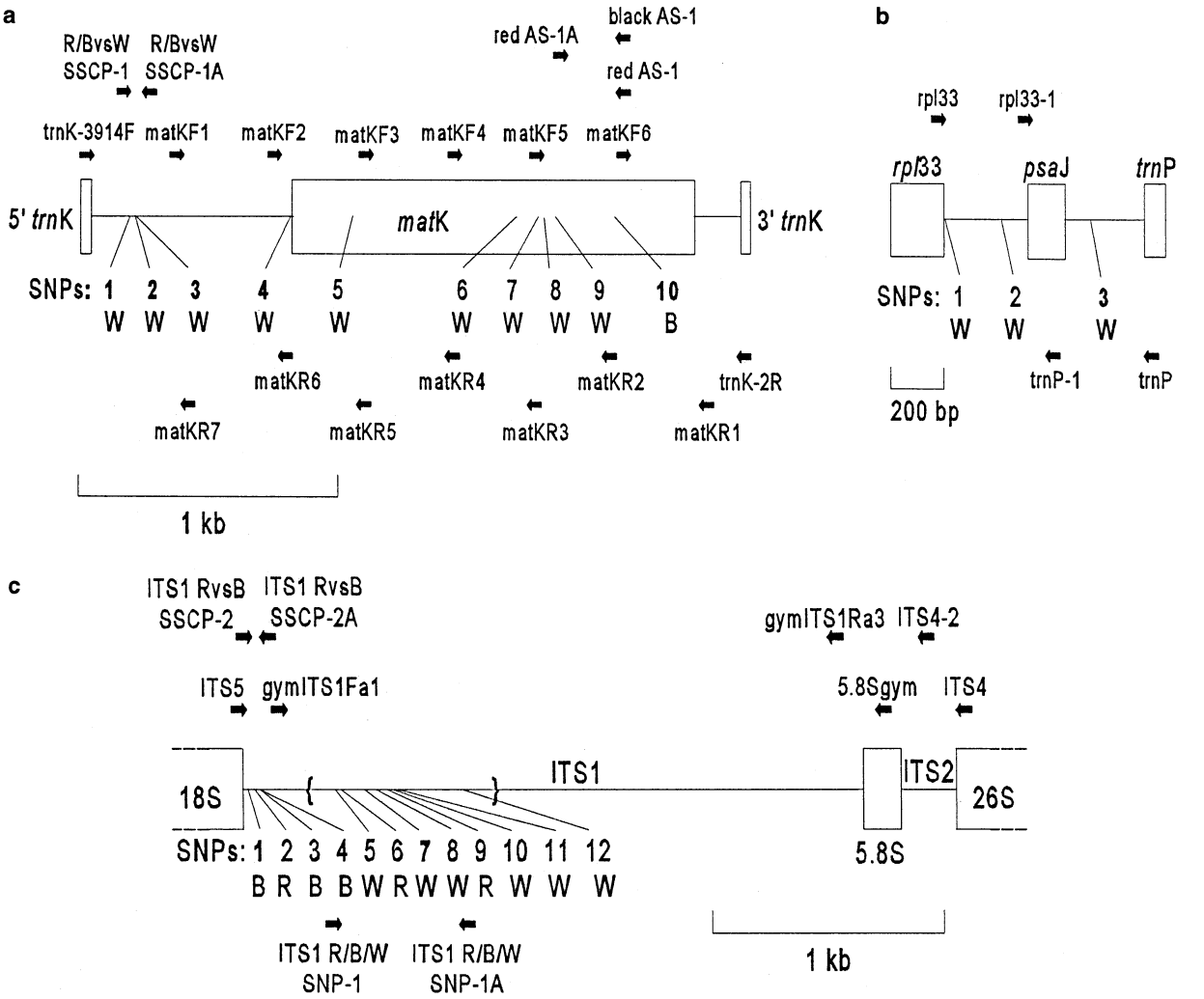


**Fig. 1a–c** Gene regions showing the relative positions of coding regions (*boxed areas*) and PCR amplification and sequencing primers. *Arrows* point in the 5′ → 3′ direction of primers. The relative positions of interspecific SNPs are shown. *Letters below* SNPs indicate the species that each SNP distinguishes: *W* distinguishes white spruce from red and black spruce, *B* distinguishes black spruce from white and red spruce, *R* distinguishes red spruce from white and black spruce. **a** *trn*K intron, **b** *rpl*33-*psa*J-*trn*P region, **c** ITS region. *Brackets* indicate the cloned 778-bp fragment of ITS1

Promega) were carried out according to the manufacturers' specifications. Resulting fragments were separated on 1.4–3% agarose; fragments smaller than 200 bp were separated on 3–4% MetaPhor® agarose (FMC BioProducts, Rockland, Me.). Single-strand conformation polymorphism (SSCP) analysis was carried out according to the protocol for MDE™ Gel (FMC). Two microliters (3–6 ng) of PCR product (119-bp fragment) was mixed with 10 µl stop solution, and the entire denatured sample was loaded onto a 1 × gel. The gel was stained in 1 µg/ml ethidium bromide for 15 min and visualized under UV light. Allele-specific PCR (ASPCR) was carried out using primers which differentially amplify specific alleles (Tables 1 and 2). All reactions contained an additional primer set as a positive control (see Wu et al. 1989). PCR products were separated by electrophoresis through 1.4% agarose.

## Results

### Chloroplast gene sequences

In order to identify potential interspecific polymorphisms, we sequenced the chloroplast *trn*K intron and the *rpl*33-*psa*J-*trn*P region from 4 individual trees representing geographically distant populations of each species (Table 3) as suggested by Baverstock and Moritz (1996). There was no intraspecific variation among these individuals in the *trn*K intron. Intraspecific variation was limited in the *rpl*33-*psa*J-*trn*P region, consisting of three *rpl*33-*psa*J IGS nucleotide substitutions among 2 of the 4 back spruce individuals, and one *psa*J-*trn*P IGS nucleotide substitution in 1 of the 4 white spruce individuals.

