ORIGINAL ARTICLE

Evolution of an Ancient Microsatellite Hotspot in the Conifer Mitochondrial Genome and Comparison with Other Plants

Juan P. Jaramillo-Correa · Erika Aguirre-Planter · Luis E. Eguiarte · Damase P. Khasa · Jean Bousquet

Abstract In plants, mitochondrial sequence tandem repeats (STRs) have been associated with intragenomic recombination, a process held responsible for evolutionary outcomes such as gene regulation or cytoplasmic male-sterility. However, no link has been established between the recurrent accumulation of STRs and increased mutation rates in specific regions of the plant mtDNA genome. Herein, we surveyed this possibility by comparing, in a phylogenetic context, the variation of a STR-rich mitochondrial intron (*nad*5-4) with eleven mtDNA genes devoid of STRs within *Abies* (Pinaceae) and its related genera. This intron has been accumulating repeated stretches, generated by at least three-independent insertions, before the split of the two Pinaceae subfamilies, Abietoideae and Pinoideae. The last of these insertions occurred before the divergence of *Abies* and produced, exclusively within this genus, a tenfold increase of both the indel and substitution rates in the STR hotspot of the intron. The regions flanking the STRs harbored mutation rates as low as those estimated in mitochondrial genes devoid of repeated stretches. Further searches in complete plant mtDNA genomes, and previous studies reporting polymorphic mtSTRs, revealed that repeated stretches are common in all sorts of plants, but their accumulation in STR hotspots appears to be taxa specific. Our study suggests a new mutagenic role for repeated sequences in the plant mtDNA.

Keywords *Abies* · Ancestral polymorphism · Mitochondrial DNA · Mutation rates · *nad*5 · Sequence-tandem-repeats

J. P. Jaramillo-Correa (✉) · E. Aguirre-Planter · L. E. Eguiarte
Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad Universitaria, Circuito Exterior, Apartado Postal 70-275, Mexico, DF, Mexico
e-mail: jaramillo@ecologia.unam.mx; jpjc@gmx.net

J. P. Jaramillo-Correa · D. P. Khasa · J. Bousquet
Canada Research Chair in Forest and Environmental Genomics, Forest Research Centre and Institute for Systems and Integrative Biology, Université Laval, Quebec, QC G1V 0A6, Canada

Introduction

The evolution of the plant mitochondrial genome (mtDNA) is thought to be mainly driven by extensive structural reorganization, recurrent incorporation of foreign DNA material, and frequent gene transfer to the nucleus, while the roles of substitutions and small indel changes are believed to be rather minor, when compared to the animal or fungal mtDNAs (e.g., Palmer et al. 2000; Barr et al. 2005). However, recent reports of high mutation rates within specific lineages (Cho et al. 2004; Mower et al. 2007; Ran et al. 2010) suggest that this type of changes also play a role in the evolution of plant mtDNA. Surprisingly, the majority of these examples involve large indels up to 3,000 bp long, instead of the short-scale length variation (i.e., indels less than 100 bp) that is most commonly reported in plant population studies (e.g., Bastien et al. 2003; Jaramillo-Correa et al. 2003; Godbout et al. 2005, 2008).

In eukaryote genomes, small indels are often associated with repetitive sequences, such as microsatellites or mobile elements (e.g., Buschiazzo and Gemmell 2006). In the mtDNA of some plants, these repetitive sequences are so abundant that they have led to important size increases and structural rearrangements (e.g., Palmer et al. 2000; Lilly

Fig. 1 Schematic organization of the sequence tandem repeats (STRs) found in the fourth intron of the mitochondrial gene *nad*5 of the genus *Abies*. The sequence and position of each STR is indicated for the two most common haplotypes (12 and 18), and the sites where substitutions are observed in other haplotypes are indicated by *arrows*. *Note* the presence of STR-2 interrupting the first array on haplotype 12 and its absence in haplotype 18

and Harvey 2001; Hecht et al. 2011). This profusion raises the question of whether such repetitive elements tend to accumulate in specific regions of the plant mtDNA, and if these regions have higher mutation rates than other parts of the genome. Such higher rates might imply that repeated sequences do not only play a key role as promoters of recombination (e.g., Allen et al. 2007; Hecht et al. 2011), but also as mutagenic hotspots (e.g., Amos et al. 2008; McDonald et al. 2011). Indeed, mutation rate differences between repetitive and non-repetitive regions have been reported for the nuclear and chloroplast genomes of plants (e.g., Thuiller et al. 2002; Jackobson et al. 2007), and also for the animal mtDNA (e.g., Broughton and Dowling 1997; Wilkinson et al. 1997).

Sequence-tandem repeats (STRs), which include both micro and minisatellites, are formed by iterations of DNA-sequence motifs. In the mitochondrial genome of different taxa, they have been associated with diverse processes including recombination, gene regulation, and some disorders (e.g., Vergnaud and Denoeud 2000; Gemayel et al. 2010). In general, STRs can arise spontaneously from unique sequences or can be brought in by mobile elements; they grow by replication slippage and/or unequal crossover during recombination, until reaching a contraction–expansion equilibrium, and finally become interrupted by the proliferation of random mutations that break up the repeat pattern (reviewed by Buschiazzo and Gemmell 2006). Some studies (e.g., Mouhamadou et al. 2007; Morse et al. 2009) have reported that STRs tend to accumulate in specific regions of the genome. Such a pattern has been interpreted as the "renaissance" of previously interrupted motifs and/or as the recurrent insertion of repetitive elements in particular hotspots (Buschiazzo and Gemmell 2006). Nevertheless, only a few STR-rich regions have

been investigated in plants, most of them in the nuclear genome (e.g., Karhu et al. 2000; Chen et al. 2002).

In firs (*Abies* Miller, Pinaceae), a predominantly temperate group of conifers, the fourth intron of the mitochondrial gene *nad*5 (*nad*5-4) represents such a STR hotspot. It contains a complex array of up to four different repeated motifs (Fig. 1a), most of them shared by distantly related taxa from Asia, the Mediterranean, and Mesoamerica (e.g., Ziegenhagen et al. 2005; Jaramillo-Correa et al. 2008; Peng et al. 2012). The concentration of that amount of STRs in a region that is only ∼800 bp long, and the availability of population data to assess its length variation in different species (Liepelt et al. 2002, 2010; Ziegenhagen et al. 2005; Jaramillo-Correa et al. 2008, 2011; Jiang et al. 2011; Wang et al. 2011; Aguirre-Planter et al. 2012; Peng et al. 2012), provides a unique opportunity for surveying the potential role of small indels in the evolution of the plant mitochondrial genome.

