Theor Appl Genet (2002) 106:40–50
DOI 10.1007/s00122-002-0986-1

I. Scotti · F. Magni · G. P. Paglia · M. Morgante

Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers

Received: 4 January 2002 / Accepted: 15 March 2002 / Published online: 4 July 2002

Abstract Trinucleotide microsatellites have proven to be the markers of choice in human genetic analysis because they are easier to genotype than dinucleotides. Their development can be more time-consuming due to their lower abundance in the genome. We isolated trinucleotide microsatellites in Norway spruce (*Picea abies* K.) using an enrichment procedure for the genomic-library construction. Here we report on the characterisation of 85 ATC microsatellite-containing clones, from which 39 markers were developed. Many of the clones showed the occurrence of tandem repeats of higher order than the trinucleotide ones, often resembling minisatellite repeats. The sequencing of a sample of the alleles at one of the loci revealed size homoplasy due to base substitutions within the microsatellite region. The presence of ATC motifs within repetitive sequence families was observed. We found a significant relationship between the level of polymorphism and the length of the microsatellite. The levels of variability for ATC trinucleotide markers were lower than those for dinucleotides, both when tested on all loci in a set of six individuals and on a subset of loci in four natural populations. This difference is most likely attributable to lower mutation rates for trinucleotide than for dinucleotide loci. The availability of markers with different mutation rates allows one to select the proper marker set to investigate population processes on different time scales.

Communicated by R. Hagemann

I. Scotti · F. Magni · G.P. Paglia · M. Morgante (✉)
Dipartimento di Produzione Vegetale e Tecnologie Agrarie,
Università degli Studi di Udine.
Via delle Scienze 208 – 33100 Udine UD, Italy
e-mail: michele.morgante@usa.dupont.com
Tel.: 302-631-2638, Fax: +1-302-631-2607

Present address:
M. Morgante, E.I. duPont de Nemours and Co. (Inc.),
DuPont Agriculture and Nutrition, Molecular Genetics,
Delaware Technology Park, P.O. Box 6104, Newark,
DE 19714-6104, USA

Keywords Norway spruce · Molecular markers · Microsatellite · Genetic variability · Trinucleotide

Introduction

Conifer species are characterised by their large genome size (10–30 × 10^9 bp/haploid genome, Murray 1998). Their genome organisation is still largely unknown, but it is nevertheless assumed that repetitive DNA makes up the bulk of their genomes. This may pose problems when trying to develop single-locus genetic markers, especially microsatellites or Simple Sequence Repeats, where unique flanking sequences are used as priming sites to specifically amplify one of the many regions in the genome containing the same microsatellite motif. Despite initial claims of an association with repetitive DNA (Ramsay et al. 1999), microsatellites have recently been shown to be preferentially associated to single-low-copy DNA at least in Angiosperms (Morgante et al. 2002). This is due to the recent amplification of repetitive DNA in species such as maize (Meyers et al. 2001) that has not allowed replication slippage events (Tautz and Schloetterer 1994) to adjust the frequency of microsatellites to an equilibrium value. The frequency with which microsatellites will occur within repetitive sequences is expected to be dependent on the age of the repetitive component. Initial analyses seem to indicate an ancient genome expansion in conifers and thus a very old origin of their repetitive DNA component (Friesen et al. 2001; Stuart-Rogers and Flavell 2001; F. Cattonaro, M. Morgante, manuscript in preparation). Difficulties in the development of microsatellite markers, due to microsatellite flanking sequences being themselves repetitive in the genome, have been encountered in conifers (Pfeiffer et al. 1997; Elsik et al. 2000). As a consequence, large numbers of microsatellite markers are discarded because they produce complex banding patterns or poor amplification and the marker development process can be very ineffective and time-consuming.

We have shown in a separate paper (Scotti et al. 2002) that dinucleotide microsatellite markers can be efficient-

ly developed in Norway spruce by assaying positive clones, taken from a library enriched for microsatellites, for their repetitive DNA content. This selection was done by dot-blot hybridisation with labelled total genomic DNA, and increased the efficiency of obtaining single-locus markers from 19% (Pfeiffer et al. 1997) to 66% of the SSRs identified. A different strategy for efficient development of single-locus microsatellites in conifers has been proposed by Elsik et al. (2000) who have identified markers from libraries obtained from the low-copy portion of the *Pinus* genome.

Most of the microsatellites so far developed in plants have been dinucleotide repeats, since they are more abundant (Morgante and Olivieri 1993) and thus easier to isolate. They suffer however from technical drawbacks for their use in genotyping, since 2-bp differences have to be resolved. Moreover, complex banding patterns ("shadow bands") arise on gels, making them sometimes difficult to score (Litt et al. 1993). For these reasons tri- and tetra-nucleotide repeats have become the markers of choice for population, linkage and forensic studies in humans as well as in other animal species (Gastier et al. 1995; Sheffield et al. 1995; Tozaki et al. 2000). An additional advantage of this class of repeats is that they are more likely than dinucleotide ones to be found within expressed regions or pseudo-genes (Wang et al. 1994; Morgante et al. 2002). Trinucleotide repeats are 3-times more frequent in transcribed than in non-transcribed regions in *Arabidopsis thaliana* as well as in maize (Morgante et al. 2002). They therefore might give a higher proportion of useful markers in species such as conifers, as a consequence of their preferential presence in low-copy number sequences. In plants, no extensive study has been carried out so far on the distribution of trinucleotide microsatellites across the genome, nor have they been developed in large numbers for any species. Small numbers have been reported in several articles: 14 in *Citrus* (Kijas et al. 1997), eight in wheat (Bryan et al. 1997), 18 in *Pinus taeda* (Elsik et al. 2000). Trinucleotide markers developed in *P. taeda* have been shown to be transferable between hard and soft pines (Kutil and Williams 2001).

In conifers, Echt and May-Marquardt (1997) estimated the relative frequency of microsatellites in the genome of two *Pinus* species, concluding that some types of trinucleotide repeats, such as AAT and ATC, are almost as frequent as dinucleotide stretches such as AG and AC. It is therefore even more attractive to attempt the development of markers from trinucleotide repeats in the *Pinaceae*, since the rarity of such repeats has often prevented their development in other species.

In this paper we report on the isolation and characterisation of a set of 85 trinucleotide (ATC) clones, from which it was possible to develop 39 markers, with a variable degree of polymorphism and banding-pattern quality. Three of these markers have also been tested on a set of four natural populations, while the whole marker set has been tested for segregation in the progeny of a controlled cross as well as for polymorphism on a set of six individuals.

Materials and methods

Plant material and DNA extraction

DNA of the tree V34 was used for the construction of the genomic library enriched in ATC repeats. This plant belongs to the Norway spruce population of Val Meledrio (TN, Italy). Thirty six trees from the Val Meledrio and three other populations [Val di Fiemme (TN), Fusine (UD), Tarvisio (UD); Italy] were also sampled for surveying polymorphism. DNA was extracted as described in Doyle and Doyle (1990) with modifications.

The progeny of a controlled cross between clones N2022 and E2006 (Skogforsk, Sweden) was used for testing the segregation of the markers. DNA was extracted from needle tissue using the Qiagen DNeasy kit.

Southern blot and hybridisation

Three micrograms of Norway spruce genomic DNA were digested to completion with the restriction enzymes *Mse*I, *Sau*3A, *Tsp*I (New England Biolabs) separately, following the manufacturer's instructions, and then blotted onto nylon membranes as described in Sambrook et al. (1989). This work was replicated four times to allow for parallel treatment in the following steps. Barley genomic DNA digested with *Tsp*I was also run side by side to provide an internal reference for signal strength.