### Screening interspecific *trn*K SNPs for species-specificity

Ten candidate interspecific single nucleotide polymorphisms (SNPs) were identified in the *trn*K intron: 4 in

**Table 1** PCR amplification and sequencing primers (*Seq* sequencing)

| Primer name | PCR/Seq[a] | 5′ Sequence 3′ |
|---|---|---|
| trnK-3914F[b] | PCR and seq | TGG GTT GCT AAC TCA ATG G |
| trnK-2R[b] | PCR and seq | AAC TAG TCG GAT GGA GTA G |
| matKF1 | PCR and seq | TCG AAT GAG TCA ATG GAG AAA G |
| matKF2 | Seq | CGC ACC ATG TAT TGT ATT ATC TCA |
| matKF3 | Seq | TGA ATC GGT ACT AGA AGG ACT GAC |
| matKF4 | Seq | AGA TTC TTC CTG TTC CTG TGG |
| matKF5 | Seq | CCA TCT TTG GAA CGA ACC |
| matKF6 | Seq | CCT TTG GTC GAG ATG GTT TAT |
| matKR1 | PCR and seq | CTC GGA TGG CAA AAT AAA AAT G |
| matKR2 | Seq | GGA TCC ACT GTA GTA ATG AAA AAT |
| matKR3 | Seq | CTA TAA GGT TCG TTC CAA AGA TG |
| matKR4 | Seq | CCA CAG GAA CAG GAA GAA TC |
| matKR5 | Seq | TCA GTC CTT CTA GTA CCG ATT CAG |
| matKR6 | Seq | CTC AGT TAT GGC CCT CGT TC |
| matKR7 | Seq | GGG AAT TCC TCG CTC GTT |
| R/BvsW SSCP-1 | PCR | ACA GCA TGT CGT TCC AAC A |
| R/BvsW SSCP-1A | PCR | ATA GAC ATT TCC CAC CCA TTT |
| red AS-1 | PCR | GGA TCC ACT GTA GTA ATG AAA AAG |
| black AS-1 | PCR | GGA TCC ACT GTA GTA ATG AAA AAT |
| red AS-1A | PCR | TTG GTC AAG ACT AAA ATG CTA GA |
| rpl33 | PCR and seq | TGT TAC TCT TAT ATC TCC GCT CTT |
| trnP | PCR and seq | CAA AAC AAA CAC GCT ACC AA |
| rpl33-1 | Seq | TGA CTT TAA GGG GAG GAC AAC |
| trnP-1 | Seq | CTG TGT TAG CTA TTT CAT CGT TCA |
| ITS5[c] | PCR and seq | GGA AGT AAA AGT CGT AAC AAG G |
| ITS4[c] | PCR and seq | TCC TCC GCT TAT TGA TAT GC |
| gymITS1Fa1 | Seq | TGT TGT CCT TGG CCT CCT |
| ITS4-2 | Seq | GAC AAT ATC ACC GCT CGC C |
| 5.8Sgym | Seq | GAT GAT TCA CGG GAT TCT G |
| gymITS1Ra3 | Seq | CCA CAA GAC ATA TGC ACT C |
| ITS1 RvsB SSCP-2 | PCR and seq | TGC GGT AGG ATC ATT GTC AGT |
| ITS1 RvsB SSCP-2A | PCR and seq | CGA TCA ACC CTC CAA AAG TG |
| ITS1 R/B/W SNP-1 | PCR and seq | CTT CGT TTG AGT CTT TGT TTT TCG |
| ITS1 R/B/W SNP-1A | PCR and seq | GGG CCA CCG GAG CAT TG |

[a] Indicates whether a primer was used for PCR, sequencing, or both
[b] Johnson and Soltis (1994)
[c] White et al. (1990)

the 5′ noncoding region and 6 in *mat*K (Fig. 1a, Table 4). Among these candidate SNPs were eight transversions and two transitions; 4 of the 6 SNPs in *mat*K are non-synonymous. One of the SNPs (trnK SNP 10) distinguishes black spruce, while the remaining 9 distinguish white spruce. TrnK SNP 10 is located at the first position of codon 410 in *mat*K, encoding isoleucine in black spruce and leucine in red and white spruce.

Five of the *trn*K SNPs were tested for species-specificity by screening for the presence or absence of the particular nucleotides (Table 4) in DNAs amplified from trees representing range-wide provenance tests. A total of 46 red spruce (3–5 from each of 11 provenances), 84 black spruce (1–4 from 30 provenances plus 1 separate sample) and 90 white spruce (4 from 22 provenances plus 2 separate samples) samples were screened. Several diagnostic methods were used, including single-strand conformation polymorphism (SSCP), restriction analysis and allele-specific PCR (ASPCR).

SSCP was used to screen trnK SNPs 1, 2 and 3 (Fig. 1a) simultaneously. At these positions, the sequence of the 4 white spruce individuals had nucleotides G, A and T, whereas the red and black spruce had T, T and G (Table 4). TrnK SNP 2 and 3 are adjacent, and SNP 1 is located 19 bp upstream. The primers R/BvsW SSCP-1 and R/BvsW SSCP-1A (Fig. 1a, Tables 1 and 2) were designed to amplify a 119-bp fragment in which trnK SNPs 1, 2 and 3 are centrally located.

Upon denaturation of the amplified fragments and separation on 1 × MDE (FMC), the white spruce single-stranded DNAs travelled at different rates than those of red and black spruce. This difference in electrophoretic mobility resulted in a distinct banding pattern for white spruce and a different pattern for red and black spruce (Fig. 2a). All of the provenance test DNAs were surveyed for these banding patterns: 100% (90/90) white spruce displayed the "close" pattern, and 100% (46/46) red and 100% (84/84) black spruce displayed the "wide" pattern (Fig. 2a, Table 4).

**Table 2** PCR profiles

| Gene or region | Primers | Fragment size | Reaction volume (µl) | *Taq* units | *Taq* extender units | $MgCl_2$ (m*M*) | Annealing temperature (°C) | Annealing time | Extension time | Cycles |
|---|---|---|---|---|---|---|---|---|---|---|
| *trn*K intron | trnK-3914F & trnK-2R[a] | ~ 2.5 kb | 50 | 2 | 2 | | 48 | 1 min, 30 s | 2 min, 30 s | 30 |
| *rpl*33-*trn*P | rpl33 & trnP | 809 bp | 50 | 2 | | 2.5 | 47 | 1 min, 30 s | 1 min | 30 |
| ITS | ITS5 & ITS4[b] | ~ 3.1 kb | 50 | 4 | 4 | | 45 | 1 min, 45 s | 3 min | 29 |
| *trn*K intron | matKF1 & matKR1 | 2072 bp | 50 | 2 | 2 | | 51 | 1 min | 2 min, 15 s | 30 |
| ITS1 | ITS1 RvsB SSCP-2<br>ITS1 RvsB SSCP-2A | 121 bp | 25 | 0.5 | | 2.5 | 50 | 1 min | 30 s | 30 |
| ITS1 | ITS1 R/B/W SNP-1<br>ITS1 R/B/W SNP-1A | 609 bp | 25 | 1 | | 1 | 51 | 1 min | 45 s | 30 |
| *trn*K intron | R/BvsW SSCP-1<br>R/BvsW SSCP-1A | 119 bp | 25 | 0.5 | | 2.5 | 47 | 1 min | 30 s | 30 |
| *mat*K | red AS-1 & red AS-1A<br>R/BvsW SSCP-1<br>R/BvsW SSCP-1A | 231 bp<br>&<br>119 bp | 25 | 1 | | 1 | 47 | 1 min | 20 s | 28 |
| *mat*K | black AS-1 & red AS-1A<br>R/BvsW SSCP-1<br>R/BvsW SSCP-1A | 231 bp<br>&<br>119 bp | 25 | 1 | | 1 | 47 | 1 min | 20 s | 28 |