In this study, we investigated the organization and evolution of this STR-rich mitochondrial intron in the genus *Abies* and its related taxa within the Pinaceae. First, we mapped its indel and substitution variation onto the phylogeny of this family, and estimated and compared its mutation rates with that of other mtDNA genes. Then, we correlated the available population data with the sequence variability (i.e., number of STRs, and repeats and substitutions per STR), and with some species distribution-traits (i.e., size and type of natural range, a proxy for species size), to examine if these variables are related to mtSTR diversity as reported for nuclear microsatellites in other species (e.g., Zhu et al. 2000; Nybom 2004). Finally, we searched the DNA sequences of some of the complete plant mtDNA genomes available to investigate how frequent STR hotspots are across taxa.

Materials and Methods

Sampling and Sequencing of mtDNA Genes

A total of 31 species and two varieties from the approximately 50 taxa currently recognized in the genus *Abies* (Liu 1971; Rushforth 1989; Farjon and Rushforth 1989) were considered together with one species of some of its related genera within the Pinaceae, namely: *Cedrus atlantica*, *Keteleeria evelyniana*, *Larix decidua*, *Picea abies*, *Pinus pseudostrobus*, *Pseudotsuga menziesii*, and *Tsuga canadensis* (Table S1). These taxa were selected in order to represent (i) as broad a phylogeographic and phylogenetic range as cross-amplification would allow, (ii) to sample the widest possible range of *nad*5-4 allele sizes, and (iii) to include both monomorphic and polymorphic taxa.

DNA was extracted with a CTAB mini-prep protocol (Vázquez-Lobo 1996) or a DNeasy Plant Mini Kit (Qiagen) from foliage collected in natural populations, arboreta, and botanical gardens (Table S1). PCR amplification was carried out in a PTC-225 thermal cycler (MJ Research) with the universal and internal primers and the conditions described elsewhere (Dumolin-Lapègue et al. 1997; Jaramillo-Correa et al. 2008). PCR products were examined by agarose gel electrophoresis (2 % in TAE) in order to verify that only one single band was amplified. Both DNA strands were then directly sequenced in an Applied Biosystems 3130xl DNA Genetic Analyser by using the appropriate primers, a Sequenase GC-rich kit (Applied Biosystems) and a dideoxynucleotide chain termination procedure. Ten additional mitochondrial regions (*ccm*C; *cox*1, *cox*3, *mat*R, *nad*1 b/c, *nad*3-*rps*12, *nad*4-3, *nad*5-1, *nad*7-1, and *SSU*-V1 region) were amplified and sequenced, for comparative purposes, following the same methods as above and using previously reported primers (see Jaramillo-Correa et al. 2003, 2008, and references therein). Sequences available in Genbank were downloaded to complete some of the datasets above (i.e., *nad*1 b/c, *nad*3-*rps*12, *nad*4-3, and *nad*5-4) and to build an additional one for the gene *atp*8 (Polezhaeva et al. 2010; Semerikova et al. 2011).

Sequence and Phylogenetic Analyses

Phylogenetic relationships were retrieved from a combined data set taken from previous studies representing the most recent calibrated phylogenies of the Pinaceae and the genus *Abies* (Gernandt et al. 2008; Xiang et al. 2009; Lin et al. 2010; Aguirre-Planter et al. 2012). DNA sequences from the mtDNA genes (including the regions flanking the *nad*5-4 STR-rich zone) were excluded from this analysis because of their low levels of polymorphism. In brief, within the genus *Abies,* only 14 parsimony informative sites were obtained for the eleven mtDNA regions sequenced

(representing more than 10,000 bp), which resulted in poorly resolved phylogenetic trees (data not shown).

DNA sequences of the gene *nad*5-4 were aligned with BioEdit (Hall 1999) and the SSRs were color-coded according to sequence resemblance (Fig. S1, see Barros et al. (2008) for an example in human SSRs). The different motifs observed were then traced onto the phylogenetic tree described above, and their putative ancestral states determined with the parsimony reconstruction method available in Mesquite (Maddison and Maddison 2011; see also Primer and Ellegren 1998 and Zhu et al. 2000). When more than one result was produced, the most parsimonious reconstructions were averaged. The order and age of each STR insertion, expansion, contraction, and interruption (i.e., substitutions) were then inferred by comparisons with previously reported divergence dates (Gernandt et al. 2008; Lin et al. 2010; Aguirre-Planter et al. 2012).

For each mtDNA region in *Abies*, the following indices were estimated: the mean number of pairwise nucleotide differences (π; Tajima 1983), the mutation-scaled effective population size derived from the number of segregating sites (θ_W; Watterson 1975), the insertion/deletion (I) and substitution (K_o) rates per site, and the average number of pairwise substitutions between species (D_{xy}). Calculations were made with DnaSP v. 5 (Librado and Rozas 2009) following Laroche et al. (1997), and using sequences from the Pinoideae subfamily (*Larix*, *Picea*, *Pinus*, and *Pseudotsuga*) as outgroups. For the particular case of *nad*5-4, such indices were also determined for each independent STR and both flanking regions.

Population mtDNA Variation and Comparative Analysis

An extensive literature survey was made in the ISI Web of Science and Scopus to gather information on the variation of *nad*5-4 in as many *Abies* taxa as possible (last update in July 2012). Only those studies that sequenced and published the observed variants were retained. Then, for each taxa, the mean number of mitotypes (nh) and the mean mtDNA diversity (H_E) was noted from the original publications, while the number of STRs, their length, and number of repeats were retrieved from the available DNA sequences. Previous works on nuclear SSRs (e.g., Zhu et al. 2000; Nybom 2004) showed that some species life-traits, such as the size and degree of fragmentation of the natural range, were related to the length and number of repeats, and variability of microsatellites. Thus, in order to test if such relationships could be also observed at the mtDNA level, the natural distribution of each retained taxon was classified as wide or restricted, and continuous or fragmented. Then, differences between categories were determined with two-independent Student's *t* tests, while the

relationships between the number of repeats, and mean nh and H_E per species were tested with Pearson's correlation coefficients (Zar 2010; see results).

