Theor Appl Genet (2003) 107:574–580
DOI 10.1007/s00122-003-1284-2

D. Bastien · J. M. Favre · A. M. Collignon · C. Sperisen · S. Jeandroz

Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce [*Picea abies* (L.) Karst.]

Received: 18 June 2002 / Accepted: 20 December 2002 / Published online: 15 May 2003

Abstract A mosaic minisatellite region has been identified in the mitochondrial genome of Norway spruce (*Picea abies*). The array was composed of three tandem repeats PaTR1 (32 bp), PaTR2a (26 bp) and PaTR2b (26 bp). PaTR2a and PaTR2b differed by one base substitution. The analysis of 92 trees covering the whole natural distribution area of the species allowed detection of 11 length variants ranging from 131 bp to 447 bp. This high intra-specific polymorphism relies on variation in the number of the tandem repeats. Population genetic parameters estimated among 14 populations suggested high population differentiation (Gst = 0.749). The phylogenetic analysis of the 11 sequenced length variants has been performed using a parsimony approach. The topology of the tree showed a good association of groups with geographical origin and a low level of size homoplasy. The phylogenetic reconstruction also suggests that this minisatellite locus has mainly evolved by an increase in the repeat copy number.

Keywords Minisatellite · VNTR · mtDNA · Heteroplasmy · *Picea abies*

Electronic Supplementary Material Supplementary material is available for this article if you access the article at http://dx.doi.org/10.1007/s00122-003-1284-2. A link in the frame on the left on that page takes you directly to the supplementary material.

Communicated by D.B. Neale

D. Bastien · J. M. Favre · A. M. Collignon · S. Jeandroz (✉)
UMR UHP/INRA, 1136, Tree-Microbe Interactions,
Université Henry Poincaré, Nancy I,
54506 Vandoeuvre Lès Nancy, France
e-mail: jeandroz@scbiol.u-nancy.fr
Tel.: +33-383-684230

C. Sperisen
Eidgenössische Forschungsanstalt WSL, Zürcherstrasse 111,
8903 Birmensdorf, Switzerland

Introduction

During the last decades, important progress haves been made in the study of the general organization and evolution of the circular mitochondrial genome of animals. Animal mitochondrial DNA (mtDNA) is characterized by a maternal, non-recombining mode of inheritance and a rapid evolution rate leading to high intra-specific polymorphism. As a result, the mitochondrial genome has been source of genetic markers extensively used in population genetics and phylogeography (Avise 1994). The sequences of the complete mitochondrial genome of more than 70 vertebrate and invertebrate species are currently available in data bases (http://www3.ebi.ac.uk./Research/Mitbase/coll.html).

Comparatively, the mitochondrial genome of plants has been less investigated. Only three complete sequences of land plants have been published, namely that of *Marchantia polymorpha* (Hepatophyta), *Arabidopsis thaliana* and *Beta vulgaris* (Anthophyta) (Oda et al. 1992; Unseld et al. 1997; Kubo et al. 2000). Furthermore, the complex organization and low rate of substitution of this molecule have strongly hampered its possible exploitation in population genetics (Palmer and Herbon 1988; Laroche et al. 1997). Intra-specific polymorphism due to large structural rearrangements as revealed by restriction Fragment length polymorphism (RFLP) probes has actually been observed in different species, but examples of convergent evolution raising doubts in data interpretation have also been mentioned (*Hevea brasiliensis*, Luo et al. 1995; *Pinus spp*, Wu et al. 1998). Point mutations would be phylogenetically more informative, but their detection requires highly sensitive methods such as Single-strand conformation polymorphism (SSCP) or sequencing.

In this general context, the identification of polymorphic mitochondrial VNTR (variable number of tandem repeats) loci could open useful possibilities. VNTR sequences are usually classified into two categories according to the length of the repeat unit: microsatellites (repeat unit of 1–5 bp) and, minisatellites (repeat unit of

6–64 bp) (see Estoup and Angers 1998, for a review). They are highly represented in the nuclear genome of many eukaryotes, but have also been described in organelle genomes.

In plants, Powell et al. (1995) were the first to demonstrate the presence of organellar microsatellites in the chloroplast DNA (cpDNA) of several pine species. More recently, it has been shown that the mtDNA of beets (*Beta vulgaris* and *B. maritima*, Nishizawa et al. 2000), Norway spruce (*Picea abies*, Sperisen et al. 2001) and several pine species (Mitton et al. 2000) could also contain tandem repeats, mainly minisatellites. To date, however, occurrence of microsatellite loci haves not yet been mentioned in the mtDNA of plants, except in the region between the genes encoding *nad3* and *rps12* in *Pinus spp*. which includes a polyG(n) (Soranzo et al. 1999).