[a] Johnson and Soltis (1994)
[b] White et al. (1990)

The restriction enzyme *Dra*I was used to screen trnK SNP 4 (Fig. 1a), which distinguishes white spruce from red and black spruce. At this position, the white spruce sequences have a G, whereas red and black spruce have a T (Table 4), part of a *Dra*I restriction site (TTTAAA). The 2072-bp fragment, amplified with primers matKF1 and matKR1 (Fig. 1a, Tables 1 and 2), also contains an invariant *Dra*I site. This restriction site acts as a positive control by generating an invariant 229-bp fragment. Therefore, *Dra*I was predicted to produce two fragments (229 bp and 1843 bp) for white spruce and three fragments (229 bp, 466 bp and 1377 bp) for red and black spruce. One hundred percent (90/90) of white spruce, 96% (81/84) of black spruce and 85% (39/46) of red spruce provenance test sample DNAs displayed the predicted fragment size patterns following *Dra*I digestion (Table 4).

The restriction enzyme *Ssp*I was used to screen trnK SNP 10 (Fig. 1a), which distinguishes black spruce from white and red spruce. At this position, the red and white spruce sequences have a C, and black spruce has an A (Table 4), part of a *Ssp*I restriction site (AATATT). The 2072-bp fragment, amplified with primers matKF1 and matKR1 (Fig. 1a), also contains three invariant *Ssp*I sites which generate three fragments (52 bp, 358 bp and 812 bp). Therefore, *Ssp*I was predicted to produce five fragments (52 bp, 358 bp, 376 bp, 474 bp and 812 bp) for black spruce and only four fragments (52 bp, 358 bp, 812 bp and 850 bp) for red and white spruce. *Ssp*I restriction analysis was used to screen a subset of the white spruce provenance test samples; 100% (43/43) of them displayed the predicted fragment size pattern (Table 4).

ASPCR was used to screen trnK SNP 10 (Fig. 1a) in the red and black spruce provenance test DNAs and those white spruce DNAs not screened with *Ssp*I. Allele-specific primers were designed whose 3′ nucleotide anneals at the position of trnK SNP 10. Primer black AS-1 (Fig. 1a) contains a 3′ T (Table 1), complementary to the A (Table 4) at this site in black spruce. Primer red AS-1 (Fig. 1a) contains a 3′ G (Table 1) that is complementary to the C (Table 4) at this site is red and white spruce. The generic upstream primer red AS-1A (Fig. 1a, Table 1) pairs with either allele-specific primer to amplify a 231-bp fragment (Table 2). An additional pair of primers, R/BvsW SSCP-1 and R/BvsW SSCP-1A (Fig. 1a, Tables 1 and 2), was added to every PCR to amplify a 119-bp fragment as a positive control on the success of the reaction (see Wu et al. 1989). Two PCRs were done for each individual sample: one containing primer red AS-1, the other containing primer black AS-1, and both containing the generic primer and the control primer pair (Table 2). Scoring amplification of the 231-bp allele-specific fragment was carried out only if the 119-bp positive control fragment successfully amplified in both reactions (Fig. 2b). In 100% (48/48) of white and 98% (45/46) of red spruce samples, the 231-bp fragment

**Table 3** Samples used for sequencing

| Species | Sample | Geographic location | GenBank accession | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *trn*K intron | *rpl*33-*trn*P | ITS1[a] | ITS1[b] | ITS1[c] | ITS[d] |
| White spruce | 64 | Durham, NH | AF133923 | AF133935 | AF117916 | AF140755 | AF119377-AF119379 | AF136610 |
| | 494 | Black Hills, SD | AF133924 | AF133936 | AF117917 | AF140756 | | AF136611 |
| | S-A | Saskatchewan | AF133925 | AF133937 | AF117918 | AF140757 | | AF136612 |
| | V-A or V-B | Alaska | AF133926 | AF133938 | AF117919 | AF140758 | | AF136613 |
| Black spruce | 63 | Durham, NH | AF133919 | AF133931 | AF117912 | AF140751 | AF119374-AF119376 | AF136614 |
| | 4274-1 | New Hampshire | AF133920 | AF133932 | AF117913 | AF140752 | | AF136615 |
| | 4962-1 | Newfoundland | AF133921 | AF133933 | AF117914 | AF140753 | | AF136616 |
| | 5004-1 | New Brunswick | AF133922 | AF133934 | AF117915 | AF140754 | | AF136617 |
| Red spruce | 2019 21-1 | Indian Gap, NC | AF133915 | AF133927 | AF117908 | AF140747 | AF119371-AF119373 | AF136618 |
| | 2027 29-4 | Pillsbury, NH | AF133916 | AF133928 | AF117909 | AF140748 | | AF136619 |
| | 2032 8-1 | Valcartier, Quebec | AF133917 | AF133929 | AF117910 | AF140749 | | AF136620 |
| | 2505 38-5 | Acadia Forest Experiment Station, NB | AF133918 | AF133930 | AF117911 | AF140750 | | AF136621 |

[a] Sequence of PCR product amplified with primers ITS1 RvsB SSCP-2 and ITS1 RvsB SSCP-2A
[b] Sequence of PCR product amplified with primers ITS1 R/B/W SNP-1 and ITS1 R/B/W SNP-1A
[c] Sequence of cloned fragment of PCR product amplified with primers ITS5 and ITS4
[d] Includes partial sequence of ITS1 (3′ end) and complete 5.8S and ITS2 sequences

amplified with the red AS-1 primer but not with the black AS-1 primer. In 98% (82/84) of the black spruce samples, the 231-bp fragment amplified with the black AS-1 primer but not with the red AS-1 primer (Fig. 2b, Table 4).

## Screening interspecific *rpl*33-*psa*J-*trn*P SNPs for species-specificity

Three candidate interspecific SNPs were identified in the *rpl*33-*psa*J-*trn*P region: 2 in the *rpl*33-*psa*J IGS and 1 in the *psa*J-*trn*P IGS (Fig. 1b). All 3 SNPs in this region are transitions, and all distinguish white spruce from red and black spruce.