Database Search for Other STR-Rich mtDNA Regions

A second literature search was performed for identifying studies reporting additional polymorphic plant mtSTRs and investigating if STR hotspots were common in other taxa. Again, only the cases with sequenced and published variants were retained, but in this case, only the type of STR (micro- or minisatellite, perfect or imperfect, and simple or compound) and the region of occurrence were noted. This data set was complemented by an intensive search of repeated motifs in 33 of the complete plant mitochondrial genomes available in GenBank (see Table 1), which was performed with the tandem repeats finder (Benson 1999) by using the default parameters. Only those sequence blocks with more than 80 % similarity and repeated at least five times where considered as STRs. When more than one sequence was available for a given genus, only the longest one was analysed (e.g. *Zea mays* spp. *mays* genotype CMS-C and *Oryza rufipogon*). The regions of occurrence for each STRs detected were determined from the original notation of each genome, complemented with blast searches, and compared across taxa.

Results

Description of the *nad*5-4 mtSTR Hotspot

Depending on the individual and species surveyed, the whole intron spanned between 832 and 1012 base pairs (bp), and contained a STR-rich region covering between 156 and 336 bp. No heteroplasmy was observed by electrophoresis, cloning or sequencing of PCR products. All the newly generated sequences for *Abies* and the other Pinaceae are available in Genbank (Accession nos. KC578686–KC578818).

Within the genus *Abies*, the extensive sequencing of 135 individuals revealed 24 different haplotypes that harbored between two and four STRs (Fig. 1). The first of these repeats was an imperfect array with two to eight repetitions with the basic motif CTATAT. The fifth of these repeats extended into a small and monomorphic AT succession, while one G/T and four A/C substitutions were also observed in different repeat copy members across haplotypes (Fig. S1). The second STR was an imperfect stretch of eight to twelve iterations of the basic sequence GATA. It was exclusive to the haplotypes harboring more than three copies of the first microsatellite (haplotypes 1–16), which was interrupted by these motifs (Fig. 1). The first three copy members of this stretch extended into small and

Table 1 Diversity indices (nh, π, θ_w), indel and substitution rates (I, K_o), and average pairwise divergence (D_{xy}) for 12 mitochondrial DNA regions across species in the genus *Abies*

Gene region	Length (bp)	N	nh	π	θ_w	I	K_o	D_{xy}^{a}
*nad*5 intron 4-STR-flanking regions	653	29	1	0	0	0	0	0.011
*nad*5 intron 4-STR1	61	29	18	0.02018	0.02577	0.0656	0.0819	–
*nad*5 intron 4-STR2	64	29	16	0.00855	0.01264	0.0983	0.0313	–
*nad*5 intron 4-STR3	31	29	2	0	0	0.0323	0	0.076
*nad*5 intron 4-STR4	31	29	2	0	0	0.0323	0	0.075
*nad*5 intron 4 Total	840	29	24	0.00475	0.00640	0.0352	0.0189	0.032
atp8	757	27	5	0.00221	0.00271	0	0.0053	0.054
*ccm*C	388	11	2	0.00258	0.00258	0	0.0026	0.010
*cox*1	576	29	2	0.00014	0.00046	0	0.0017	0.048
*cox*3	621	29	3	0.00024	0.00076	0	0.0032	0.017
*mat*R	474	22	1	0	0	0	0	0.009
*nad*1 intron b/c	1814	29	9	0.00033	0.00131	0.0017	0.0022	0.016
*nad*3–*rps*12 i.r.[b]	252	24	1	0	0	0	0	0.032
*nad*4 intron 3	1791	16	4	0.00086	0.00096	0.0033	0.0028	0.029
*nad*5 intron 1	698	29	5	0.00363	0.00439	0.0043	0.0057	0.017
*nad*7 intron 1	716	29	2	0.00002	0.00039	0.0014	0.0013	0.012
SSU *rRNA* V1 region	458	29	3	0.00297	0.00324	0.0022	0.0059	0.035

Estimates for the STR-rich *nad*5-4 were made for the whole intron, for each independent STR, and for the flanking regions

[a] Determined using sequences from the Pinoideae subfamily (*Pinus*, *Larix*, *Pseudotsuga* and *Picea*) as outgroups

[b] Intergenic region

monomorphic AT successions with some T/G substitutions in three particular haplotypes (5–7). The third and fourth STRs were both imperfect AT repeats (Fig. 1) that were located 30 and 100 bp downstream the end of the first STR, respectively. The former was present in all but three haplotypes (14, 15, and 21) and harbored one TTT and two AAA motifs that interrupted an array of six to nine repeats. The later was observed in all 21 haplotypes and was formed by 12 or 13 repetitions that were disrupted by one TT and one AA motif, and one A/C substitution (Figs. 1, S1).

Most of the intron, especially the sequences flanking the STRs, appeared to be well conserved within the Pinaceae (Fig. S1). No within-species polymorphism was observed outside the genus *Abies*, in spite of an intensive screening of the natural distributions of three additional taxa (*Pinus pseudostrobus*, *Picea abies*, *Tsuga canadensis*; data not shown). The basic sequence pattern of the first (1–3 repeats), third (9–13 repeats), and fourth (12–13 repeats) STRs were present in all species surveyed except *Picea abies*, while a proto-STR (2 repeats) of the second motif was observed in *Tsuga*. Blast searches and visual comparisons with the homologous intron of other plants revealed that the regions flanking the STRs were also well conserved in the gymnosperm *Cycas taitungensis* and in many angiosperms. However, although a proto-STR (1 repeat) of the first array, together with four imperfect AT-tandem repeats, were detected in *Cycas*, the zone containing the tandem stretches was highly variable outside the Pinaceae and was virtually impossible to align (data not shown). Interestingly, one of the tandem arrays found during the analyses of the complete mtDNA genomes of *Zea* and *Tripsacum* was located in this same region, *nad*5-4 (see below). Nevertheless, no sequence similarities were observed between this array and the conifer STRs (Table S2), thus suggesting independent origins for these motifs.