Two contrasting models are traditionally proposed to describe the molecular evolution of micro- and minisatellites regions: the stepwise mutation model (SMM) in which VNTR variations areis explained by the addition/subtraction of a single unit from the current allele size and the infinite allele model (IAM) where the current allele size can be changed by any number of units with constant mutating probability. Close to IAM, is the K allele model (KAM) in which any allele can mutate towards a K fixed number of new allele states. A mixed two-phase model (TPM) has also been described (Di Rienzo et al. 1994). Experimental data have shown that changes in one and/or several repeat units are common in nuclear minisatellites and could fit the SMM or IAM as well (Jeffreys et al. 1988; Estoup and Angers 1998 for review). Multiple and complex molecular mechanisms involving intra-allelic (slipped-strand mispairing, unequal sister chromatid exchange) and inter-allelic (unequal crossing over, gene conversion) events have been proposed to govern size variation in microsatellite replication slippage (Haber and Louis 1998; Buard and Jeffreys 2000).

Critical evaluation of VNTRs as practical tools has been made by Lunt et al. (1998) and Estoup and Angers (1998). The high variability of VNTRs offers numerous advantages for molecular ecology studies, although the possible occurrence of recurrent haplotypes or recurrent alleles (homoplasy) can be a strong limiting unfavorable factor. The probability of size homoplasy depends on the model of molecular evolution of VNTRs. IAM and KAM with large K value are expected to result in low or nil homoplasy because any new allele created by mutation is distinct from the existing ones. In contrast, SMM and TPM will favor homoplasy. In a recent study of cpDNA microsatellite variation in *Pinus contorta*, the occurrence of recurrent haplotypes has been demonstrated, resulting in a poorly resolved phylogeny (Marshall et al. 2002). The high mutation rate observed in mitochondrial minisatellites could also contribute to obscure the phylogenetic signal (Faber and Stepien 1998).

In the investigation reported here, we characterized a new polymorphic region composed of two tandem repeat motifs in the mtDNA of Norway spruce (*Picea abies*, a major conifer species widely distributed over plains and medium-altitude mountains of the European continent (Schmidt-Vogt 1974). Complementarily, we carried out a preliminary test on natural populations analysis and a genealogical study of the allelic molecular variation to evaluate the potential of this region to be used for investigating intra-specific variation in this forest tree species.

Materials and methods

Plant material and DNA extraction

Molecular analyses were carried out on 92 trees originating in the three domains usually distinguished within the natural distribution area of *Picea abies* (Schmidt-Vogt 1974; Huntley and Birks 1983): 30 from the baltico-nordic (Fennoscandia and European Russia), 32 from the hercyno-carpathian (Central Europe and Carpathian mountains) and 30 from the alpine (alpine and dinaric massifs) domains. Each tree represented a distinct provenance (see A1). They were sampled in the French IUFRO provenance test 1964/1968 (Krutzsch 1992) located in the Amance forest near Nancy (Lorraine, north-eastern France). Complementarily, we analysed 14 populations distributed over the whole natural range of the species, each represented by ten trees. Material was collected in the Hungarian IUFRO provenance test (ten provenances) located in the vicinity of Màtrafured and four natural populations (see A1).

Total DNA was extracted from needles according to the procedures described by Collignon and Favre (2000) or by Sperisen et al. (2000).

Polymerace chain reaction (PCR) amplification and electrophoresis

The polymorphic mitochondrial region (locus *mh44*) was identified using the specific mtDNA primers 44F (ATGACTGGAAGAATTGCTCAC) and 44R (TTCACTTGATACTCACCCCC) (Jeandroz et al. 2002). Each amplification reaction contained 10 m*M* Tris-HCl, 50 m*M* KCl, 100 μM each of dNTPs, 2 m*M* $MgCl_2$, 0.2 μM of primers, 25 ng of genomic DNA template and 0.5 U of *Taq* DNA polymerase (Invitrogen, Carlsbad, Calif.) in a total volume of 25 μl. Amplification was carried out in a DNA thermal cycler (BioRad iCycler, Hercules, Calif.) with the following program: an initial denaturation at 94 °C for 4 min, then 35 cycles of 45 s at 94 °C, 45 s at 56 °C, 45 s at 72 °C and a final elongation at 72 °C for 10 min. Five microliters of the PCR products were electrophoresed on a 2% agarose gel, then stained with ethidium bromide. Gels were photographed under UV light, and fragment sizes were estimated using the Base image IR software (LiCor, Lincoln, Neb.).

Cloning and sequencing of the PCR products

PCR products of different length were cloned using the TOPO-TA cloning kit (Invitrogen) according to the manufacturer's recommendations. DNA sequencing was performed from one clone per each variant by Genome express (Grenoble, France) using the M13 universal primers. Sequences were aligned with the PILEUP and LINEUP software of the GCG package (Devereux et al. 1984).