The four membranes were hybridised to $(AAG)_{10}$, $(AAC)_{10}$, $(ACC)_{10}$ and $(ATC)_{10}$ probes, radioactively labelled as described in Tenzer et al. (1999). Hybridisations were performed in 5 × SSC, 1 × Blocking reagent (Boheringer GmbH, Mannheim, Germany), 0.1% Laurylsarkosine, and 0.02% SDS at 45 °C. Washing was carried out at 45 °C in 0.5 × SSC, 0.1% SDS.

Construction of the genomic library and microsatellite sequence isolation

A genomic library was constructed for isolating trinucleotide ATC/GAT microsatellites. The library was prepared as in Tenzer et al. (1999), using an $(ATC)_{10}$ oligonucleotide as a probe for the enrichment step and for the selection of positive clones. Plasmids were purified (Wizard Minipreps; Promega Corp.) and sequenced on an ALF automated sequencer (Pharmacia Biotech, Inc.) using ThermoSequenase fluorescent-labelled primer cycle sequencing kit (Amersham Pharmacia Biotech).

Sequence similarity searches were performed using BlastN (score for matches: 4; score for mismatches: –5; gap opening penalty: 20).

Microsatellite primer design

Oligonucleotide primers flanking the microsatellite regions were designed using the computer program PRIMER (version 0.5; Whitehead Institute for Biomedical Research, Cambrige, Mass.). The theoretical annealing temperature of the primers was chosen to be 58 °C for all markers, and tested in the range of 50–63 °C.

Microsatellite PCR amplification and SSR scoring

For each primer pair, amplification conditions were tested on a panel of six individual trees from different natural populations of Norway spruce belonging to the alpine range of the species, including the tree used for library construction. The products were resolved on agarose gels or on polyacrylamide gels, depending on their pattern. PCR amplifications were performed in a final volume of 10 µl containing 1.5 ng of genomic DNA, 200 µM of each dNTP, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM $MgCl_2$, 0.25 µM of each forward and reverse primers, and 0.4 U of Ampli*Taq* Gold *Taq* polymerase (Perkin Elmer, Inc., Foster City,

Calif.). In the case of acrylamide-gel separation, one of the two primers was end-labelled in the following way: 2.5 pmol of the primer were combined with 1 μCi of [γ^{33}P]ATP and 0.5 U of T4 polynucleotide kinase (New England BioLabs, Inc., Missisauga, Canada) in 1 × kinase buffer and incubated for 35 min at 37 °C; the enzyme was deactivated by heating for 10 min at 70 °C. The labelling reaction was then combined with the PCR mix. Amplifications were performed in 9600 or 9700 Gene Amp PCR systems (Perkin Elmer, Inc., Foster City, Calif.) with the following profile: 1 cycle of hot start (95 °C for 10 min); 7 cycles of touch down [95 °C for 30 s, (Ta+7 °C) for 30 s $\Delta\downarrow$1.0 °C, 72 °C for 30 s]; 28 cycles of amplification (95 °C for 45 s, Ta for 45 s, 72 °C for 45 s); 1 cycle of final extension (72 °C for 10 min). Non-radioactive amplification products were separated on 3.5% Metaphor agarose gels (FMC BioProducts, Inc.) in 1 × TBE, stained with 1 × ethidium bromide; gels were visualised on a UV lamp and photographed on a Polaroid film. Radioactively labelled amplification products were run on acrylamide gels and visualised as described in Paglia et al. (1998).

For sequencing of microsatellite alleles, a set of plants was chosen from a larger data pool available in our lab. The SSR fragments were produced using *Pfu* polymerase (Promega Corp., Madison, Wis.) and cloned into the pGEM Easy Vector System (Promega Corp., Madison, Wis.), following the manufacturer's instruction. Plasmids were purified (Wizard Minipreps; Promega Corp.) and sequenced as described above. For each individual, multiple clones have been sequenced to check for the consistency of the sequence in relation to possible mutations introduced in the PCR amplification.

Markers SpAGC1, SpAGC2, SpAGG3, SpAC1F7, SpAGD1 and SpAC1H8, used for comparison in the diversity survey, are described in Pfeiffer et al. (1997).

Population parameters and statistical tests have been obtained using the program GENEPOP (Raymond and Rousset 1995).

Results

Southern hybridisation of four trinucleotide microsatellite probes to Norway spruce total genomic DNA (Fig. 1) shows that ATC and ACC stretches appear to be more abundant than AAG and AAC in spruce, especially in comparison to barley. The differences in signal intensity could be partly ascribed to differences in T_m among probes; therefore we included an internal reference in the form of barley genomic DNA. ATC repeats were shown to be very abundant also in pines (Echt and May-Marquardt 1997). The digestions with the enzymes *Sau*3A and *Tsp*I show some discrete bands when probed with AAG, ACC and ATC oligonucleotides, indicating that at least part of the repeats belong to multi-copy genomic fragments. We chose to construct the library for ATC repeats because of their abundance and because, as opposed to ACC repeats, no discrete band that could be indicative of repetitive DNA sequences was visible in the selected size fraction (200–700 bp) of the *Tsp*I digest.

After plaque screening of the genomic library, 121 clones were picked and excised in vivo. One hundred fragments were sequenced; all clones were unique and 85 of them (85%) displayed the expected ATC stretch.

A survey of the sequences shows some peculiar features of these clones. In several cases the repeat turned out to be embedded either into an ATC-rich region (EATC2G09; Fig. 2a) or into a higher-order tandem repeat, with the consensus sequence $(GAT)_n$–(G/T)CCA-

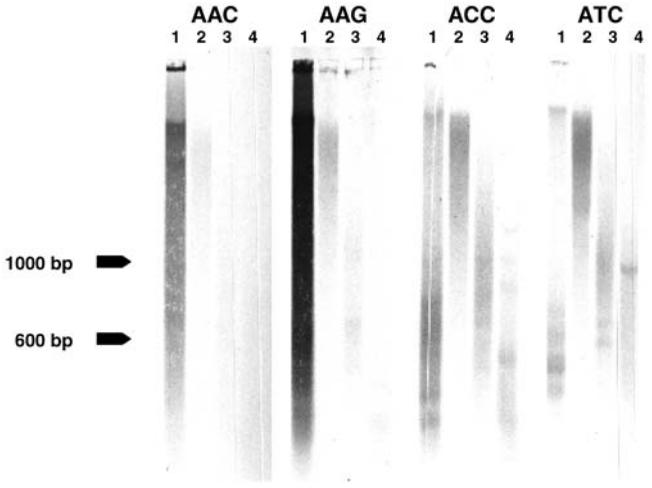

Fig. 1 Hybridisation with microsatellite probes on Southern blots of digested Norway spruce DNA. In each blot: *lane 1*: control (barley DNA, *Tsp*I-digested); *lane 2*: Norway spruce DNA, *Mse*I-digested; *lane 3*: Norway spruce DNA, *Sau*3A-digested; *lane 4*: Norway spruce DNA, *Tsp*I-digested

CTTCA(A/G)CTTCTTGCTGAA (EATC2A03; Fig. 2b). In a few cases the sequence of the clone was such that *in silico* translation in some of the six possible reading frames produced a polypeptide enriched in one amino acid. Only part of the sequence was, however, made of the selected repeat (e.g. clone EATC3H02; Fig. 2c), and the same amino acid was coded in the same sequence also by synonymous codons (Fig. 2d).