Evolution of the *nad*5-4 STR-Rich Region

Tracing the variable STR characters onto the Pinaceae phylogeny (Fig. 2) revealed that the insertion and proliferation of the first, third, and fourth-repeated motifs preceded the divergence of the subfamilies Abietoideae (i.e. *Abies*, *Cedrus*, *Keteleeria*, and *Tsuga*) and Pinoideae (i.e. *Larix*, *Picea*, *Pinus*, and *Pseudotsuga*) of the Pinaceae. Interestingly, most of the substitutions interrupting these arrays were observed in the branches leading to the Pinoideae subfamily and the genus *Cedrus*. The insertion of the second repetitive motif apparently occurred after the differentiation of *Cedrus* from the other Abietoideae, while its proliferation took place exclusively within the genus *Abies*. This genus seemingly had two ancestral haplotypes (12 and 18), which are currently distributed in species from Eurasia and North America, and which, respectively, showed the

presence (haplotype 12 and derived forms) and absence (haplotype 18 and derived forms) of the second STR (Figs. 1, 2, 3). The different polymorphisms that characterized the remaining haplotypes in *Abies* were restricted to a few species or populations within species, and generated a significant phylogeographic structure (Fig. 3; see Liepelt et al. 2002, 2010 and Jaramillo-Correa et al. 2008 for more details), suggesting a more recent origin.

Sequence and Population Diversity Across *Abies* Taxa

As a whole, the *nad*5-4 intron had higher diversity and mutation rates in the genus *Abies* than other mitochondrial genes (Table 1). In particular, both the indel and substitution rates were an order of magnitude higher than those of other introns, while all the coding regions surveyed (*atp*8, *ccm*C, *cox*1, *cox*3, *mat*R), excepting SSU *rRNA* V1, were monomorphic across species. However, when analysing each *nad*5-4 region independently, it was observed that this increase of the mutation rates (for both indels and substitutions) actually occurred around the first two STRs, while the remaining of the intron, including the regions with the last two-repeated stretches, were far less variable across *Abies* taxa. Nevertheless, it is noteworthy that these last two regions were highly divergent (D_{xy} values in Table 1) when compared to the outgroups (i.e., the taxa in the Pinoideae subfamily), while the zones of the intron devoid of STRs had similar D_{xy} estimates than the observed for other mtDNA genes (Table 1). Pairwise divergence could not be estimated for the first two stretches given their short size (STR-1) and complete absence (STR-2) in the outgroups, respectively.

On the other hand, and after collecting data from eight population studies implicating 24 *Abies* taxa (i.e., Liepelt et al. 2002, 2010; Ziegenhagen et al. 2005; Jaramillo-Correa et al. 2008, 2011; Jiang et al. 2011; Wang et al. 2011; Aguirre-Planter et al. 2012; Peng et al. 2012), no significant differences were detected for the number of STRs within *nad*5-4, their length, or number of repeats, across species distributed in wide, restricted, continuous, or fragmented ranges (Fig. 4a). Similar results were obtained for the number of mitotypes (*nh*) and average mtDNA diversity (H_E) (Fig. 4b), which suggests that, contrary to nuclear STRs, species effective sizes appear to be independent of the mtSTRs length and diversity for this particular locus.

STR Mining in Other Plant Mitochondrial Regions

The initial literature review resulted in 45 plant mtSTRs, most of which (i.e., 21) were reported for wheat (*Triticum aestivum*) and its wild relatives (Ishii et al. 2006). Six minisatellites have also been reported in *Brassica napus*, five in *Oryza sativa* and its relatives (Honma et al. 2011), four in *Beta vulgaris* (Nishizawa et al. 2000), and nine
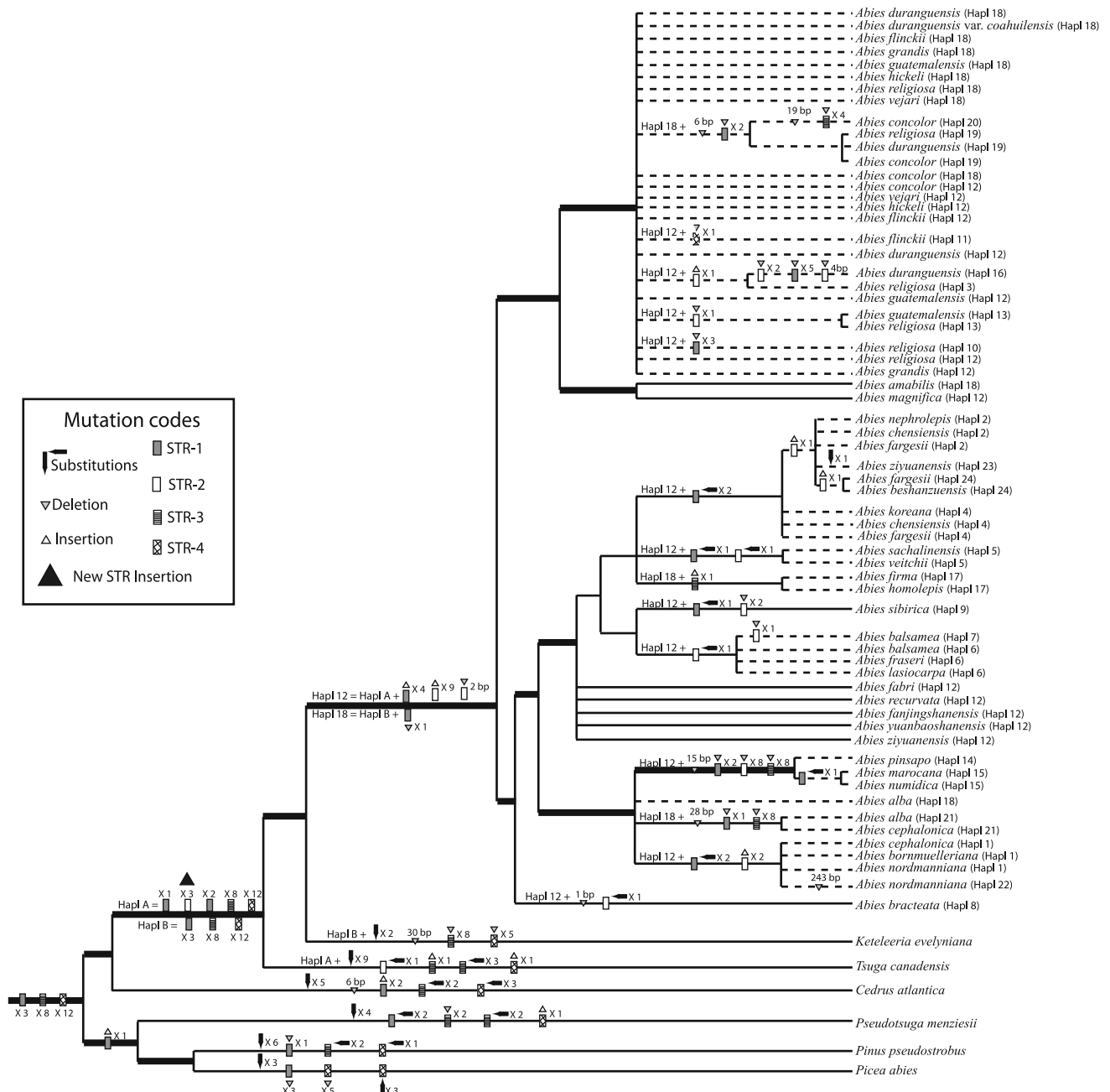
Fig. 2 Tandem-repeats found in the fourth intron of the mitochondrial gene *nad*5 mapped onto the phylogeny of the Pinaceae, with emphasis on the genus *Abies* (adapted from Gernandt et al. (2008), Lin et al. (2010), and Aguirre-Planter et al. 2012). STRs are represented by squares with different fillings. Only the mutations leading to new motifs on each STR are mapped on the phylogenetic tree. Ancestral states were determined by parsimony reconstruction using Mesquite (Maddison and Maddison 2011). *Branches* exhibiting more than one haplotype indicate putative ancestral polymorphisms. *Thick horizontal lines* represent branches supported by bootstrap values over 95, *thinner lines* have values between 50 and 95, and *dashed lines* have no statistical support (i.e., bootstrap below 50) and are shown only to ease the visualisation of particular mutations