Data analysis

Population genetic parameters (Hs, Ht and Gst) were calculated using the HAPLODIV software (Pons and Petit 1995). Phylogenetic analyses of size variants were estimated using parsimony analysis with PAUP 3.1.1 (Swofford 1993). Data sets were analyzed using the

heuristic search. Characters were considered as ordered or unordered to recreate the constraints imposed by SMM or IAM. a 50% majority rule consensus tree, rooted trees using the mid-point rooting option wereas chosen to represent the phylogeny.

Results

Identification and variation of minisatellite sequences

The analysis of the 92 single trees at the *mh44* locus allowed identification of 11 length variants ranging from 131 bp to 447 bp. An example of this intra-specific length polymorphism is shown Fig. 1. Nine individuals giving non-exploitable amplification (no amplifications or faints bands) profiles were discarded from the analyses. These 11 length variants were then cloned. Sequencing of the cloned fragments (accession numbers AY184274–AY184284 in Genebank database) revealed the presence of two repeat units of different length: PaTR1 (32 bp) and PaTR 2 (26 bp) (Fig. 2). One base substitution at the first position allowed us to further differentiate PaTR2a and PaTR2b (Fig. 2A). These three repeat units are characterized by a consensus sequence [GGG(T,A)GAGGAAGAA] at their 3′ end. Remarkably, the last six of bases of this consensus sequence (GAAGAA) were identical to the six bases immediately upstream of the minisatellite array (Fig. 2).

Intra-locus organization of these repeat units and alignment of the different length variants is shown in Fig. 3. The shortest length variant (131 bp) is composed

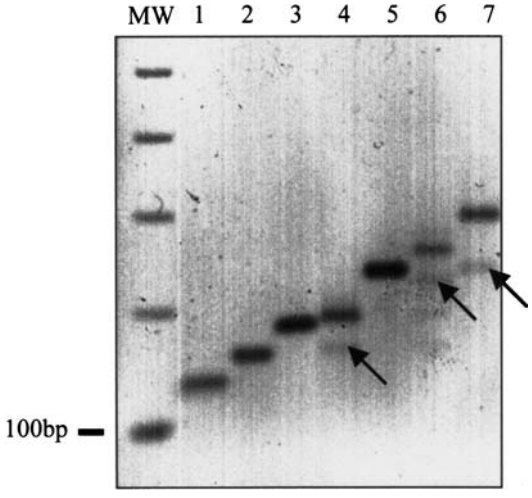

Fig. 1 Ethidium bromide-stained 2% agarose gel after electrophoresis of the PCR products derived from the *mh44* locus of seven *Picea abies* trees (*lanes 1–7*). *MW* 100-bp-size ladder (Invitrogen). *Arrows* indicates "extra bands"

A.

PaTR1	TTGCTCACCTGAAGCATGAGGGTGAGGAAGAA	32bp
PaTR2a	CTGCTCGACCAACGGGAGAGGAAGAA	26bp
PaTR2b	TTGCTCGACCAACGGGAGAGGAAGAA	26bp

B.

```
                    ATGACTGGAAGAA    13
TTGCTCACCTGAAGCATGAGGGTGAGGAAGAA     45
      CTGCTCGACCAACGGGAGAGGAAGAA     71
TTGCTCACCTGAAGCATGAGGGTGAGGAAGAA    103
TTGCTCACCTGAAGCATGAGGGTGAGGAAGAA    135
TTGCTCACCTGAAGCATGAGGGTGAGGAAGAA    167
TTGCTCACCTGAAGCATGAGGGTGAGGAAGAA    199
      TTGCTCGACCAACGGGAGAGGAAGAA    225
      TTGCTCGACCAACGGGAGAGGAAGAA    251
      TTGCTCGACCAACGGGAGAGGAAGAA    277
      TTGCTCGACCAACGGGAGAGGAAGAA    303
TGCAAGCATAACTAGATAAGATGGGTCTTCCA    335
AATGAGCTGCCATCACCGATTATGAACTCCGA    367
TGGGGGTGAGTATCAAGTGAAA              389
```

Fig. 2 **A** Nucleotide sequence of the three tandem repeats (PaTR1, PaTR2a and PaTR2b) identified in the 11 length variants at locus *mh44* in the different size variants. **B** Complete sequence of the 389-bp variant (individual 240). The 6-bp motif located upstream the array (*in bold*) and at the end of each repeat unit is *underlined*