The restriction enzyme *Msp*I was used to screen rpl33-trnP SNP 3 (Fig. 1b). At this position, the white spruce sequences have an A, whereas red and black spruce have a G (Table 4), part of a *Msp*I restriction site (CC**G**G). The 809-bp fragment, amplified with primers rpl33 and trnP (Fig. 1b, Tables 1 and 2), also contains an invariant *Msp*I site, which generates a 407-bp fragment. Therefore, *Msp*I was predicted to generate two nearly equal-sized fragments (407 bp and 402 bp, seen as one bright band on an agarose gel) for white spruce and three fragments (407 bp, 227 bp and 175 bp) for red and black spruce (Fig. 2c). One hundred percent (90/90) of white spruce, 100% (84/84) of black spruce and 100% (46/46) of red spruce provenance test sample DNAs displayed the predicted fragment size patterns following *Msp*I digestion (Table 4).

## ITS sequences

An approximately 3.1-kb fragment encompassing the ITS region (ITS1, 5.8S and ITS2) was amplified and partially sequenced (Fig. 1c, Tables 1 and 2). Sequences of 4 individual trees of each species (Table 3) were obtained, including complete sequences for 5.8S and ITS2 plus 300–550 bp from the 5′ end of ITS1 and 300–650 bp from the 3′ end of ITS1 (Fig. 1c). The 5.8S gene is 162 bp and ITS2 is 236 bp in each of the species. The endpoints of spruce 5.8S and ITS2 were determined by comparison to the ITS region sequence of *Pinus pinea* (stone pine) (GenBank X87936; Marrocco et al. 1996).

The rDNA repeat is present in multiple copies in the nuclear genome; it is assumed that the PCR-amplified DNAs represent ratios similar to the genomic copies. The sequences were obtained using PCR product as template, and therefore they represent the main fraction of the template DNAs. The sequence chromatograms, especially for ITS1, contained a significant amount of background; in many locations, there were two distinct peaks of two different nucleotides, one on top of the other. This result can be explained by heterogeneity of the template DNAs. This background caused the sequence signals to deteriorate rapidly, sometimes limiting chromatogram analysis to only 150–200 bases per reaction.

Primers gymITS1Fa1 and gymITS1Ra3, which anneal to opposite ends of ITS1 (Fig. 1c), yielded limited sequence from PCR product as template. One explanation for this sequencing difficulty is the heterogeneity of

**Table 4** Results of population studies

| Gene region | *trn*K | | *trn*K | | *trn*K | | *rpl*33-*trn*P | | ITS | | ITS | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Location | 5′ *trn*K intron | | 5′ *trn*K intron | | *mat*K | | *psa*J-*trn*P IGS | | ITS1 | | ITS1 | |
| Species | SNPs 1, 2 & 3 | SSCP analysis | SNP 4 | *Dra*I analysis | SNP 10 | ASPCR or *Ssp*I analysis | SNP 3 | *Msp*I analysis | SNP 1 | *Bst*UI analysis | SNP 7 | *Bsp*1286I analysis |
| White | G AT[a] | 90/90[b] (100%) | G | 90/90 (100%) | C | 90/90 (100%) | A | 90/90 (100%) | G | 89/90 (99%) | C | 89/89[c] (100%) |
| Black | T TG | 84/84 (100%) | T | 81/84 (96%) | A | 82/84 (98%) | G | 84/84 (100%) | A | 84/84 (100%) | G | 84/84 (100%) |
| Red | T TG | 46/46 (100%) | T | 39/46 (85%) | C | 45/46 (98%) | G | 46/46 (100%) | G | 44/46 (96%) | G | 46/46 (100%) |

[a] Nucleotide positions of trnK SNP 2 and trnK SNP 3 are adjacent
[b] Proportion of provenance test samples whose DNAs scored positive for the corresponding SNP allele
[c] This region would not amplify with the DNA from one white spruce individual