tandemly repeated stretches in various conifers (Tables 2, S2). However, it must be noted that half of the STRs reported for *Beta*, and most of those described for *Brassica* and *Oryza* did not meet our selection criteria (80 % homology and at least five-repeated stretches) and were excluded, thus reducing the final set to 33 STRs (Table 2).

The *in silico* search yielded 309 additional STRs for taxa as varied as liverworts, mosses, cycads or grasses, whose complete mtDNA genomes ranged between ~100 and ~1,000 kb long (Table 2, S3). The method used was able to recover all repeated stretches previously reported in the literature for those plants with a complete mtDNA

Fig. 3 Minimum spanning network of the haplotypes found in the fourth intron of the mitochondrial gene *nad*5 of the genus *Abies*. Each mtDNA type is represented by a circle whose filling indicates its geographical location. *Lines* connecting mitotypes denote mutations that are represented by *three different symbols* (see "mutation codes"). The mtDNA types observed in the six outgroups are shown as dashed circles, and the missing mitotypes are represented by *small white dots*
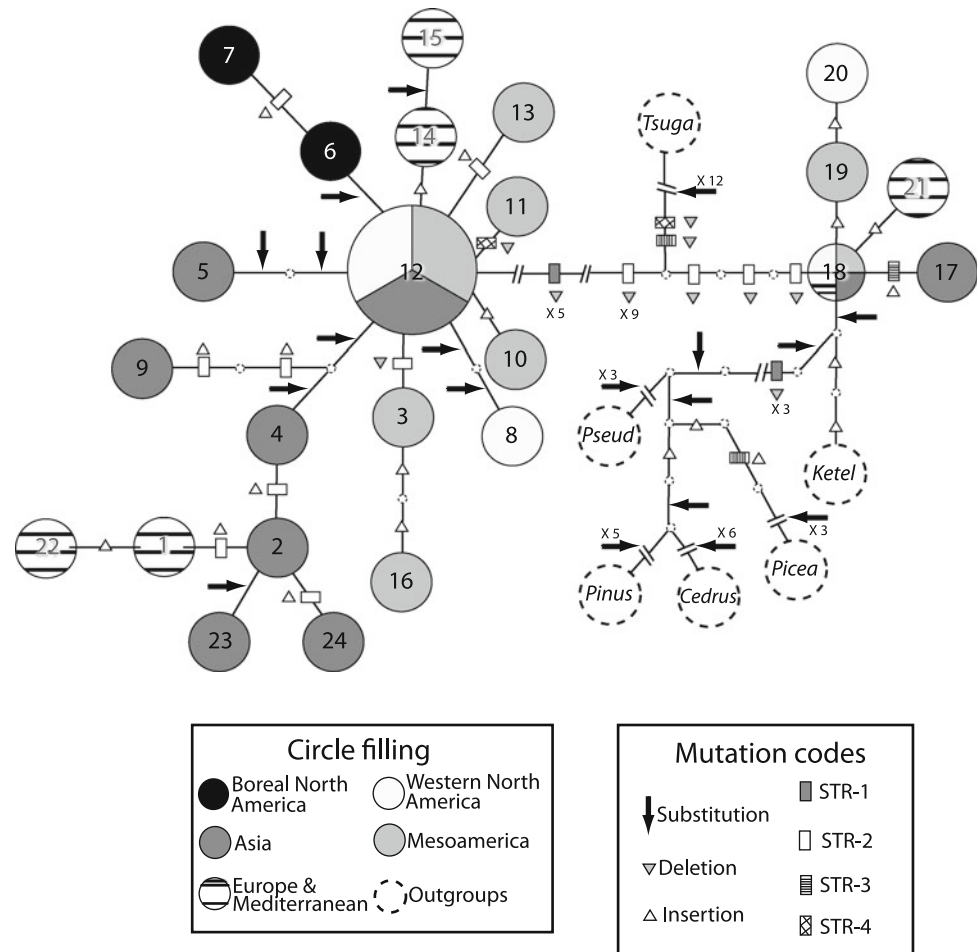
genome (e.g., *Beta vulgaris*, *Brassica napus*, or *Oryza*). However, only those tandem repeats meeting the pre-established selection criteria were retained. The taxa exhibiting the highest number of STRs were *Selaginella moellendorfii* (Selaginellaceae), *Tripsacum dactyloides*, *Zea mays* (Poaceae), *Cucurbita pepo* (Cucurbitaceae), and *Cycas taitungensis* (Cycadaceae). A positive but non-significant correlation was observed between the genome length and the number of STRs ($r^2 = 0.249$; t test $P = 0.17$), which became significant once the atypical *S. moellendorfii* was removed from the dataset ($r^2 = 0.475$; t test $P < 0.01$; see "Discussion" section).