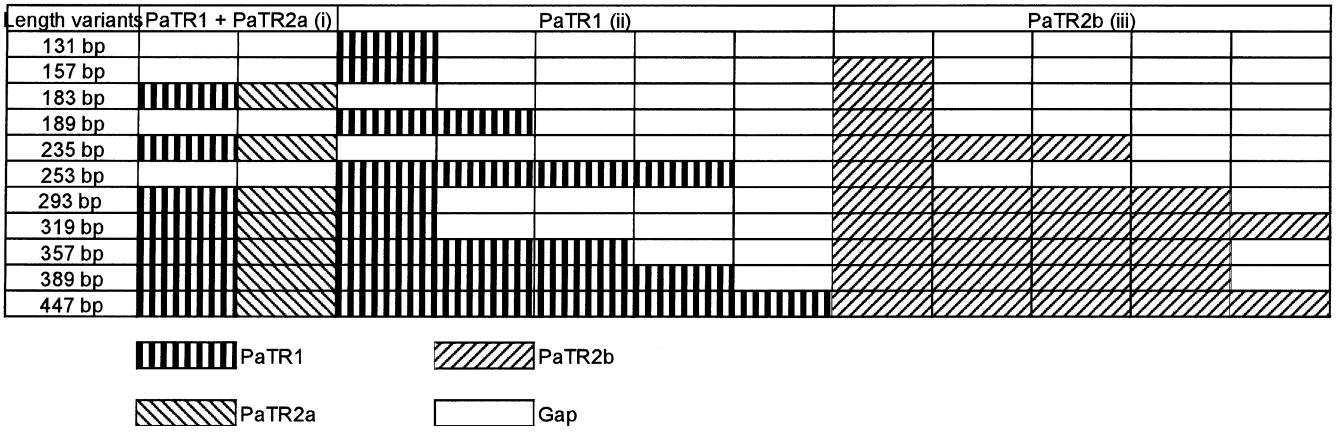

Fig. 3 Schematic representation of the minisatellite region found in the 11 length variants at locus *mh44*. The diagram describes the number, type, interspersion pattern and alignment of the repeat units (PaTR1, PaTR2a and PaTR2b). The different repeat motifs of the minisatellite region are represented by *patterned boxes*. *White boxes* represent gaps introduced for alignment and corresponding to indels of the repeated motifs. (*i*), (*ii*) and (*iii*) refer to the three characters used for parsimony reconstruction of length variants evolution (see Results)

Table 1 Variants frequencies (%) at the *mh44* locus in the 14 *Picea abies* populations investigated and number of trees assayed (*n*)

Population	Country	*n*	Domain	Variant length (bp)									
				131	157	183	235	293	319	335	357	366	447
1	Austria	8	Ap	0.0	0.0	0.0	87.5	0.0	0.0	0.0	12.5	0.0	0.0
2	France	10	Ap	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
3	France	10	Ap	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Italy	10	Ap	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
5	Serbia	10	Ap	0.0	0.0	0.0	10.0	0.0	90.0	0.0	0.0	0.0	0.0
6	Belarus	8	B	0.0	75.0	0.0	12.5	12.5	0.0	0.0	0.0	0.0	0.0
7	Poland	10	B	0.0	90.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0
8	Russia	10	B	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	Sweden	10	B	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	Sweden	9	B	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	Germany	8	HC	0.0	0.0	0.0	75.0	0.0	0.0	0.0	0.0	12.5	12.5
12	Poland	9	HC	0.0	0.0	0.0	0.0	0.0	11.1	0.0	88.9	0.0	0.0
13	Poland	9	HC	0.0	0.0	11.1	11.1	0.0	33.3	44.5	0.0	0.0	0.0
14	Romania	10	HC	90.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0

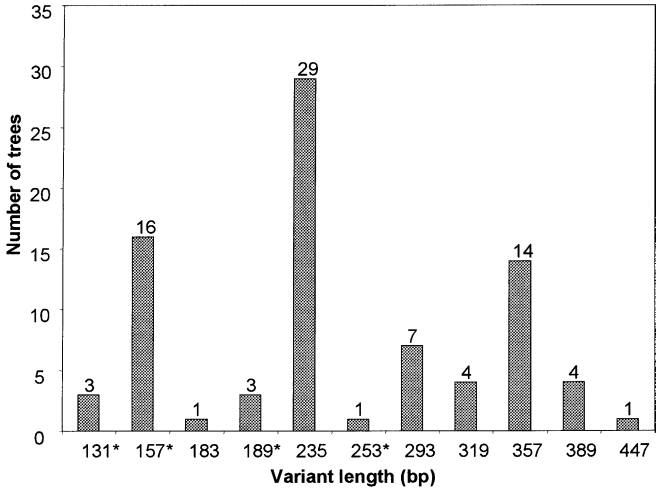

Fig. 4 Histogram showing repartition of the 11 size variants isolated in 92 *P. abies* trees from the whole natural distribution area. *Columns* indicate the number of individuals for each size variant. *Asterisk* indicates the size variants only observed in the baltico-nordic domain

of a single PaTR1 repeat. The variants of 157 bp, 189 bp and 253 bp comprise a series of $(PaTR1)_{1–4}$ and $(PaTR2b)_1$. The remaining seven variants contain the three repeat units in different combination patterns: $(PaTR1)_1$ + $(PaTR2a)_1$ + $(PaTR1)_{1–5}$ + $(Pa\ TR2b)_{4–5}$ in the 293-, 319-, 357-, 389-, and 447-bp variants, respectively, and $(PaTR1)_1$ + $(PaTR2a)_1$ + $(PaTR\ 2b)_{1–3}$ in those of 183 bp and 235 bp. Base substitutions were not detected in the regions flanking the array.