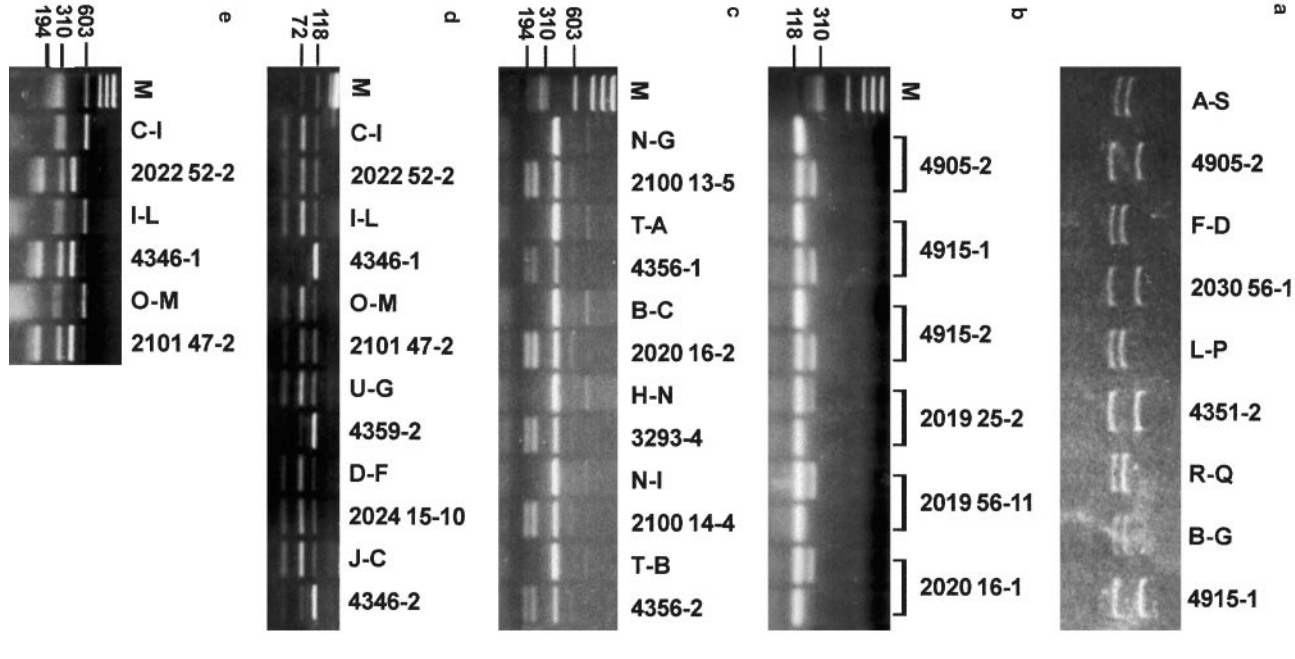


**Fig. 2a–e** Representative results from screening provenance test samples. *Alphabetic labels* represent white spruce samples, *Numeric labels* beginning with *2* represent red spruce samples and *numeric labels* beginning with *3 or 4* represent black spruce samples. The marker (*M*) is $\phi$X 174DNA/*Hae*III with the sizes of some fragments labeled in base pairs. **a** SSCP analysis of trnK SNPs 1, 2 and 3. **b** ASPCR analysis of trnK SNP 10. In the *left lane* for each sample are PCR products from primer red AS-1 plus the positive control primer pair. In the *right lane* for each sample are PCR products from primer black AS-1 plus the positive control primer pair. The 119-bp band (present in every lane) is the positive control PCR product. The 231-bp band is the product of the allele-specific primers. **c** *Msp*I analysis of rpl33-trnP SNP 3. **d** *Bst*UI analysis of ITS SNP 1. **e** *Bsp* 1286I analysis of ITS SNP 7

the rDNA copies. Another possible explanation is the presence of subrepeats within ITS such as those reported in stone pine by Marrocco et al. (1996) and *Pseudotsuga* and *Larix* by Gernandt and Liston (1999). Preliminary results from sequencing multiple cloned ITS *Sau*3AI fragments of different lengths suggest that subrepeats also exist in red, black and white spruce ITS1 (data not shown). During cycle-sequencing using PCR product as template, if a primer anneals to subrepeats, then it would anneal to multiple locations of the template DNA. This would cause multiple sequences to be represented in the chromatograms, making the data increasingly difficult to analyze as the subrepeat sequences diverge.

There were no interspecific differences between white, black and red spruce in the 5.8S gene and ITS2, and in the final 300 bp of the 3′ end of ITS1. The 5′ end of ITS1 is more variable than its 3′ end. Four candidate interspecific SNPs were identified in the initial 100 bp from the 5′ end of ITS1 (Fig. 1c). They are all transitions, including three that distinguish black spruce and one that distinguishes red spruce.

## Screening interspecific ITS SNP 1 for species-specificity

The restriction enzyme *Bst*U1 was used to screen ITS SNP 1, which distinguishes black spruce from red and white spruce. At this position, the black spruce sequences have an A, whereas red and white spruce have a G (Table 4), part of a *Bst*U1 restriction site (C**G**CG). This is the only *Bst*U1 site in the 121-bp fragment amplified with primers ITS1 RvsB SSCP-2 and ITS1 RvsB SSCP-2A (Fig. 1c, Tables 1 and 2). Therefore, *Bst*U1 was predicted to generate two fragments (44 bp and 77 bp) for red and white spruce but not to cleave the 121-bp fragment amplified from black spruce. When the 121-bp fragment amplified from DNAs of provenance test trees was digested with *Bst*U1 however, a fraction (0–20%) of the PCR product was cleaved in many of the back spruce samples, and a fraction (10–30%) of the PCR product was not cleaved in the red and white spruce samples (Fig. 2d). These results suggest that the copies of the rDNA repeat within an individual are heterogeneous for this *Bst*U1 site. The observed heterogeneity was considered for individual amplified DNAs that were scored for ITS SNP 1. The majority ($\geqslant$ 70%) of the PCR product was cleaved into two fragments in 99% (89/90) of white and 96% (44/46) of red spruce provenance test sample DNAs, and the majority ($\geqslant$ 80%) of the PCR product was not cleaved in 100% (84/84) of the black spruce provenance test sample DNAs (Fig. 2d, Table 4).

## Sequences of cloned ITS fragments

In order to extend the sequence of ITS1, the approximately 3.1-kb PCR product encompassing ITS1, 5.8S and ITS2 was digested with *Sau*3AI. Some of the *Sau*3AI fragments were cloned and sequenced from 1 individual of each species (Table 3). The longest clone, 778 bp, is contiguous with the 5′ ITS1 sequence obtained with primers ITS5 and gymITS1Fa1 (Fig. 1c). Three 778-bp clones from each individual were sequenced. Alignment of these sequences revealed a significant amount of intra-individual variation, including single nucleotide transitions, transversions and insertion/deletions. There were 29 SNPs among the three clones from white spruce sample 64, 12 SNPs among the three clones from black spruce sample 63, and 16 SNPs among the three clones from red spruce sample 2019 21-1. These intra-individual polymorphisms further support the hypothesis that there is considerable heterogeneity among copies of ITS1.

Eight candidate interspecific SNPs were identified in the 778-bp *Sau*3AI clone, including three transversions, four transitions and one insertion/deletion (Fig. 1c). Two of the SNPs distinguish the red spruce sample, while the other 6 distinguish the white spruce sample. In order to confirm these candidate SNPs we sequenced a 609-bp fragment (amplified with primers ITS1 R/B/W SNP-1 and ITS1 R/B/W SNP-1A) from 4 individuals of each species (Table 3) using PCR product as template. This sequence confirmed the cloned fragment sequences at all positions for ITS SNPs 6-12.

## Screening interspecific ITS SNP 7 for species-specificity

The restriction enzyme *Bsp*1286I was used to screen ITS SNP 7 (Fig. 1c), which distinguishes white spruce from red and black spruce. At this position, the white spruce sequences have a C, whereas the red and black spruce sequences have a G (Table 4), part of a *Bsp*1286I restriction site (GD**G**CHC). The 609-bp fragment, amplified with primers ITS1 R/B/W SNP-1 and ITS1 R/B/W SNP-1A (Fig. 1c, Tables 1 and 2), also contains another *Bsp*1286I site, generating a 322-bp positive control fragment. Therefore, *Bsp*1286I was predicted to generate three fragments (322 bp, 168 bp and 119 bp) from red and black spruce amplified DNAs and only two fragments (322 bp and 287 bp) from white spruce DNAs. Interestingly, the positive control *Bsp*1286I site coincides with an intra-individual polymorphic site, as evidenced by the alignment of cloned 778-bp fragment sequences from 1 individual. Hence, cleavage at the positive control site did not occur in a fraction (up to 35%) of the amplified DNAs from each individual sample. Therefore, some 609-bp fragments are observed in digests of white spruce DNAs, and some 441-bp fragments are observed in digests of red and black spruce DNAs (Fig. 2e). The nucleotide at the ITS SNP 7 position, however, appeared to be homogeneous within each individual. One hundred percent (89/89) of white spruce provenance test sample DNAs were not cleaved

at the ITS SNP 7 site; 100% (84/84) of black spruce and 100% (46/46) of red spruce sample DNAs were cleaved at the ITS SNP 7 *Bsp*1286I site (Table 4). Data for all of the SNPs listed in Table 4 from screening provenance test samples in population studies were analyzed with a $\chi^2$ test of independence ($P < 0.0001$).