Sequence annotations, complemented by further blast analyses, revealed that most of these STRs (94 %) were located in mitochondrial intergenic regions, followed by group-II introns (6 %) (Tables 2, S3). Only two STRs were observed in coding regions, one in *Brassica napus* (*orf*101a) and the other in *Raphanus sativus* (*orf*138). The first STR was conserved in other species of the genus *Brassica*, while the second one was exclusive to a male-sterile variety of radish (data not shown). Other than the fourth intron of the gene *nad*5, for which STRs were

observed in *Abies*, *Zea*, and *Tripsacum*, repeated stretches were also retrieved across taxa in the following introns: *nad*1 b/c (six Eurasian *Picea*, *Pinus ponderosae* complex, *Triticum–Aegilops* complex, *Selaginella*), *nad*4-1 (most taxa in the Poaceae), and *nad*4-2 (*Cycas*, *Phoenix* and *Selaginella*). However, compound STRs were found only in the first case, in specific taxa of the genera *Picea* and *Pinus* (Mitton et al. 2000; Sperisen et al. 2001). Within species, 17 intergenic regions bore more than one set of repeated motifs (Table 2); some examples include: *rps*2B2–*rrn*26-2 in *Z. mays* with six STRs, *nad*1–*ycf*2 in *C. pepo* with five stretches, and *cob*–*atp*1 in *S. moellendorfii* with four different motifs. Nevertheless, these STRs were dispersed in regions spanning more than 3 kb, except for the last case where the repeated stretches were restricted to a segment of only 330 bp (see Hecht et al. 2011).

Discussion

In this study, we showed that STRs can accumulate and be maintained over long periods of time (i.e., between 180 and
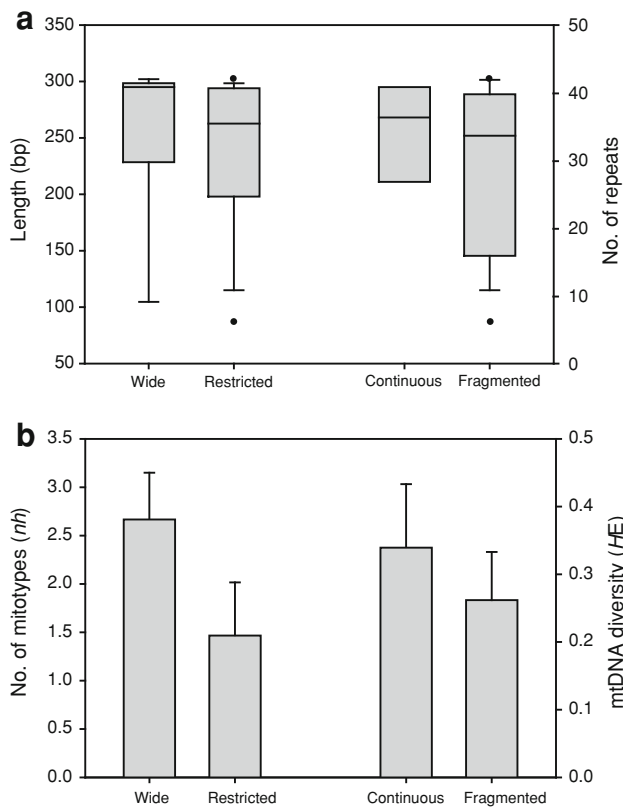
Fig. 4 a Length (*left y* axis) and number of repeats (*right y* axis), and **b** number of mitotypes (*left y* axis) and mtDNA diversity (*right y* axis) observed for the fourth intron of the mitochondrial gene *nad*5 in population studies of *Abies* species distributed in wide or restricted (*left*), and continuous or fragmented (*right*) natural ranges. *Boxes*, *lines within boxes*, and *black dots* in **a** represent 25–75th percentiles, and mean and outlier values, respectively. *Error bars* in **a** and **b** denote 95 % confidence intervals

130 Ma; Aguirre-Planter et al. 2012) in short regions of the plant mitochondrial genome, and that this accumulation can translate in increased mutation rates (both indels and substitutions). The population data for the case study region (intron *nad*5-4) further suggested that mtSTR variability in *Abies* is quiet independent of the species' effective sizes, thus indicating that its variability is the result of the differential fixation of ancestral polymorphisms, instead of the accumulation of more recent mutations following speciation. Also, searches in complete plant mtDNA genomes revealed that although STRs can accumulate in all sorts of plants, proliferation of STR hotspots appears to be taxa specific.

Recurrent STR Insertions in the Plant Mitochondrial Intron *nad*5-4

The co-location of various STRs in short-genomic regions (i.e., less than 1 kb) has been interpreted as previously interrupted motifs that resurge, by recurrent mutations, as linked sets of tandem repeats and/or as a the continual

Table 2 Summary of the number and location of plant mitochondrial sequence tandem repeats (mtDNA STRs) found and meeting selection criteria (see below) with two different search methods

	Search method		
	Literature review[a]	Complete mtDNA genome search[b]	Total
No. of taxa/genomes surveyed	52	31	83
Total Nb of STRs	45	309	354
No. of STRs retained[c]	33	309	342
No. of compound STRs	3	0	3
No. of STRs in introns	8	18	26
No. of STRs in intergenic regions	24	290	314
No. of STRs in exons	1	2	3
No. of regions with various STRs	5	17	24

[a] See Table S2 for a complete description of these STRs

[b] As performed with the software Tandem Repeats Finder (Benson 1999). See Table S3 for a complete description of these STRs

[c] Only those stretches with 80 % sequence similarity and repeated at least five times were considered as STRs

insertion of repeated stretches in mutational hotspots (e.g., Mouhamadou et al. 2007; Morse et al. 2009). The basic repeat sequence of the STRs found in the Pinaceae *nad*5-4 was very different in three out of four cases, which suggests at least three-independent insertions during the evolution of this family. A putative fourth insertion could be at the origin of the last-repeated stretch, although the separate evolution of a previously interrupted motif cannot be ruled out, given that the third and fourth STRs share the same basic sequence (AT; see Figs. 1 and S1).

Further evidence for recurrent STR insertion in this intron derives from the analysis of complete plant mtDNA genomes, which revealed another microsatellite in the homologous intron of *Zea* and *Tripsacum* (Table S3). Our blast analyses showed a proto-minisatellite formed by two repeats of the basic motif (CAATCC) in *Sorghum* and other Poaceae (data not shown), which indicates that the proliferation of this STR occurred after the divergence of the lineage leading to *Zea* and *Tripsacum*, some 12 Ma (Gaut et al. 2000; Swigoňová et al. 2004).