The different length variants detected were unequally represented among the 83 trees analyzed. Figure 4 shows the number of individuals for each class of variants. Three size classes, were dominant, namely those of 157, 235 and 357 bp. Interestingly, each of these seemed to be preferentially associated to one specific domain of the species' natural range. The 157-bp variant was exclusively found in the baltico-nordic provenances (16/16), while the 235- and 357-bp variants were mainly observed in the alpine (25/29) and hercyno-carpathian (11/14) provenances, respectively.

Several individual profiles (10/83) consisted of more than one length variant (Fig. 1). Their electrophoretic pattern showed one main band and one or two additional faint fragments. These extra bands were cloned and sequenced. The sequences revealed a high similarity with the corresponding main band, differing only by the number of PaTR1 or PaTR2 repeats.

Of the 11 length variants detected in the 83 single trees, nine were also found in the 14 populations. One additional length variant of 366 bp was found (not sequenced). In addition, five trees were found to show intra-individual variation. In this case, the dominant size variants were taken to be representative. Two individuals showing ambiguous profiles (two or three bands of same intensity) were discarded from the analyses. The frequencies of the length variants in the populations are given in Table 1. In each population, one dominant length variant was observed (frequency $\geq$ 75%), except in the case of population 13 (two dominant variants). Population genetic parameters estimated at this *mh44* minisatellite locus were: Hs = 0.196 (±0.061), Ht = 0.782 (±0.060) and Gst = 0.749 (±0.067).

Parsimony reconstruction of size variants evolution

The organization pattern of the repeat units in the 11 length variants was encoded using three characters (Fig. 3). The first character, (i), was the presence (encoded 1)/absence (0) of the initial $(PaTR1)_1$ + $(PaTR2a)_1$ repeat association. The second, (ii), and third, (iii), characters were the number of PaTR1 and PaTR2b repeats in the middle and 3′ terminal regions of the array, respectively. These two latter characters displayed six or five states, respectively. Following this method, the 131-bp variant was encoded 010 and the 447-bp variant, 155. A double weight was assigned to character (i) considering that the coincident duplication of PaTR1 + PaTR2a is a priori more informative than the simple variable number

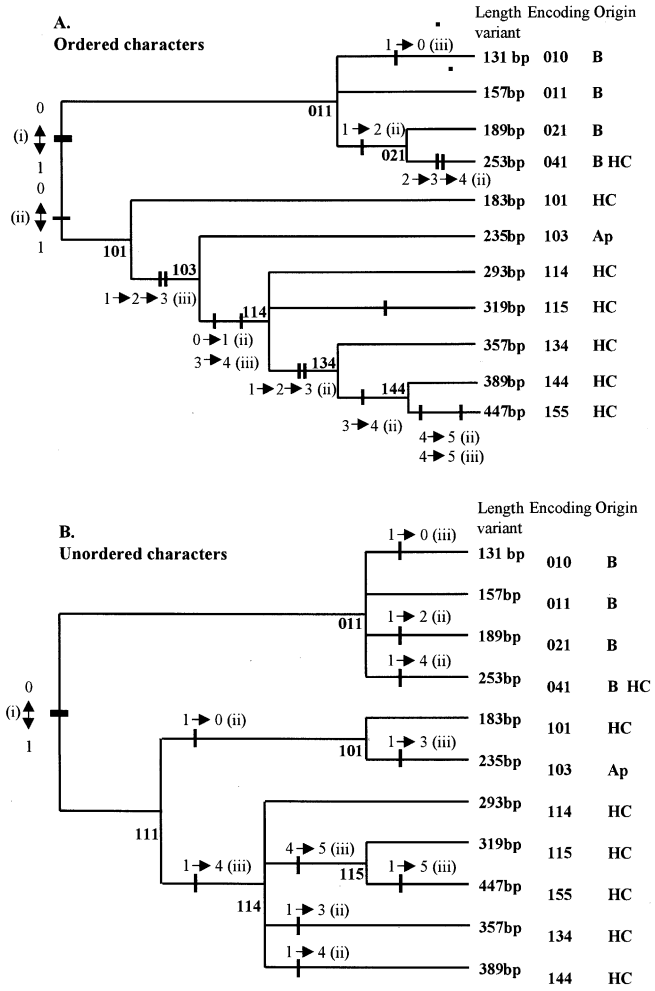

Fig. 5 **A** Most parsimonious tree obtained from analysis of the 11 size variants detected at the minisatellite locus *mh44* when characters are treated as ordered. **B** 50% majority rule consensus tree obtained from parsimony analysis of the 11 size variants detected at the minisatellite locus *mh44* when characters are treated as unordered. Characters (*i*), (*ii*) and (*iii*) used for phylogenetic analysis are indicated in Fig. 3 (see Results for a description). Length of variants, corresponding encoding of characters and geographical origin of trees are indicated. *B* Baltico-nordic domain, *HC* hercyno-carpathian domain, *A* alpine domain. Changes from one to another character state are indicated on each branch