### Testing species-specific SNPs as a tool to identify species from single needles

In order to use these molecular tools to identify species in a blind test, 45 anonymous white, black and red spruce samples were provided by Dr. Robert Eckert. DNA was extracted from as little as one needle per sample. SSCP analysis of trnK SNPs 1, 2 and 3, *Msp*I analysis of rpl33-trnP SNP3 and *Bsp*1286I analysis of ITS SNP 7 were carried out in order to identify the white spruce samples. Then ASPCR analysis of trnK SNP 10 and *Bst*UI analysis of ITS SNP 1 was used to distinguish the black spruce samples from the red and white spruce. One hundred percent (45/45) of the anonymous samples were identified correctly.

## Discussion

We sequenced (1) the chloroplast *trn*K intron including *mat*K, (2) the region between chloroplast *rpl*33 and *trn*P including *psa*J and (3) nuclear rDNA 5.8S, ITS2, and portions of ITS1 in 4 individuals each of white, black and red spruce. We identified one nuclear and four chloroplast SNPs which distinguish white spruce from black and red spruce, plus one nuclear and one chloroplast SNP which distinguish black spruce from white and red spruce. These markers are strictly associated with species, as determined by surveying trees representing range-wide provenance tests of each species. In a blind test, the species-specific nuclear and chloroplast markers were used to correctly identify anonymous samples. These molecular markers can be used to identify the species of a white, black and red spruce tree from single needles.

### Levels of sequence variation in spruce

The percentage divergence of all genes sequenced was remarkably low between the spruce species. For *mat*K, there was only 0.1% divergence between red and black spruce, 0.3% between white and red spruce and 0.4% between white and black spruce. This is markedly lower than the 1.1% *mat*K variation between lodgepole pine and black pine (Hilu and Liang 1997).

Between 5% and 15% sequence divergence between the taxa in question is thought to provide a sufficient number of characters for phylogenetic analysis (Olmstead and Palmer 1994). The relatively high level of variation between some angiosperm species makes *mat*K appropriate for resolving phylogenetic relationships (Johnson and Soltis 1995). Lodgepole pine and black pine are believed to be distantly related members of subgenus *Pinus* based on ITS sequences (Liston et al. 1999). White and black spruce are also thought to be distantly related based on RFLPs of chloroplast DNA (Sigurgeirsson and Szmidt 1993). The low levels of variation between distantly related species within these two genera suggest that *mat*K may not be sufficiently divergent to infer relationships among species of *Pinus* or among species of *Picea*.

The noncoding regions of the *trn*K intron displayed no more variation than *mat*K between spruce species. There was 0% divergence between red and black spruce, 0.4% between white and red spruce and 0.4% between white and black spruce. *Pinus trn*K intron noncoding regions also show slightly less variation than *mat*K having 1% divergence between lodgepole pine (GenBank X57097; Lidholm and Gustafsson 1991) and black pine (GenBank D17510; Tsudzuki et al. 1992). The combined noncoding intergenic spacers between *rpl*33 and *trn*P also had very low levels of variation between spruce species: 0% divergence between red and black spruce, 0.5% between white and red spruce and 0.5% between white and black spruce.

There was no variation in the nuclear rDNA ITS2, 5.8S gene or the 3′ end of ITS1 among these three spruce species. This is strikingly different than the level of 5.8S and ITS2 variation among species of *Pinus* (Liston et al. 1999).

Variation was present in the 5′ end of ITS1 (5′ 330 bp): approximately 0.9% divergence between red and black spruce, 2.2% between white and red spruce and 2.2% between white and black spruce. There was also variation in the cloned ITS1 fragment between species: as much as 1.4% divergence between red and black spruce, 2.4% between white and red spruce and 1.4% between white and black spruce. These values may be inflated, however, due to the observed heterogeneity of ITS within an individual.

### Heterogeneity of ITS

Nuclear rDNA is present in multiple copies arranged in tandem repeats (Hamby and Zimmer 1992). Angiosperms have thousands of copies located at one or a few chromosomal loci (nucleolus organizer regions, NORs) (Hamby and Zimmer 1992). Gymnosperms have been shown to possess many more copies; in red and black spruce there are as many as $10^6$ copies rDNA per nuclear genome (Bobola et al. 1992b). Using in situ hydridization Brown et al. (1993) identified 12-14 rDNA chromosomal loci in white spruce.

There have been previous reports of heterogeneity among rDNA repeats in conifers (Bobola et al. 1992b;

Beech and Strobeck 1993; Karvonen and Savolainen 1993; Liston et al. 1996) in both ITS and IGS. Here we present further evidence that the copies of ITS1 are heterogeneous within a spruce individual. (1) Sequence chromatograms from using PCR product as template showed multiple peaks at single positions. (2) Restriction endonucleases *Bst*U1 and *Bsp*1286I each cleaved PCR products differently within screened individuals. (3) There was significant variability between sequences of cloned ITS1 fragments within an individual. In fact, there was 0.8–1.7% divergence between the clones of a red spruce individual, 0.6–1.2% divergence between the clones of a black spruce individual and 1.3–3.2% divergence between the clones of a white spruce individual.

This heterogeneity may be explained by the large number of rDNA chromosomal loci in spruce. Liston et al. (1996, 1999) pointed out that the large number of rDNA loci might slow the process of concerted evolution which tends to homogenize the rDNA repeats within an individual. The high level of heterogeneity of ITS1 within an individual makes estimation of the levels of intra- and interspecific variation difficult. It also complicates the process of identifying species-specific SNPs in the ITS region as markers to distinguish between species. Furthermore, such high levels of intra-individual variation would complicate phylogenetic analysis using ITS1 sequences.

## Species-specific markers to distinguish between spruce species

Perron et al. (1995) stressed the importance of identifying molecular markers from trees whose species has been carefully identified a priori. They chose populations from outside the sympatric zone, where the extent of natural hydridization of red and black spruce is probably low, and conducted a five-character morphological analysis on each tree. Only trees with a morphological composite index specific to either species were used in their study (Perron et al. 1995). In our study, each tree of the provenance tests was typed using six morphological characters (Gordon 1976; Eckert 1990). Additionally, any suspected hybrids and/or introgressed individuals from the provenance tests identified with nuclear and organelle RFLPs (Bobola et al. 1996) were excluded from the population studies.