Like many other plant mitochondrial introns, *nad*5-4 is a group-II intron that has apparently lost its self-splicing ability (Kelchner 2002). This type of introns share a basic secondary structure, with a relatively well conserved core region and six helical domains with different degrees of divergence, and within which indel variation and insertions of mobile elements are rather common (e.g., Laroche and Bousquet 1999; reviewed by Bonen 2008). Indeed, polymorphic STRs have been observed within plant group-II

introns of mitochondrial (e.g., *nad*1 b/c in *Picea*; Sperisen et al. 2001, *nad*7-1 in *Pinus*; Godbout et al. 2005, 2008), and chloroplast genes (*pet*D, Löhne and Borsch 2005), while 6 % of the STRs recovered herein from the analyses of complete mtDNA genomes were located in this type of introns (see Table S2). Explanations to account for the accumulation of STRs in group-II introns include relaxed evolutionary constrains, as the self-splicing functions were taken over by external elements (Carrillo et al. 2001), and intragenomic recombination among short-direct repeats (e.g. Woloszynska et al. 2001). Whether the STRs detected within the *Abies nad*5-4 are the result of such rearrangements is still a hypothesis to test, and for which it might be necessary to sequence complete mtDNA genomes across different lineages for detecting the recombination hotspots and their flanking sequences (see Allen et al. 2007, for an example in maize). Nevertheless, mtDNA recombination has already been reported in conifer hybrid zones, but without identifying the specific mechanism involved (Jaramillo-Correa and Bousquet 2005).

Increased Mutation Rates in the *nad*5-4 STR Hotspot of *Abies*

In the Pinaceae, most of the mutations observed in the *nad*5-4 STR hotspot occurred at different times (Fig. 2). For instance, the second array was inserted after the separation of *Cedrus* from the other Abietoideae, about 100 Ma (Gernandt et al. 2008; Lin et al. 2010), while most mutations within this STR occurred exclusively in the genus *Abies*, during the last 80–40 Ma (see Aguirre-Planter et al. 2012 for divergence dates of this genus). As many exclusive substitutions can also be observed within STRs 3 and 4 in *Pinus* (data not shown), which probably took place during the last 80–70 Ma, after the divergence of this genus (Gernandt et al. 2008). Such heterogeneous patterns of mutation rate (i.e., indel and substitution rates) are not rare in plant mitochondrial genes, and are often taxon- or region-specific. For example, in the gene *rps*3, decreased mutation rates were observed for the first intron in the Betulaceae (Laroche and Bousquet 1999) when compared to annual plants such as *Petunia*, while in its third exon, augmented mutation rates were noted in the Podocarpaceae and Araucariaceae relative to other gymnosperms (Ran et al. 2010). However, none of these cases involved the insertion and proliferation of multiple STRs, such as for *nad*5-4 in the Pinaceae.

Both the indel and substitution rates estimated for *nad*5-4 were one order of magnitude higher in the STR hotspot than in its flanking regions, or than in other mtDNA genes devoid of STRs (Table 1). Indeed, the rates for this STR hotspot in *Abies* were of the same order as those previously estimated for the homologous intron among dicots (Laroche et al. 1997), which implies that as many mutations

have accumulated in the same mtDNA region during the last 80–40 Ma in a single genus (*Abies*) as in the last 120–100 Ma among dicot families (Laroche et al. 1997; Chaw et al. 2004). This observation further suggests that the evolutionary pace of typically slowly evolving plant mtDNA regions can be triggered by recurrent STR insertions, such as observed for the nuclear genome (e.g., Amos et al. 2008). This trend indicates an expansion of the role of repeated sequences in the evolution of plant mtDNA: from promoters of intragenomic recombination (e.g., Palmer et al. 2000; Barr et al. 2005; Allen et al. 2007) to mutagenic hotspots (present study).

Accumulation of STRs in the Plant mtDNA Genome

If recurrent STR insertions can increase the mutation rate of a specific plant mtDNA region, it then follows to verify how frequent STR hotspots are across whole plant mtDNA genomes, and if they do tend to accumulate in specific genomic zones and/or particular species. This study revealed that repeated stretches can be found in the mtDNA of all sorts of plants (from mosses and liverworts, to monocots and dicots), with no apparent bias toward a particular group (Table S3; see also Sperisen et al. 2001). Analogous examinations of EST databases revealed similar trends for the nuclear genome (e.g., Wang et al. 1994; Victoria et al. 2011). However, in this study we did find that plants with larger mtDNA genomes tend to have more and larger STRs than plants with smaller genomes, while such a relationship was not apparent for the nuclear genome (Morgante et al. 2002; Victoria et al. 2011). Indeed, the differential accumulation of micro- and minisatellites only accounted for a fraction of the nuclear size variation across species (e.g., Morgante et al. 2002; Morgante 2006), while dissimilar storing of STRs have enlarged fivefold the mtDNA genome of species like cucumber, with respect to their related taxa (e.g., Lilly and Havey 2001; Alverson et al. 2010). However, it must be noted that some species, such as *S. moellendorfii*, have gathered unusually high amounts of STRs (Table S3), while conserving mtDNA genomes of similar size than that of their relatives (Hecht et al. 2011).

Across plant taxa, STRs seemed more prone to accumulate in intergenic mtDNA regions than in introns (see below), while they are very rare in exons (i.e., only 3 among 342 STRs; Table 2). This pattern was expected given the highly variable and dynamic structure of the plant intergenic mtDNA regions. For instance, comparisons of mtDNA genomes across varieties of wheat or maize (Ogihara et al. 2005; Allen et al. 2007) revealed that gene synteny only exists across very short regions, thus implying a rapid evolution through abundant recombination. Indeed, in maize, such rearrangements seem to add and remove

rapidly diverging STRs across the whole mtDNA genome at the short evolutionary time scale (Allen et al. 2007).

Within introns, only a few of the STRs observed were compound or mosaic stretches (i.e., 5 out 309), and all of them were found in conifers (Mitton et al. 2000; Sperisen et al. 2001; Bastien et al. 2003; present study). Outside the Pinaceae, cases of co-located STRs can be found throughout the mtDNA genome of specific taxa, such as *S. moellendorfii* (especially in intron *nad*4-2, Hecht et al. 2011), and *C. taitungensis* (Chaw et al. 2008). In this last species, the recurrent association of the same two elements (Tai and Bpu) across its whole mtDNA genome suggests that STR hotspots could be very common in specific lineages under certain evolutionary circumstances, such as ancestral transposon invasions (Chaw et al. 2008). Given the amount of introns with compound STRs they bear, it could be hypothesized that this phenomenon may also apply to conifers, but this cannot be verified until the first complete mtDNA genomes are sequenced for these taxa.