of each repeat unit [characters (ii) and (iii)]. These three characters were firstly treated as ordered characters under the assumption that to get from one state to another state, the character must proceed necessarily through intermediate states. Then, in a second step, the characters were considered to be unordered, i.e. capable of moving directly to any other state. In other words, a character in state 0 can change to state 1, 2, 3, 4 or 5 with the equal probability of 1.

The phylogenetic trees obtained are presented in Fig. 5. The first analysis (ordered characters) led to one single parsimonious tree of 17 steps (Fig. 5A). Size variants are distributed into two major groups characterized by the presence or absence of the associated (PaTR1) + (PaTR2a) repeats at the 5′ end of the array and one change from state 0 to state 1 for the second character. Lineages with 011 and 101 variants can be considered to be putative common ancestors of these two groups. Recurrent changes in the second and third characters led to several homoplasic changes (consistency index 0.706). The second analysis (unordered characters) resulted in 32 most parsimonious trees of (12 steps. The resulting 50% majority rule consensus tree is presented shown in Fig. 5B. Size variants were again distributed into the same two groups but only separated by the presence or absence of the associated (PaTR1) + (PaTR2a) repeats. These groups corresponded to two major lineages with 011 and 111 variants as putative common ancestors. Multiple combinations between the three characters were observed, leading to lower homoplasy than in the first analysis (consistency index 0.917). Indeed, only one homoplasic change could be observed for the second character (state 1 to state 4 in the 253- and 389 bp variants).

Interestingly, the two groups of length variants observed in both phylogenetic reconstructions corresponded to two different geographic origins: the length variants of the first group (131, 157, 189 and 253 bp) were found in individuals originating in the northern and north-eastern parts of the natural distribution area of the species (baltico-nordic and north of hercyno-carpathian domains), while the second group was composed of trees of central and south-eastern Europe (alpine and hercyno-carpathian domains).

Discussion

The length polymorphism detected at the *mh44* locus of the mitochondrial genome of *Picea abies* results from the variable number of two tandem repeats of 32 bp (PaTR1) and 26 bp (PaTR2), respectively, that occur in variable combinations within the array. The lengths of these repeats allows us to classify them as minisatellites. Sequence similarity between PaTR1 and PaTR2 suggests that they share a common origin. The BLAST survey of nucleotide sequences (Genbank and EMBL databases) indicated that the first 21-bp fragment of PaTR2 is similar to the mitochondrial tandem repeat TR2 (32bp) identified in beets (Nishizawa et al. 2000). In addition, it is interesting to notice that the presence of a short 5- to 6-bp motif upstream of the array and at the 3′ end of each repeat unit (GAAGAA in the case of *mh44*) has already been observed in many nuclear minisatellites of yeast and human (Haber and Louis 1998) as well as in chloroplast and in other plant mitochondrial minisatellites (Nishizawa et al. 2000; Cafasso et al. 2001). According to Haber and Louis (1998), such a motif could be involved in a molecular mechanism generating generation of the variation in the number of repeat units through replication slippage or unequal crossing-over.

The mosaic organization of this minisatellite locus is original and differs from the two other cases reported to date in plants which, interestingly, are also found in

coniferous species, i. e. *Pinus ponderosa* (Mitton et al. 2000) and *P. abies* itself (Sperisen et al. 2001). In both species, the mosaic minisatellites were located within the *nad1b/c* mitochondria locus. In *P. ponderosa*, the array mainly consisted of two alternating tandem repeats of 32 bp and 34 bp, while in *P. abies*, two homogeneous adjacent arrays of 32 bp and 34 bp repeats occurred.

The estimates of population genetic parameters obtained here are in good agreement with the theoretical expectation for maternally inherited genomes (Petit et al. 1993). Indeed, the within-population diversity is indeed low, with one size variant fixed in six out of the 14 populations analyzed. Moreover, genetic differentiation among populations at the *mh44* locus appeared to be higher than that of biparentally inherited markers such as allozymes (Gst = 0.052; Lagercrantz and Ryman 1990) and random amplified polymorphic DNA (Gst = 0.063; Collignon and Favre 2000) or paternally inherited cpDNA markers (Rst = 0.11; Vendramin et al. 2000).