Clone EATC1E03 had a structure combining the features of the clones displayed in Fig. 2b and c. Moreover, the size of the cloned fragment (175 bp) was very different from the most common fragment sizes detected in a panel of trees used for surveying marker information content (Fig. 3 and below). To unveil the origin of this large difference we cloned and sequenced some alleles, which allowed us to detect a duplication in the 175-bp fragment encompassing the ATC stretch itself. In this case (Fig. 4), an ATC microsatellite (with variability across alleles) is included in a mini-satellite repeat unit, the sequence of which is $(ATC)_n$(GTC). This locus shows size variability across individuals due to the number of mini-satellite repeat units. Moreover, there is sequence variability in the mini-satellite repeat unit, which adds to the complexity and variability of this sequence.

Fifty three of the 85 clones showing the microsatellite repeat were chosen for the development of markers, based on the length of the microsatellite stretch and on

Fig. 2 **a** Sequence of the clone EATC2G09. The repeat units are in *bold* and the primer sequences are in *italics*. **b** Sequence of the clone EATC2A03. The microsatellite stretches are in *bold*, the higher-order repeat units are *boxed*, and the primer sequences are in *italics*. **c** Sequence of the clone EATC3H02. The microsatellite stretch is in *bold*. **d** *In silico* translation of clone EATC3H02 in two of the six possible reading frames, coding for a long isoleucine- or aspartate-rich stretch (+1 and –2 frame, respectively). Other translations contain at least one stop codon ▶

*TCAACTTCAGCAACACTGTCA*GTCAT**ATC**TATAGGTC**ATC**AAAAGTTTTTCTC**ATCATC**AAT**ATC**AGTAGAAACTAAAC**ATCATCATCATC**TGCAACTTGGAGCTGGAGTAAC**A**

TCATCATCATCAACAACATTAAGTTGGAAC**ATCATCATC**ACCAACAACACGAAGCTGGAAC**ATC**AAC**ATC**TATATGTGGCAATAGAATGAGGCAGGTGC**ATC**AAAAACAAC**ATC**

TGTGGTATTCTGACAGACCGT

a

*AGGAGTCAATGTCGTACCCA*CTAATCTGAACAACTCCTGCCAAAACGTACTCATGAATTT**GATGATGATGAT**GCCACTTCATCTTCTTGCTGAA

GATGATGATGATGATTCCACTTCAGCTTCTTGTTGAAGTT**GATGATGAT**GCCACTTCAACTTCTTGCTGAA**GATGATGATG**AAGTT**GATGATGA**

TGCCACTTCAACTTCTTGCTGAA**GATGATGATGATGAT**GCCACTTCAGCTCCTTGCCGAA**GATGAT**TAT**GATGATGATGAT**GTCATTTCATCTT

TCATGCCGAATATGATGA*TGTTACTTTAGCCTCACGCTG*

b

TTCATCATCATCACCATCATTGTCACCATTGTCATCATTTTCATCACCATCATCATCCTTGTCATCATCGTCGTC**ATCATCATCATCATCATCATCATCATCATCATCATCATCAT**

CATCATCATCATCATCATCATCATCGGCGTCGTCGTCGCCATCATTATCATCGTCTTCCTTTTCATCATTGTCATCACCATAATCAACATTGTCATCCTCCTCCTCGTCGTCATCA

TCATCACAATCATCATCGTCGTCATCATTTTGATCATTGTCATCAACATCATCATTGACATTGTCATTTTCATCATCATTGTCATCACTGTCATCAACATCATCGTTATCATCATCATC

ATTGTCATCATCATAATTCGCACTGAGTCCA

c

+1 Frame:

```
  1  ttc atc atc atc acc atc att gtc acc att gtc atc att ttc atc acc atc atc atc ctt
       F   I   I   I   T   I   I   V   T   I   V   I   I   F   I   T   I   I   I   L
 61  gtc atc atc gtc gtc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc
       V   I   I   V   V   I   I   I   I   I   I   I   I   I   I   I   I   I   I   I
121  atc atc atc atc atc atc atc atc atc ggc gtc gtc gtc gcc atc att atc atc gtc ttc
       I   I   I   I   I   I   I   I   I   G   V   V   V   A   I   I   I   I   V   F
181  ctt ttc atc att gtc atc acc ata atc aac att gtc atc ctc ctc ctc gtc gtc atc atc
       L   F   I   I   V   I   T   I   I   N   I   V   I   L   L   L   V   V   I   I
241  atc aca atc atc atc gtc gtc atc att ttg atc att gtc atc aac atc atc att gac att
       I   T   I   I   I   V   V   I   I   L   I   I   V   I   N   I   I   I   D   I
301  gtc att ttc atc atc att gtc atc act gtc atc aac atc atc gtt atc atc atc atc att
       V   I   F   I   I   I   V   I   T   V   I   N   I   I   V   I   I   I   I   I

361  gtc atc atc ata att cgc act gag tcc a  388
       V   I   I   I   I   R   T   E   S
```

-2 Frame:

```
388 t gga ctc agt gcg aat tat gat gat gac aat gat gat gat gat aac gat gat gtt gat gac
        G   L   S   A   N   Y   D   D   D   N   D   D   D   D   N   D   D   V   D   D
327   agt gat gac aat gat gat gaa aat gac aat gtc aat gat gat gtt gat gac aat gat caa
        S   D   D   N   D   D   E   N   D   N   V   N   D   D   V   D   D   N   D   Q
267   aat gat gac gac gat gat gat tgt gat gat gat gac gac gag gag gag gat gac aat gtt
        N   D   D   D   D   D   D   C   D   D   D   D   D   E   E   E   D   D   N   V
207   gat tat ggt gat gac aat gat gaa aag gaa gac gat gat aat gat ggc gac gac gac gcc
        D   Y   G   D   D   N   D   E   K   E   D   D   D   N   D   G   D   D   D   A
147   gat gat gat gat gat gat gat gat gat gat gat gat gat gat gat gat gat gat gat gat
        D   D   D   D   D   D   D   D   D   D   D   D   D   D   D   D   D   D   D   D
 87   gat gat gat gat gac gac gat gat gac aag gat gat gat ggt gat gaa aat gat gac aat
        D   D   D   D   D   D   D   D   D   K   D   D   D   G   D   E   N   D   D   N
 27   ggt gac aat gat ggt gat gat gat gaa  1
        G   D   N   D   G   D   D   D   E
```

d

Table 1 List of the markers. EMBL = EMBL accession number; Exp. size = size of the cloned fragment; Ta = annealing temperature; Na = number of alleles; C = complex pattern; N = not amplifying; Segr = segregation tested