Perron et al. (1995) identified RAPD markers which distinguish red and black spruce from only six black populations (12 trees) and three red spruce populations (12 trees), and verified them in ten black and nine red spruce $F_1$ progeny from interspecific crosses. They characterized four RAPD markers that were present in 100% of surveyed red spruce (21 total provenance and $F_1$ progeny trees) and absent in 100% of the surveyed black spruce (22 total trees). The three black-distinguishing markers were absent in 100% of the red spruce surveyed but only present in 75–92% of surveyed black spruce provenance trees and in 70–100% of black spruce $F_1$ progeny. Perron et al. (1995) referred to all of these markers as "species-specific".

In this study, we used SNPs as molecular markers to distinguish between white, black and red spruce. A total of 25 candidate interspecific SNPs were identified in the three sequenced gene regions. Of these, 18 distinguished white spruce, 4 distinguished black spruce and 3 distinguished red spruce. The abundance of white-distinguishing markers relative to the number of red- or black-distinguishing markers is not surprising because white spruce is not as closely related as red and black spruce (Sigurgeirsson and Szmidt 1993).

Eight of the candidate SNPs were screened in 11 red spruce populations (46 trees), 30 black spruce populations (84 trees) and 22 white spruce populations (90 trees) in order to assess their degree of species-specificity. These trees represent a considerable portion of the species' ranges.

During the Pleistocene Epoch, spruce populations were displaced in response to glaciation events. Repeated alterations in the geographic distribution of these trees influenced the amounts of genetic variability throughout species (Critchfield 1984). Therefore, range-wide sampling is important for this type of population study in order to include samples from both current and previously disjunct populations that may be experiencing or have undergone genetic bottlenecks.

The nuclear (ITS SNP 7) and chloroplast (trnK SNPs 1, 2 and 3, and rpl33-trnP SNP 3) SNPs which distinguish white spruce are 100% species-specific. The results of the indirect screening methods suggested that all of the white spruce trees surveyed had the white spruce-type nucleotide and that all of the black and red spruce trees surveyed had the black/red spruce-type nucleotide (Table 4). TrnK SNP 4, which also distinguishes white spruce, was 100% consistent in surveyed white spruce trees and 96% consistent in surveyed black spruce. As only 85% of the red spruce trees displayed the red/black spruce-type base, trnK SNP 4 was not considered "species-specific".

Chloroplast trnK SNP 10, which distinguishes black spruce from white and red spruce, is located at the first position of codon 410 in the *mat*K gene. This marker was consistent among 100% of the white spruce surveyed and among 98% of the red and black spruce. Only 1 red spruce tree (out of 46) appeared to have the black spruce-type base, and only 2 black spruce trees (out of 84) appeared to have the red/white spruce-type base.

Nuclear ITS SNP 1 has a more complicated species profile due to the heterogeneity of ITS1 within an individual. At this position, sequence chromatograms (using PCR as template) show a G in white and red spruce and an A in black spruce. *Bst*U1 digestion of PCR product, however, indicates that some of the white and red spruce DNAs have a G at this site and

that some of the black spruce DNAs do not have a G at this site. Although the rDNA copies are not homogeneous within an individual at this site, this marker is still useful for distinguishing between the spruce species because the majority of amplified DNAs have a G in white and red spruce individuals, and the majority of amplified DNAs have an A in black spruce individuals. These restriction profiles were consistent in 99% of surveyed white spruce, 100% of black spruce and 96% of red spruce. Therefore, only 1 white spruce tree (out of 90) and 2 red spruce trees (out of 46) had a majority of DNAs lacking a G at this site.

The validity of the nuclear and chloroplast markers was confirmed by using each of the species-specific SNPs to identify anonymous white, black and red spruce samples in a blind test. This test was carried out with DNA from as little as a single needle per sample, thereby demonstrating the utility of these markers when abundant tissue is not available.

We have identified seven species-specific SNP markers that can reliably identify white, black or red spruce from a single needle. SNPs are easier to identify and more robust than other molecular tools, such as isozymes (Eckert 1989), RFLPs (Bobola et al. 1992a, b, 1996) and RAPDs (Perron et al. 1995), previously used to distinguish spruce species. Furthermore, screening SNP markers with indirect PCR-based methods enables an efficient high throughput of samples.

**Acknowledgements** We wish to thank Dr. R.T. Eckert, Dr. G.R. Furnier and Dr. G.E. Rehfeldt for providing plant materials. We are grateful to Dr. C.D. Neefus for assisting with the statistical analysis and illustrations. We also thank Dr. J.J. Collins, Dr. R.H. Cote and Dr. T.D. Kocher for their critical review of the manuscript. This research was supported by a grant from the New Hampshire Agricultural Experiment Station (NHAES) MacIntire-Stennis Project MS27. The experiments carried out for this project comply with the current laws of the USA.

## References

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann Mo Bot Gard 82:247–277

Baverstock PR, Moritz C (1996) Project design. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics. Sinauer Assoc, Sunderland, Mass., pp 17–27

Beech RN, Strobeck C (1993) Structure of the intergenic spacer region from the ribosomal RNA gene family of white spruce (*Picea glauca*). Plant Mol Biol 22:887–892

Bobola MS, Eckert RT, Klein AS (1992a) Restriction fragment variation in the nuclear ribosomal DNA repeat unit within and between *Picea rubens* and *Picea mariana*. Can J For Res 22:255–263

Bobola MS, Smith DE, Klein AS (1992b) Five major nuclear ribosomal repeats represent a large and variable fraction of the genomic DNA of *Picea rubens* and *P. mariana*. Mol Biol Evol 9:125–137

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. Can J For Res 26:433–443

Brown GR, Amarasinghe V, Kiss G, Carlson JE (1993) Preliminary karyotype and chromosomal localization of ribosomal DNA sites in white spruce using fluorescence in situ hybridization. Genome 36:310–316

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Retting JH, Conti E, Palmer JD, Manhart JR, Systma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang GY, Plunkett GM, Soltis PS, Swensen WM, Williams SE, Gadek PA, Quinn CJ, Equiarte LE, Golenberg E, Learn GHJ, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann Mo Bot Gard 80:528–580

Chaw SM, Zharkikh A, Sung HM, Lau TC, Li WH (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol 14:56–68

Critchfield WB (1984) Impact of the Pleistocene on the genetic structure of North American conifers. In: Lanner RM (ed) Proc 8th N Am For Biol Workshop. International Union of Forest Research Organizations, Birmensdorf, Switzerland, pp 70–118

D'Aquila RT, Bechtel LJ, Videler JA, Eron JJ, Gorczyca P, Kaplan JC (1991) Maximizing sensitivity and specificity of PCR by pre-amplification heating. Nucleic Acids Res 19:3749

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–131

Donoghue MJ, Sanderson MJ (1992) The suitability of molecular and morphological evidence in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 340–368

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Eckert RT (1989) Genetic variation in red spruce and its relation to forest decline in the Northeastern United States. In: Bucher JB, Bucher-Wallin I (eds) Air pollution and forest decline. (Proc 14th Int Meet Specialists Air Pollution Effects For Ecosystems). International Union of Forest Research Organizations P2.05, Birmensdorf, Switzerland, pp 319–324

Eckert RT (1990) Genetic variation in red spruce pollution response. Final report. Spruce-Fir Research Cooperative, Northeast Forest Experiment Station, Radnor, Pa.