Intra-individual size variation was identified in about 10% of the trees analyzed. Intra-individual variation of VNTR loci is common in mtDNA of animals and usually interpreted to be a consequence of heteroplasmy, which is the coexistence of several differing mtDNA molecules in each individual cell (see Lunt et al. 1998 for a review). This phenomenon has been little documented in plants. It has been attributed to biparental inheritance of chloroplasts in *Chamaecyparis obtusa* (Cupressaceae) (Shiraishi et al. 2001). Heteroplasmy in mtDNA can also arise from length variation generated by recombination across repeated sequences (André et al. 1992). For instance, in maize, the NCS2 mutant contains a mixture of mutated and normal mitochondrial genomes. The NCS2 mutation is lethal, and the surviving NCS2 plants are heteroplasmic (Yamato and Newton 1999). In this case, heteroplasmy is uniparentally (maternally) inherited. In *P. abies* also, the mtDNA is maternally inherited (Grivet et al. 2000). Therefore, in this species, heteroplasmy would be unlikely to result from biparental origin of the mitochondrial genome but is more probably the consequence of intra- or intermolecular recombination events between mtDNA submolecules. The analysis of controlled crosses should enable the accumulation of more information on the nature of these intraindividual variation.

The phylogenetic analysis of the 11 sequenced length variants was performed using a parsimony approach. Characters were treated (1) as ordered to recreate the constraints imposed by the SMM model of VNTR molecular evolution (Marshall et al. 2002) and (2) as unordered to simulate the constraints of the IAM or KAM, which are better well suited for microsatellite arrays containing several types of repeat units, as observed in the *mh44* locus (Lunt et al. 1998). The two resulting phylogenetic trees areis quite similar to that obtained from the polymorphic sites of the *nad1*b/c intron (Sperisen et al. 2001) which distinctively divides *P. abies* populations into two lineages. As predicted by the IAM model, a low level of size homoplasy was observed in the second tree (unordered characters). The size distribution of variants (Fig. 4) in each lineage was intermediate between the expectations of the SMM and IAM (one or multiple steps are involved to explained phylogenetic relationships between variants). This suggests that mutation processes are certainly more complex than has been assumed in these two extremes models. The topology of the phylogenetic trees also showed a good association of groups with geographical origin, with one lineage corresponding to the baltico-nordic populations plus some trees from the northern limit of the hercyno-carpathian domain and the other one to the total alpine and the majority of the hercyno-carpathian populations. This strong congruence between phylogeny and geography is in close agreement with what couldan be expected from diversity analyses developed using maternally inherited genetic markers in forest trees (Dumolin Lapègue et al. 1997; Newton et al. 1999). The genetic variation at the locus *mh44* can be considered to be phylogenetically informative. Furthermore, it is important to note that the hypothetical ancestors determined at the root of each lineage correspond to short length variants. This strongly suggests that the *mh44* locus has mainly evolved by an increase in the repeat copy number. Only two changes resulted from a decrease of the repeat copy number. Interestingly, the prevalent length variant in the baltico-nordic populations (157 bp) was found to be in the position of the hypothetical ancestor of the corresponding lineage in both phylogenetic trees. However, results did not allow us to determine which variant was closest to the common ancestor of the lineages. Further investigation of the molecular evolution of these length variants will be needed to elucidate more precisely the phylogenetical relationships between the *P. abies* populations.

In conclusion, it clearly appears that the mitochondrial locus analyzed in this study displays multiple VNTR length variants well-suited for intra-specific discrimination of populations that have recently diverged. In addition, the low level of homoplasy and the high population differentiation observed confirm its utility for phylogeography analysis. Easily scorable on agarose gels, this maternally inherited locus should prove to be useful for diversity studies in this tree species.

Acknowledgements This study has been financially supported by the CTPS (Comité Technique Permanent de la Sélection des Plantes cultivées. France). CS was financially supported by the Swiss Federal Office for Education and Science through the participation in the EU-project "Dynamics of forest tree biodiversity: linking genetic, palaeogenetic and plant historical approaches" (EVK2-1999-00036). We would like to thank M.F. Arnould, A. André, and R. Kissling for their technical assistance, and G. Besnard and two anonymous reviewers for their constructive comments.