SSR name	EMBL	Forward primer	Reverse primer	Repeat	Exp. size	Ta	Na	Segr.
EATC1A02	AJ296686	CGTCATCCTCGTCATCGT	TGATTATGGTGATGCAATGAA	$(CAT)_4\ (CCT)_2\ (CAT)_4$ $CGT(CAT)_2\ (CCT)_3$ $(CAT)_2\ N_{21}\ (CAT)_4$	226	60	C	Y
EATC1A05	AJ296687	CATTTCTTAGAATGTTTTTTCATCT	TATAGGGTCTCCTCCATCCC	$(ATC)_5\ N_{18}\ (ATC)_6$ $TTC(CTC)_{10}\ (ATC)_3$ $TTC(ATC)_2\ TTCCTC(ATC)_7$	216	53	N	
EATC1A07	AJ296688	TGACAATGACAATGACGATGA	TCATCTTGCACCTCCTCATC	$(ATC)_7$	150	63	4	
EATC1A10	AJ296689	TCTTATGAGAAGAATGTTCACTTCC	ATAAGGTGAAGAACCTCAAGAGC	$(CAT)_4\ CGT(CCT)_2\ (CAT)_4$	104	60	C	
EATC1B02	AJ296690	TGGCATGAGATTTATGTGGTT	GTGTGCCACTCAACCTCAC	$(ATC)_7\ (AT)_3$	199	57	3	Y
EATC1B03	AJ296691	GACACCGATGATGAAAATGAC	GAAAGGACATATGAAAATTACATGG	$(ATC)_5$	297	63	1	
EATC1B04	AJ296692	CAGTTACCCTCTTAAGGTCATCG	CTAGGTATGCGAAGGGGC	$(CGT)_{16}\ (TCA)_{51}\ (TCC)_3$	268	53	N	
EATC1C01A	AJ296693	TCTTAATCAACATTGTCATTGTCC	GTGGCCAACACATGGTATG	$(TCA)_4\ N_{50}\ (TCA)_4$	139	61	1	
EATC1C02	AJ296694	GACGACGACAATGATAACCAT	TGTCATCACCATAATCAACATT	$(CAT)_5\ TGCAGTCGT(CAT)_3$ $CGT(CAT)_3\ GGT(CAT)_7$	148	53	N	
EATC1C07	AJ296695	TGCTACTTCAACTCCTTGCTG	TCTCCGGCATGAAAGATG	$(CAT)_6\ N_{29}\ (ATC)_5$ $N_{58}\ (CAT)_4\ N_{58}\ (ATC)_4$	258	57	C	Y
EATC1C09	AJ296696	GCATCAACATGGATTAGTATAACA	ATGATGAGCCAACGTCGT	$(CAT)_7\ N_{68}\ (CAA)_4$	254	60	C	Y
EATC1C10A	AJ296697	CAGTGCCAATTATAGTATTTCAACA	CTTCCAATGATTTATAGATGTGTCA	$(GA)_8$	153	53	N	
EATC1C10B	AJ296697	ACATTTGGAACATTGATTGACA	GACACATCAGTGAGAAGGGG	$(CAT)_{10}$	179	53	N	
EATC1D01	AJ296698	ACGGTGATGAAAATGAGCAC	CTTCCCAAATGTTCCAACC	$(ATC)_3\ AT(ATC)_6$	205	60	C	
EATC1D02A	AJ296699	TTGTCATCGTCGTCATTGTC	TTTTAGCCTCTGTTTTCTAGCG	$(TCC)_4\ N_{15}\ (TCA)_{16}$ $TCC(TCA)_5\ (TCC)_4\ (TCA)_4$	216	60	3	
EATC1D03A	AJ296700	GTTGTCGTCATCGTTGACATT	ATAATGTTGATTATGGTGATGACA	$(CAT)_6\ N_{49}\ (CAT)_4$	141	60	C	Y
EATC1D03B	AJ296700	TTTTTTCACACTTATCATCACTGG	TCCTTTGATCATCTTGACAATG	$(CAT)_4\ (CCT)_4$	194	60	2	
EATC1D10	AJ296701	GATGTCGATGTTGTTCCAGAC	CATCACAGATCAAGGTTGGAG	$(ATC)_4$	176	61	1	
EATC1E03	AJ296702	CCCCTTATTCCTAACGTCAAA	TACCAGTGGTGACAACGATG	$(CAT)_4\ CGT(CAT)_8$ $CGT(CAT)_4\ CGT(CAT)_4$ $CGT(CAT)_4$	175	57	6	
EATC1F03B	AJ296703	GACAATGACAATGTTTATTATGGTG	TTATTTTCTTCCTCATCATCATTG	$(TCA)_4\ N_{21}\ (TCA)_7$	128	59	1	
EATC1F07B	AJ296704	AACCTCTAATACGCCAACTCC	ATGGAAACTGCGATGACAAT	$(TCA)_{18}$	186	63	7	Y
EATC1F08	AJ296705	CCACTTCAGCTCCTAAGTCCT	TTGGAACACTGACTGGCATAT	$(ATC)_5\ N_{23}\ (ATC)_5$ $N_{23}\ (ATC)_5$	257	60	C	
EATC1F09	AJ296706	ACAATGACACGGATGGAAAC	TACCTACTCCTCCTTCGCCT	$(TCA)_4\ N_{12}\ (TCA)_{10}$	113	59	6	
EATC1G02	AJ296707	TTTATGTTCACTGATGTTGATGC	CATTAACAACATGGAGCTGGA	$(CAT)_9\ N_{22}\ (CAT)_5$ $N_{22}\ (CAT)_4\ N_{25}\ (CAT)_4$	211	60	3	
EATC1G03	AJ296708	ACTAGTACTCGTTCAAAACAACCC	GATGGTGCCAGTATCAGGAC	$(ATC)_9\ AAT(ATC)_2$	187	53	N	
EATC1G09	AJ296709	GCCACTTCAGCCTCTTGC	AGCAAGGACTCAGGAGCTG	$(ATC)_6$	66	60	C	
EATC2A03	AJ296710	AGGAGTCAATGTCGTACCCA	CAGCGTGAGGCTAAAGTAACA	$(ATC)_4\ N_{31}\ (ATC)_5$ $N_{102}\ (ATC)_5\ N_{22}\ (ATC)_4$	265	53	N	
EATC2A04	AJ296711	ATGTTGTTCAGCCTCGTGTT	TGGCATCATTATCATCATCTTC	$(ATC)_5\ N_{22}\ (ATC)_4$	159	60	C	
EATC2A05	AJ296712	TGCTGTGATGATGTGTTTCC	TAGCAAGAAGCTAAAGTGGCA	$(ATC)_4$	196	60	C	
EATC2B01	AJ296713	AGCGTGAGGCTAAAGTAGCA	TGTTTTGATATGGTTGGTGTTG	$(CAT)_6\ N_{52}\ (ATC)_4\ (AAC)_3$	212	53	N	
EATC2B02	AJ296714	GATGGATCTATGTTGGTTCACC	TTGGTCTCAAGGGAAGTTAATC	$(CAT)_8$	185	57	3	
EATC2B03	AJ296715	TTAGCATGAAGCTGAAGTAGCA	CAACCTCGTGTTGGTGATATT	$(ATC)_4\ N_{22}\ (ATC)_5\ N_{19}\ (ATC)_5$	165	60	C	
EATC2B04	AJ296716	TGCCAATTTGTCTCATTGTG	CATCATCCAGTGAAGTCTTGTG	$(CAT)_8$	134	53	N	
EATC2C01	AJ296717	AGCTACCACGTCAAACCTTG	CTTATACAATCCCCTGACTCCC	$(AT)_7\ (CAT)_8$	104	57	5	
EATC2C03	AJ296718	AACACAGAGGTTGGAGCATC	ATTGTTGTGTAGTGTGTCAGCC	$(CAT)_4\ N_{29}\ (CAT)_5\ N_{29}\ (CAT)_5$	299	53	N	
EATC2C04	AJ296719	AACGATGACACCGATGATGA	ATCTTTGTCATTTCCATTGTCA	$(ATC)_5$	104	53	N	