Ellsworth DL, Rittenhouse KD, Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. Biotechniques 14:214–217

Fowler DP, Park YS, Gordon AG (1988) Genetic variation of red spruce in the Maritimes. Can J For Res 18:703–709

Furnier GR, Stine M (1995) Interpopulation differentiation of nuclear and chloroplast loci in white spruce. Can J For Res 25:736–742

Furnier GR, Stine M, Mohn CA, Clyde MA (1991) Geographic patterns of variation in allozymes and height growth in white spruce. Can J For Res 21:707–712

Gernandt DS, Liston A (1999) ITS region evolution in *Larix* and *Pseudotsuga*. Am J Bot (in press)

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. Mol Biol Evol 11:769–777

Gordon AG (1976) The taxonomy and genetics of *Picea rubens* and its relationship to *Picea mariana*. Can J Bot 54:781–813

Hamby RK, Zimmer EA (1992) Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 50–91

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. Quart Rev Biol 66:411–453

Hilu KW, Liang H (1997) The *mat*K gene: sequence variation and application in plant systematics. Am J Bot 84 : 830–839

Johnson LA, Soltis DE (1994) *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Syst Bot 19 : 143–156

Johnson LA, Soltis DE (1995) Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. Ann Mo Bot Gard 82 : 149–175

Karvonen P, Savolainen O (1993) Variation and inheritance of ribosomal DNA in *Pinus sylvestris* L. (Scots pine). Heredity 71 : 614–622

Landegren U, Nilsson M, Kwok PY (1998) Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res 8 : 769–776

Li JH, Bogle AL, Klein AS (1997) Interspecific relationships and genetic divergence of the disjunct genus *Liquidambar* (Hamamelidaceae) Rhodora 99 : 229–240

Li JH, Bogle AL, Klein AS (1999) Phylogenetic relationships of the Hamamelidaceae: evidence from the nucleotide sequences of the plastid gene *mat*K. Plant Syst Evol (in press)

Lidholm J, Gustafsson P (1991) A three-step model for the rearrangement of the chloroplast *trn*K-*psb*A region of the gymnosperm *Pinus contorta*. Nucleic Acids Res 19 : 2881–2887

Liston A, Robinson WA, Oliphant JM, Alvarez-Buylla ER (1996) Length variation in the nuclear ribosomal DNA internal transcribed spacer regions of non-flowering seed plants. Syst Bot 21 : 109–120

Liston A, Robinson WA, Piñero D, Alvarez-Buylla ER (1999) Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phyl Evol 11 : 95–109

Little EL Jr (1971) Conifers and important hardwoods. (Atlas of United States trees). U.S. Department of Agriculture, Forest Service, Washington, D.C. Miscellaneous Publ No. 1146, vol 1

Manley SAM (1971) Identification of red, black and hybrid spruce. Can Dep Environ For Can For Serv, Ottawa, Canada

Marrocco R, Gelati MT, Maggini F (1996) Nucleotide sequence of the internal transcribed spacers and 5.8S region of ribosomal DNA in *Pinus pinea* L. DNA Seq 6 : 175–177

Micheli MR, Bova R, Pascale E, D'Ambrosio E (1994) Reproducible DNA fingerprinting with the random amplified polymorphic DNA (RAPD) method. Nucleic Acids Res 22 : 1921–1922

Morgenstern EK, Farrar JL (1964) Introgressive hybridization in red spruce and black spruce. Technical Report No. 4. Faculty of Forestry, University of Toronto, Toronto, Ontario, pp 1–46

Muralidharan K, Wakeland EK (1993) Concentration of primer and template qualitatively affects products in random-amplified polymorphic DNA PCR. Biotechniques 14 : 362–364

Okayama H, Curiel DT, Brantly ML, Holmes MD, Crystal RG (1989) Rapid, nonradioactivite detection of mutations in the human genome by allele-specific amplification. J Lab Clin Med 114 : 105–113

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81 : 1205–1224

Perez de la Rosa J, Harris SA, Farjon A (1995) Noncoding chloroplast DNA variation in Mexican pines. Theor Appl Genet 91 : 1101–1106

Perron M, Gordon AG, Bousquet J (1995) Species-specific RAPD fingerprints for the closely related *Picea mariana* and *P. rubens*. Theor Appl Genet 91 : 142–149

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM (1993) The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. Genomics 16 : 325–332

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

Soltis PS, Doyle JJ, Soltis DE (1992) Molecular data and polyploid evolution in plants. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 177–201

Sommer SS, Cassady JD, Sobell JL, Bottema CD (1989) A novel method for detecting point mutations of polymorphisms and its application to population screening for carriers of phenylketonuria. Mayo Clin Proc 64 : 1361–1372

Steele KP, Vilgalys R (1994) Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastis gene *mat*K. Syst Bot 19 : 126–142

Sugita M, Shinozaki K, Sugiura M (1985) Tobacco chloroplast tRNA Lys (UUU) gene contains a 2.5-kilobase-pair intron: an open reading frame and a conserved boundary sequence in the intron. Proc Natl Acad Sci USA 82 : 3557–3561

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17 : 1105–1109

Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, Wakasugi T, Sugiura M (1992) Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trn*Q, *trn*K, *psb*A, *trn*I and *trn*H and the absence of *rps*16. Mol Gen Genet 232 : 206–214

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23 : 4407–4414

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91 : 9794–9798

White EE, Watkins RF, Fowler DP (1993) Comparative restriction site maps of chloroplast DNA of *Picea abies*, *Picea glauca*, *Picea mariana*, and *Picea sitchensis*. Can J For Res 23 : 427–435

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Inns M, Gelfand D, Sninsky J, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 315–322

Wu DY, Ugozzoli L, Pal BK, Wallace RB (1989) Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proc Natl Acad Sci USA 86 : 2757–2760