References

André C, Levy A, Walbot V (1992) Small repeated sequences and the structure of plant miochondrial genomes. Trend Genet 8:128–132

Avise JC (1994) Molecular markers, natural history and evolution. Chapman & Hall, New York

Buard J, Shone AC, Jeffreys AJ (2000) Meiotic recombination and flanking marker exchange at the highly unstable human minisatellite CEB1 (D2S90). Am J Hum Genet 67:333–344

Cafasso D, Pellegrino G, Musacchio A, Widmer A, Cozzolino S (2001) Characterization of a minisatellite repeat locus in the chloroplast genome of *Orchis palustris* (Orchidaceae). Curr Genet 39:394–398

Collignon AM, Favre JM (2000) Contribution to the postglacial history at the western margin of *Picea abies* natural area using RAPD markers. Ann Bot 85:713–722

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the WAX. Nucleic Acids Res 12:387–385

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Dumolin Lapègue S, Demesure B, Fineschi S, LeCorre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics 121:1475–1487

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: Carvalho GR (ed) Advances in molecular ecology. IOS Press, pp 55–86

Faber JE, Stepien CA (1998) Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the pike-perches *Stizostedon*. Mol Phylogenet Evol 10:310–322

Grivet D, Jeandroz S, Favre JM (1999) *Nad1*b/c intron polymorphism reveals maternal inheritance of the mitochondrial genome in *Picea abies*. Theor Appl Genet 99:346–349

Haber JE, Louis EJ (1998) Minisatellite origins in yeast and humans. Genomics 48:132–135

Huntley B, Birks HJB (1983) An atlas of past and present pollen maps for Europe, 0–13,000 years ago. Cambridge University Press, Cambridge

Jeandroz S, Bastien D, Chandelier A, Du Jardin P, Favre JM (2002) A set of primers for amplification of mitochondrial DNA in *Picea abies* and other conifer species. Mol Ecol Notes 2:389–392

Jeffreys AJ, Royle NJ, Wilson V, Wong Z (1988) Spontaneous mutations rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 3332:278–281

Krutzsch P (1992) IUFRO's role in coniferous tree improvement: Norway spruce (*Picea abies* (L) Karst.). Silvae Genet 41:143–150

Kubo T, Nishizawa S, Sugarawa A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$ (GCA). Nucleic Acids Res 28:2571–2576

Lagercrantz U, Ryman N (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. Evolution 44:38–53

Laroche J, Li P, Maggia L, Bousquet J (1997) Molecular evolution of angiosperm mitochondrial introns and exons. Proc Natl Acad Sci USA 94:5722–5727

Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Mol Ecol 7:1441–1455

Luo H, Van Coppenolle B, Seguin M, Boutry M (1995) Mitochondrial DNA polymorphism and phylogenetic relationships in *Hevea brasiliensis*. Mol Breed 1:51–63

Marshall HD, Newton C, Ritland K (2002) Chloroplast phylogeography and evolution of a highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). Theor Appl Genet 104:367–378

Mitton JB, Kreiser BR, Rehfeldt GE (2000) Primers designed to amplify a mitochondrial nad1 intron in ponderosa pine, *Pinus ponderosa*, limber pine, *P. flexilis*, and Scots pine, *P. sylvestris*. Theor Appl Genet 101:1269–1272

Newton AC, Allnut TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intra-specific variation and the conservation of tree species. Trends Ecol Evol 14:140–145

Nishizawa S, Kubo T, Mikami T (2000) Variable number of tandem repeat loci in the mitochondrial genomes of beets. Curr Genet 37:34–38

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of Liverwort *Marchantia polymorpha* mitochondrial DNA. J Mol Evol 223:1–7

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol 28:87–97

Petit RJ, Wagner DB, Kremer A (1993) Finite island model for organelle and nuclear genes in plants. Heredity 71:630–641

Pons O, Petit RJ (1995) Estimation variance and optimal sampling of gene diversity. I. Haploid locus. Theor Appl Genet 90:462–470

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: application to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–7763

Schmidt-Vogt (1978) Monographie der *Picea abies* (L.) Karst. Unter Berück sichtigung genetischer und züchtensher Aspekte. Forst Wissenschaft Centralbl 97:281–302

Shiraishi S, Maeda H, Toda T. Seido K, Sasaki Y (2001) Incomplete paternal inheritance of chloroplast DNA recognized in *Chamaecyparis obtusa* using an intra-specific polymorphism of the trnD-trnY intergenic spacer region. Theor Appl Genet 102:935–941

Soranzo N, Provan J, Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. Genome 42:158–161

Sperisen C, Gugerli F, Büchler U, Mátyás G (2000) Comparison of two rapid DNA extraction protocols for gymnosperms for application in population genetic and phylogenetic studies. For Genet 7:133–136

Sperisen C, Büchler U, Gugerli F, Matyas G, Geburek T, Vendramin GG (2001) Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. Mol Ecol 10:257–263

Swofford DL (1993) PAUP: Phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign, Ill

Unseld M, Marenfield JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet 15:57–61

Vendramin GG, Anzidei M, Madaghiele A, Sperisen C, Bucci G (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). Genome 43:68–78

Wu J, Krutovskii KV, Strauss SH (1998) Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. Genetics 150:1605–1614

Yamato KT, Newton KJ (1999) Heteroplasmy and homoplasmy for maize mitochondrial mutants: a rare homoplasmic nad4 deletion mutant plant. J Hered 90:369–373