Table 1 (continued)

SSR name	EMBL	Forward primer	Reverse primer	Repeat	Exp. size	Ta	Na	Segr.
EATC2C06	AJ296720	TATGTGTCACTGCATTGGGA	TGTGTGGGAATACATGTCCA	$(CAT)_7$	158	58	3	
EATC2D04	AJ296721	TATCGAAACCGACATCGATG	TGGCCAATCCTTACAATGTT	$(ATC)_9$	104	53	N	
EATC2D05	AJ296722	AAAAGCAAAGAGCAACTCCTC	CGGATCATGAGAACGAGAAG	$(ATC)_{11}$	231	53	N	
EATC2E07	AJ296723	GATGACACCACTGATGAAAATG	TTGTCACCATCGTCATGATC	$(CAT)_5$	175	57	1	
EATC2E08	AJ296724	ATCATTGCTCACAACATGTAGC	ATCCATGTTGATGCAGATAATG	$(ATC)_5$	163	57	1	
EATC2F07	AJ296725	AGCTTAGTGTTGACGTCGATG	AATATCATCATCCAAAGCAAGG	$(ATC)_4\ N_{33}\ (TCA)_4$	175	60	2	Y
EATC2F08	AJ296726	ACTAGTGTCTTCATGTCATTCT	GATAATGAAAATGAAAAATGATC	$(ATC)_{12}$	208	53	N	
EATC2G05	AJ296728	TGGAGCATGGGTAAATCG	TACCTCACACCCGTGAGAAT	$(AAT)_5\ (CAT)_{16}\ CAA(CAT)_4$	240	53	2	Y
EATC2G06	AJ296729	CCAATAGACCACTTCGGTCT	AGTACCTGTCATCTCGAGCAC	$(ATC)_{71}$	332	53	N	
EATC2G08	AJ296730	GTGTTACCAGCCTTTGTGCT	GTGGCATAATCATCATCAACTG	$(CAT)_4$	129	57	1	
EATC2G09	AJ296731	TCAACTTCAGCAACACTGTCA	ACGGTCTGTCAGAATACCACA	$(ATC)_4\ N_{22}\ (ATC)_4$	249	57	1	
EATC2H03	AJ296732	CCAATGCAACATTTGTATCATC	GGTCGACAGATGAAGAAGATG	$(ATC)_7$	206	58	C	
EATC2H05	AJ296733	ACATCATCATCTGCAGTAGGG	CCGTGTTGTGATTTTGTTGA	$(ATC)_4$	136	58	1	
EATC2H06	AJ296734	ATCCATGTCGATGCAGTTG	ACTCAGGAGCTGAAGTGGC	$(ATC)_5\ N_{22}\ (ATC)_5$	123	58	1	
EATC3B02	AJ296735	AGCTTCTTGCTGAAGATGATG	ATCATTCACATCATGGAGCTAG	$(CAT)_4$	130	60	C	
EATC3C05	AJ296736	TTAGTGGACGTTCATCATCATC	TCACAATCACTTTTTTTAGTCGC	$(CAT)_{10}\ TAT(CAT)_{10}\ N_{78}\ (AT)_{20}$	296	57	7	
EATC3G04	AJ296738	ATGATTTCATTTCGATTATGGG	CACACTCAGGAAAACCCATC	$(CAT)_{12}$	132	57	5	
EATC3H02	AJ296740	CATTGTCATCATTTTCATCACC	CGACGATGATGATTGTGATG	$(CAT)_{24}$	235	53	N	
EATC3H03	AJ296741	TCATCCTCCTTCTCCACATC	AAGAGGATGAGAAGGGTGATG	$(CCT)_4\ (CAT)_6$ TGTCAT$(CCT)_4\ (CAT)_4$ AGT $(CAT)_5\ N_{36}\ (CAT)_5\ N_{59}\ (CAT)_6$	255	57	5	

Table 2 Basic parameters of the markers as computed on the global population. N = number of individuals; H_e = Nei's expected heterozygosity (Nei 1973); N_A = number of observed alleles

Locus	N	H_e	N_A
SpAGC1	143	0.543	19
SpAGC2	140	0.860	23
SpAGG3	146	0.918	21
SpAC1F7	142	0.540	10
SpAGD1	136	0.956	39
SpAC1H8	129	0.943	41
Mean (SE)		0.793 (0.198)	25.50 (12.09)
EATC1B02	135	0.471	6
EATC1G02	119	0.506	7
EATC1E03	133	0.547	7
Mean (SE)		0.508 (0.038)	6.67 (0.47)

the availability of flanking regions for primer design. All sequences with less than four perfect consecutive repeat units were discarded. Twenty four clones carry simple perfect microsatellites, while the others are imperfect, compound or very complex (Table 1). The ATC motif was often accompanied in the compound microsatellites by other trinucleotide motifs differing by a single base, most frequently CTC (seven sequences) followed by AAC (two sequences), GTC and ATA (one sequence each). In three other sequences the AT dinucleotide was found next to the ATC motif. For two clones (EATC1C10 and EATC1D03), two primer pairs were designed because of the presence of two separate repeats, so that the total number of primer pairs tested was 55.

The selected clones were compared to each other and to the GenBank nucleotide sequence database as described. Figure 5 displays the similarity level for each pair of microsatellite clones. A group of 18 clones appears to form a cluster; eight more clones form four doublets. Moreover, five of these clones display significant similarity to classes of Norway spruce repeated DNA, such as portions of retroelements (although the cluster of microsatellite clones as a whole does not show similarity to any known class of repetitive elements). Some of the clones with high scores for similarity, however, also show significant similarities (as measured by the E value) to sequences from species other than conifers (see Fig. 5). These similarities, as well as those between the same sequences and other spruce ATC clones, seem to be non-specific and caused simply by their overall low sequence complexity that is not adequately filtered by the BLAST algorithm for masking simple sequences (see Fig. 5).

The markers were amplified on a panel of six trees as described in the Materials and methods. The primers were designed to have an annealing temperature of 58 °C, but most of them were tested experimentally in the range of 53–63 °C to optimise PCR conditions.

After optimisation of the annealing temperatures, 26 primer pairs (47.3%) produced a simple banding pattern, with no more than two bands per sample. These bands should be attributable to single loci. Thirteen pairs (23.6%) produced a complex pattern with up to 20

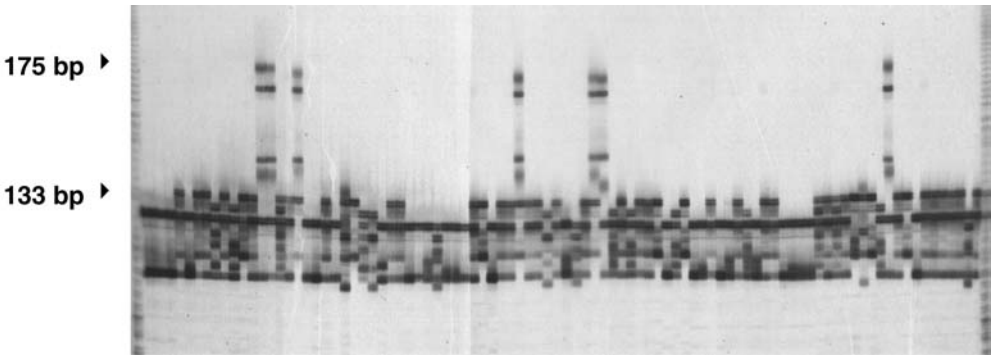

Fig. 3 Acrylamide gel run of marker EATC1E3 on a sample of individuals from the stand of Val Meledrio (IT)

```
133 : CCCCTTATTCCTAACGTCAAAAGCTATTACACCAAATCAATGAAGAATAGGAATGTCATCATCATCATCGTCATCATCATCATCATCATCATC----GTTCTCATCGTCATCCTCATCGTTGTCACCACTGGTA
133 : CCCCTTATTCCTAACGTCAAAAGCTATTACACCAAATCAATGAAGAATAGGAATGTCATCATCATCATCGTCGTCATCATCATCATCGTCATC----GTTCTCATCGTCATCCTCATCGTTGTCACCACTGGTA
133 : CCCCTTATTCCTAACGTCAAAAGCTATTACACCAAATCAATGAAGAATAGGAATGTCATCATCATCATCGTCATCGTCATCATCATCATCATC----GTTCTCATCGTCATCCTCATCGTTGTCACCACTGGTA
136 : CCCCTTATTCCTAACGTCAAAAGCTATTACACCAAATCAATGAAGAATAGGAATGTCATCATCATCATCGTCATCATCATCATCATCATCATCATC-GTTCTCATCGTCATCCTCATCGTTGTCACCACTGGTA
175 : CCCCTTATTCCTAACGTCAAAAGCTATTACACCAAATCAATGAAGAATAGGAATGTCATCATCATCATCGTCATCATCATCATCATCATCATC---ΔATTCTCATCGTCATCCTCATCGTTGTCACCACTGGTA

Δ= GTCATCATCATCATCGTCATCATCATCATCGTCATCATCATC
```

Fig. 4 Alignment of five alleles of the locus EATC1E03. Allele sizes are displayed on the left. Primer sequences are in *italics*. Microsatellite regions are *underlined*. The "Δ" in the sequence of allele 175 indicates the insertion, displayed in the lower part of the figure. Residues that deviate from the consensus are highlighted in *bold*

bands, in which allelism of the bands could not be determined without the analysis of segregation. In some cases, the pattern included an evenly spaced ladder (data not shown), which might suggest that the primers amplify regions composed of a variable number of tandem repeats, present in multiple copies in the genome or displaying the priming sites within the repeat. Sixteen primer pairs produced no band (29.1%), and may have been designed on chimerical clones in which recombination had occurred during the enrichment process or due to errors in the sequence or primer design. Ten of the primer pairs that produced a simple pattern showed only one band in the test panel of six trees. Only one of the monomorphic pairs has more than five repeat units. In the 26 markers producing a simple pattern, a significant correlation ($r = 0.487$; $P < 0.001$) could be found between the length of the longest uninterrupted repeat and the number of bands amplified by the corresponding primer pair. The correlation between the quality of PCR pattern (no amplification vs monomorphic vs. polymorphic vs. complex) and the assignment of clones to genomic fractions (repeated vs. single copy) was tested by Fisher's exact test on the corresponding contingency table, in which clones are grouped according to the two classification criteria. No significant departure from random distribution was found ($P = 0.055$; 1,000,000 iterations in the Markov chain), though the majority of polymorphic markers (12 out of 16) came from the 26 single-copy clones and the majority of complex patterns (9 out of 13) were produced by primer pairs designed on the 27 multiple-copy clones.

Three markers (EATC1B02, EATC1E03, and EATC1G02), that amplified respectively 3, 6 and 3 alleles in the test panel of trees, were also tested in a more extensive survey on 36 plants from each of four natural populations (Val Meledrio, Val di Fiemme, Tarvisio, Fusine; Italy). For this purpose, six dinucleotide microsatellites [from the set reported in Pfeiffer et al. (1997)] were also amplified from these samples in order to compare the variability in the two types of microsatellites. Expected heterozygosity and the number of observed alleles in this sample are shown in Table 2. The trinucleotide loci showed on average a considerably lower level of polymorphism than the dinucleotide ones both in terms of the number of alleles per locus as well as of the expected heterozygosity. It is interesting to notice that at one locus (EATC1E03) almost all the allelic variants were already detected in a panel of only six trees (six alleles compared to seven in the larger sample set; compare Table 1 and Table 2). An example of one of these markers, run on acrylamide, is shown in Fig. 3.

In order to test the inheritance of the new markers, they were all scored in a controlled F1 cross between two Swedish Norway spruce provenances. Only markers that were heterozygous in either parent could be checked for segregation in this cross. Eight such markers were scored (Table 1). Four of these belong to the set of primer pairs producing complex patterns. The multi-locus nature of these bands was clearly indicated by the independent segregation of up to seven bands in a single pattern (data not shown). For all eight primer pairs, Mendelian segregation of at least one PCR product was observed. All the multi-band markers were scored from acrylamide gels, while it was impossible to univocally score them on agarose. These markers will be included in a genetic linkage map which is under construction (Scotti et al. in preparation).

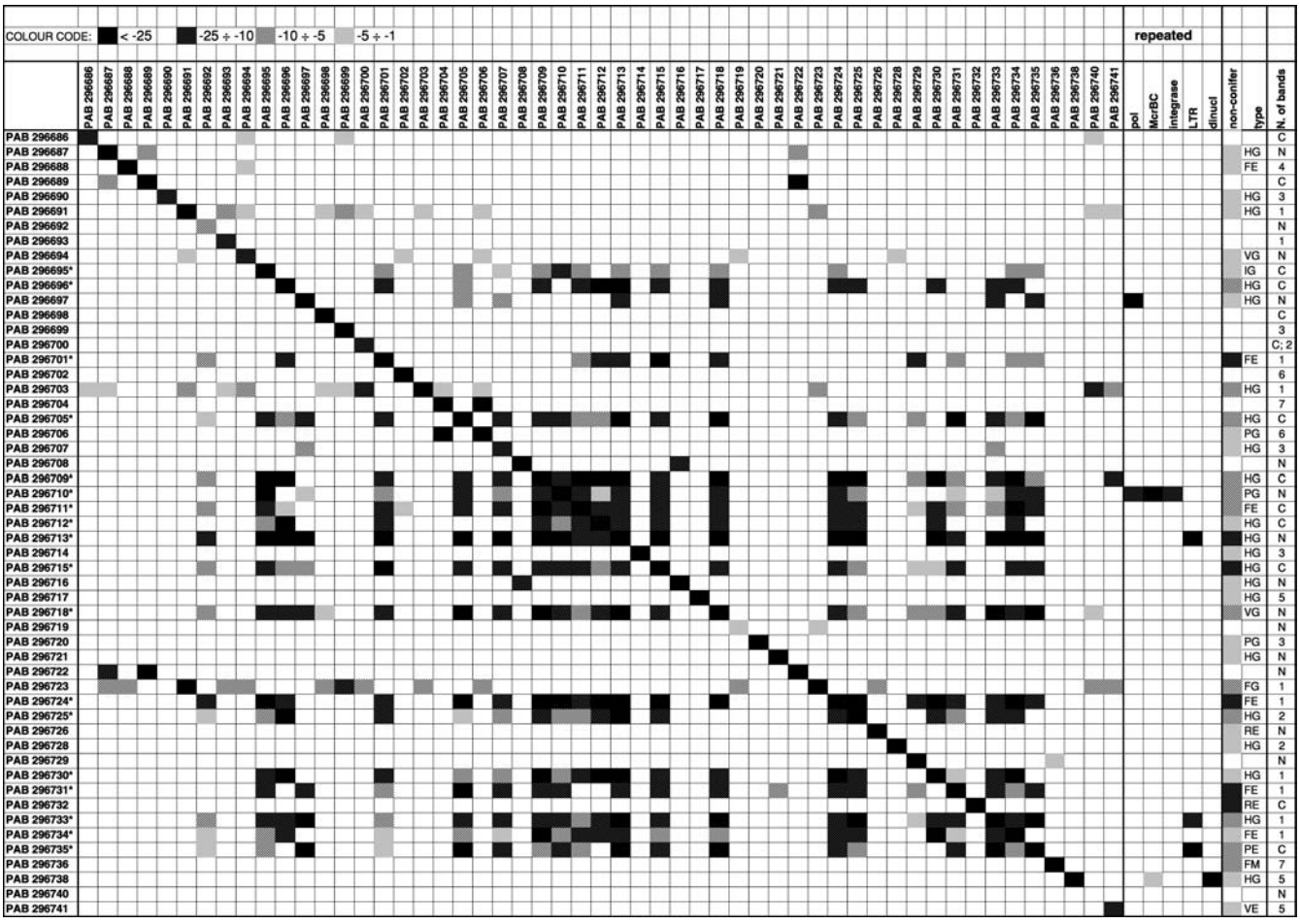

Fig. 5 Similarity matrix for the 53 microsatellite clones. Similarity is expressed here as Log(E-value). *Pol/integrase/LTR*: similarity to portions of retrotransposable elements (EMBL clones: AF150992–993; AF152541 to -546; AF153059; AF180427 to -443; AF180923 to -937; AF187273 to -300; AF364405 to -407; AF367768 to -771; AY032680-681; PAB243312 to -319; PAB288018 to -053; PAB288171 to -181; PAB290585 to -595; PAB290661 to -676; PAJ224363 to -368); *Mcr*BC: similarity to clones obtained by restriction with the enzyme *Mcr*BC, corresponding to the hyper-methylated fraction of the genome (EMBL clones AF305098 to AF305192); *dinucleotide*: similarity to dinucleotide microsatellite containing clones (EMBL clones: PAB292690 to -739; PAB292740 to -742; PAB197477 to -481; PAB131096 to -109; PAG821 to PAG856). *Non-conifer*: most significant match with a non-conifer sequence. *Type*: class of the non-conifer sequence showing the most significant score; P/V/H/I/F/R = plant/vertebrate/human/invertebrate/fungal/protozoan; G/E/M = genomic/expressed/mitochondrial. Clones marked with "*" belong to a group of related sequences (see text)

Discussion

Despite their complex structure, the efficiency of the whole process of marker development from ATC microsatellite-containing sequences was satisfactory: 47.3% of the primer pairs produced a simple pattern (although a subset of these, corresponding to 18.2% of the pairs, was monomorphic in our screening). This figure is intermediate between what was observed for a random set of dinucleotide markers (19%; Pfeiffer et al. 1997) and the result for a set of dinucleotide markers enriched for single-copy clones through hybridisation to labelled total genomic DNA (66%; Scotti et al. 2002). This shows that the genomic distribution of ATC microsatellites is different from that of the dinucleotide ones and suggests that trinucleotide repeats are preferentially found in the single- or low-copy number fraction of the genome in conifers. This is confirmed, through a different approach by Elsik and Williams (2001) for *P. taeda*, and finds further support in other species (Young et al. 2000; Morgante et al. 2002). The fact, however, that several primer pairs designed on these clones produce complex and multi-locus patterns is an indication that ATC microsatellites can also be found within repetitive sequences. The evenly spaced ladders obtained with some primer pairs can be taken as a further indication of the presence of tandemly repeated arrays of motifs containing microsatellites. The presence of a group of 18 clones showing sequence similarity, plus four doublets, strongly suggests the presence of repeated sequences containing trinucleotide SSRs. The PCR-based enrichment procedure that was performed on genomic DNA during the library construction may have caused preferential amplification of specific repeats. The observed sample may therefore not be fully representative of the diversity of ATC microsatellites in the genome and does not allow for inferences on the ab-

solute genomic abundance of these repetitive families. As shown in the segregation analysis, at least some of the multi-band markers are also useful in mapping and can be counted somehow among the successfully developed ones. This group of markers, however, can only be scored on acrylamide gels, while the reduced stuttering, and the distance between alleles, favours the use of high-resolution agarose gels with single-locus trinucleotide SSRs, circumventing the need for isotopes or silver staining.

The microsatellite clones reported in this paper show some peculiar features. On one hand, a subset of the clones displays tandem repeats of motifs larger than the SSR units; on the other hand, a subset of the clones can be translated *in silico* to a biologically meaningful protein sequence; in some cases the same clones share both characteristics (such as EATC1E03). Clone EATC3H2, which displays a region rich in the degenerated repeat TCN (Fig. 2c), can be translated in long amino-acid stretches, with the longest uninterrupted ones coding for an isoleucine-rich or an aspartate-rich stretch (Fig. 2d). The most frequent trinucleotide we found next to ATC in complex repeats is CTC. Young et al. (2000) reports a similar situation in yeast. In *A. thaliana* the open reading frames in full-length cDNA sequences preferentially have GAT stretches over ATC ones in the sense strand (Morgante et al. 2002). These stretches probably code for aspartic acid (GAT), since the two other reading frames correspond either to a stop codon (TGA) or to the ATG start codon (methionine). Although simulated protein translations are just speculative clues, these features might suggest that the distribution of ATC repeats is not entirely random, and that they are potentially found in expressed regions. In *Arabidopsis* ATC is the second most-abundant microsatellite type in coding sequences and its frequency is 3-times higher than in non-coding regions (M. Morgante, M. Hanafey, W. Powell, unpublished results). We have found large numbers of sequences containing ATC stretches longer than ten repeats in public databases of expressed sequences [e.g. the EST (Expressed Sequence Tag) branch of the EMBL database] across all Eukaryote kingdoms (data not shown). Young et al. (2000) show that, in yeast, trinucleotide repeats are preferentially found in the coding region of regulatory genes, even though they are unlikely to correspond to the most-functionally constrained domains of the coded protein. This is consistent with the observation of length variability, since the length of the portions of a protein, acting as a linker between functional domains, does not appear to be a determinant for protein function (with few exceptions; Young et al. 2000).

Trinucleotide microsatellites showed a lower polymorphism level than dinucleotide ones in Norway spruce. The trinucleotide microsatellites amplifying a single-locus produced on average 2.96 alleles on a test panel of six trees, as opposed to 4.63 alleles per locus produced by dinucleotide markers on a similar panel (data not shown; see Scotti et al. 2001). Similar differences in levels of polymorphism between the two classes were found in different species (Huttel et al. 1999; Rossetto et al. 1999). Even with lower levels of diversity than the dinucleotide SSRs, trinucleotide markers have proven useful to discriminate cultivars of plant species with a rather narrow genetic base, such as soybean (Song et al. 1999). When a smaller set of loci was applied to a larger sample of four populations of Norway spruce, trinucleotide markers displayed even lower levels of variation in comparison to dinucleotide ones, both in terms of the number of alleles per locus and diversity as measured by Nei's expected heterozygosity. Neutral theory predicts that diversity is the product of mutation rate by effective population size. The behaviour of the two marker classes in these four populations could therefore be attributed either to different mutation rates or to non-neutrality of trinucleotides, if they are within coding sequences. An uninterrupted open reading frame was observed in only one of the markers used, EATC1E03, but not in the two others, that had multiple stop codons and no ORF longer than 50 amino acids in all reading frames (data not shown). The levels of variability were very similar among the three markers (actually EATC1E03 was the most variable), thus making it unlikely that they may be under direct selective constraints. This may point to a difference in mutation rate and suggest distinct applications for the two types of markers. Over longer evolutionary times it may be more suitable to look at more stable, though polymorphic, markers such as trinucleotide SSRs, while hypervariable markers would suffer from high "background mutation noise". The availability of markers with large differences in mutation rates allows one to select the proper marker set to investigate population processes on different time scales.

The level of polymorphism of this class of trinucleotide repeats has been found in this study to be correlated with the length of the longest uninterrupted microsatellite repeat. In the literature, proportionality between repeat length and marker polymorphism has been reported in some cases (Taramino et al. 1997) but excluded in others (Donini et al. 1998). The rather high proportion of monomorphic markers may be caused by the small number of repeat units and by the fact that these are often interrupted. Moreover, trinucleotide microsatellites generally appear to be more stable compared to dinucleotide stretches of the same size (Echt et al. 1996; Chakraborty et al. 1997). Chakraborty et al. (1997) report the mutation rate to be inversely related to repeat unit length for neutral microsatellite loci, while the situation is different for disease-related loci. It has been observed (Kijas et al. 1995; Scotti et al. 1999) that the use of short probes for the selection of positive clones from libraries tends to select for fragments with shorter, interrupted repeats. An option to solve this is to hybridise at higher stringency with longer probes, while in this study we used an $(ATC)_{10}$ for both library construction and library screening.

As is shown by the example of EATC1E03, there can be a considerable amount of sequence variation within a locus in addition to that detected by size variation (Fig. 4

shows three alleles all of size 133 bp but with base substitutions in the microsatellite region). This situation, known as size homoplasy, has been studied in *Drosophila* (Colson and Goldstein 1999), where complex mutation processes have been identified within the repeat stretches of microsatellites. This has two major implications. The first one is that a large amount of evolutionary information could be hidden behind size variation. Loci such as EATC1E03 can be very informative because of the multiple independent events that might have occurred within the repeat, as they can be deduced by the sequence variability found even in a small sample set (Fig. 4). The second implication is for the applicability of the Stepwise Mutation Model (Kimura and Ohta 1978), since length of the alleles alone may not be representative of their evolutionary relationships.

With the addition of this set of trinucleotide microsatellites to the dinucleotide ones that we have already published (Pfeiffer et al. 1997; Scotti et al. 2002) we have developed a total of 177 markers for Norway spruce. Sixty nine of them in total are single-locus co-dominant multi-allelic markers and are thus suitable for population genetic analyses. Fifty five other markers that amplify multiple loci can be used for genetic mapping studies (Paglia et al. 1998) and/or provide dominant markers. A large set of sequence tagged site markers (STSs; Olson et al. 1989) is thus now available for genetic analysis in Norway spruce.

Acknowledgements This project was funded by the European Union grant BIO CT-960706 "Tree biodiversity". The authors thank Nicoletta Felice and Stefania degli Ivanissevich for technical assistance, and Gunnar Jansson of Skogforsk (Sweden) for providing plant material of the controlled cross.

References

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor Appl Genet 94:557–563

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deaka R (1997) Relative mutation rates at di-, tri-, and tetra-nucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046

Colson I, Goldstein DB (1999) Evidence for complex mutations at microsatellite loci in Drosophila. Genetics 152:617–627

Donini P, Stephenson P, Bryan GJ, Koebner RMD (1998) The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley. Genetc Res Crop Evol 45:415–421

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Echt C, May-Marquardt P (1997) Survey of microsatellite DNA in pine. Genome 40:9–17

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996) Characterisation of microsatellite markers in eastern white pine. Genome 39:1102–1108

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

Elsik CG, Williams CG (2001) Families of clustered microsatellites in a conifer genome. Mol Genet Genomics 265:535–542

Friesen N, Brandes A, Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers. Mol Biol Evol 18:1176–1188

Gastier JM, Pulido JC, Sunden S, Brody T, Buetow KH, Murray JC, Weber JL, Hudson TJ, Sheffield VC, Duyk GM (1995) Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. Hum Mol Genet 4:1829–1836

Huttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217

Kijas JMH, Fowler JCS, Thomas MR (1995) An evaluation of sequence tagged microsatellite markers for genetic analysis within *Citrus* and related species. Genome 38:349–355

Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. Theor Appl Genet 94:701–706

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci USA 75:2868–2872

Kutil BL, Williams CG (2001) Triplet-repeat microsatellites shared among hard and soft pines. J Hered 92:327–332

Litt M, Kramer P, Hauge XY, Weber JL, Wang Z, Wilkie PJ, Holt MS, Mishra S, Donis-Keller H, Warnich L (1993) A microsatellite-based index map of human chromosome 11. Hum Mol Genet 2:909–913

Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution and transcriptional activity of repetitive elements in the maize genome. Genome Res 11:1660–1676

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

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Murray BG (1998) Nuclear DNA amounts in Gymnosperms. Ann Bot 82 (Supplement A):3–15

Nei M (1973) Analysis of genetic diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Olson M, Hood L, Cantor C, Botstein D (1989) A common language for physical mapping of the human genome. Science 245:1434–1435

Paglia GP, Olivieri AM, Morgante M (1998) Towards second generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.). Mol Gen Genet 258:466–478

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

Ramsay L, Macaulay M, Cardle L, Morgante M, degli Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–425

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rossetto M, McLauchlan A, Harriss FCL, Henry RJ, Baverstock PR, Lee LS, Maguire TL, Edwards KJ (1999) Abundance and polymorphism of microsatellite markers in the tea tree (*Melaleuca alternifolia*, Myrtaceae). Theor Appl Genet 98:1091–1098

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Scotti I, Magni F, Fink R, Powell W, Binelli G, Hedley PE (1999) Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. Genome 43:41–46

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

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, Pulido JC, Yandava C, Sunden SL, et al. (1995) A collection of tri- and tetra-nucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Hum Mol Genet 4:1837–1844

Song QJ, Quigley CV, Nelson RL, Carter TE, Boerma HR, Strachan JL, Cregan PB (1999) A selected set of trinucleotide simple sequence repeat markers for soybean cultivar identification. Plant Var Seeds 12:207–220

Stuart-Rogers C, Flavell AJ (2001) The evolution of Ty1-copia group retrotransposons in gymnosperms. Mol Biol Evol 18:155–163

Taramino G, Tarchini R, Ferrario S, Lee M, Pe' ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Tautz D, Schlötterer C (1994) Simple sequences. Curr Opin Genet Dev 4:832–837

Tenzer I, degli Ivanissevich S, Morgante M, Gessler C (1999) Identification of microsatellite markers and thier application to population genetics of *Venturia inaequalis*. Phytopathology 89:748–753

Tozaki T, Inoue S, Mashima S, Ohta M, Miura N, Tomita M (2000) Sequence analysis of trinucleotide repeat microsatellites from an enrichment library of the equine genome. Genome 43:354–365

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1–6

Young ET, Sloan JS, Van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. Genetics 154:1053–1068