On the other hand, it has been proposed that STRs and other major mutations that alter genome size can accumulate more readily in species with small mitochondrial effective populations sizes (N_e) (e.g., Lynch and Conery 2003; Boussau et al. 2011). Nevertheless, our analyses in *Abies* did not support such a hypothesis. Indeed, species with restricted or fragmented distributions (and thus small N_e) had a similar number of repeats, total STR length, *nh* and H_E, than their counterparts with wider and continuous ranges (and thus larger N_e; Fig. 4). These results point to a differential fixation of ancestral polymorphisms in isolated taxa, rather than to the accumulation of new mutations following speciation (Boussau et al. 2011), which should be less frequent in species with long-generation times and large ancestral N_e, such as forest trees (Bouillé and Bousquet 2005; see Petit and Hampe 2006 for a review).

Perspectives

In this study, we highlighted the role of STRs as drivers of evolution in the plant mtDNA genome. For instance, we showed that STRs can accumulate and be retained in short regions of the plant mtDNA genome over long periods of time, which differs from the traditional view that small repeated sequences are inserted and deleted on the short-term evolutionary scale, and are seldom retained among distantly related taxa (e.g., Laroche et al. 1997; Allen et al. 2007). Thus, given that STRs are ideal landmarks for surveying structural mtDNA changes (e.g., Lilly et al. 2001; Ogihara et al. 2005; Allen et al. 2007), their study through phylogenetically controlled analyses could expand such comparisons to higher phylogenetic levels, such as attempted herein for a single intron (see also Gugerli et al. 2001; Ran et al. 2010).

Furthermore, we revealed that recurrent STR accumulations can increase the mutation rate of particular mtDNA regions, which is worth investigating at the genome-wide scale. So far, there are no complete mtDNA genomes for conifers, which hamper such a study in these taxa. But this could be easily examined in the Poaceae, for which more than 12 complete genomes are already available (see Table S3, and references therein). However, mtDNA STR variation is still to be surveyed through phylogeographic studies in these species, such as done for some conifers (e.g., Sperisen et al. 2001; Godbout et al. 2005, 2008; Jaramillo-Correa et al. 2008). Primers for specific STRs are already available for *Oryza* (Honma et al. 2011) and *Triticum* (Ishii et al. 2006), but some more could be easily designed for other species (e.g., *Zea*) and transferred to related taxa.

Finally, the differential accumulation of STR hotspots across plants, and its putative correlation with plant life-history traits and ancestral N_e remain open for further investigation. Additional completely sequenced mtDNA genomes should also be necessary to formally address these questions.

References

Aguirre-Planter E, Jaramillo-Correa JP, Gómez-Acevedo S et al (2012) Phylogeny, diversification rates and species boundaries of Mesoamerican firs (*Abies*, Pinaceae) in a genus-wide context. Mol Phyl Evol 62:263–274

Allen JO, Fauron CM, Minx P et al (2007) Comparison among two fertile and three male-sterile mitochondrial genomes in maize. Genetics 177:1173–1192

Alverson AJ, Wei X-X, Rice DW et al (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27:1436–1448

Amos W, Flint J, Xu X (2008) Heterozygosity increases microsatellite mutation rate, linking it to demographic history. BMC Genet 9:72. doi:10.1186/1471-2156-9-72

Barr CM, Neiman M, Taylor DR (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol 168:39–50

Barros P, Blanco MG, Boán F, Gómez-Márquez J (2008) Evolution of a complex minisatellite sequence. Mol Phyl Evol 49:488–494

Bastien D, Favre JM, Collington AM et al (2003) Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce [*Picea abies* (L.) Karst.]. Theor Appl Genet 107:574–580

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucl Acids Res 27:573–580

Bonen L (2008) *Cis-* and *trans-*splicing of group II introns in plant mitochondria. Mitochondrion 8:26–34

Bouillé M, Bousquet J (2005) Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. Am J Bot 92:63–73

Boussau B, Brown JM, Fujita MK (2011) Nonadaptive evolution of mitochondrial genome size. Evolution 65:2706–2711

Broughton RE, Dowling TE (1997) Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow *Cyprinella Spiloptera*. Mol Biol Evol 14:1187–1196

Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryote genomes. BioEssays 28:1040–1050

Carrillo C, Chapdelaine Y, Bonen L (2001) Variation in sequence and RNA editing within core domains of mitochondrial group II introns among plants. Mol Gen Genet 264:595–603

Chaw SM, Shih AC-C, Wang D et al (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol 25:603–615

Chen X, Cho YG, McCouch SR (2002) Sequence divergence of rice microsatellites in *Oryza* and other plant species. Mol Genet Genomics 268:331–343

Cho Y, Mower JP, Qiu Y-L, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci USA 101:17741–17746

Dumolin-Lapègue S, Pemonge M-H, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. Mol Ecol 6:393–397

Farjon A, Rushforth KD (1989) A classification of *Abies* Miller (Pinaceae). Notes of the Royal Botanical Gardens Edinburgh 46:59–79

Gaut BS, Le Tierry d'Ennequin M, Peek AS, Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA 97:7008–7015

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu Rev Genet 44:445–477

Gernandt DS, Magallón S, López GG et al (2008) Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. Int J Plant Sci 169:1086–1099

Godbout J, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. Mol Ecol 14:3497–3512

Godbout J, Fazekas A, Newton CH, Yeh FC, Bousquet J (2008) Glacial vicariance in the Pacific Northwest: evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distinct and widely separated refugia. Mol Ecol 17:2463–2475

Gugerli F, Sperisen C, Büchler U et al (2001) The evolutionary split of the Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. Mol Phyl Evol 21:167–175

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows95/98/NT. Nucl Acid Symp Ser 41:95–98

Hecht J, Grewe F, Knoop V (2011) Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant DNA recombination in early Tracheophytes. Genome Biol Evol 3:344–358

Honma Y, Yoshida Y, Terachi T et al (2011) Polymorphic minisatellites in the mitochondrial DNAs of *Oryza* and *Brassica*. Curr Genet 4:261–270

Ishii T, Takahashi C, Ikeda N et al (2006) Mitochondrial microsatellite variability in common wheat and its ancestral species. Genes Genet Syst 81:211–214

Jackobson M, Sáll T, Lind-Halldén C, Halldén C (2007) Evolution of chloroplast mononucleotide microsatellites in *Arabidopsis thaliana*. Theor Appl Genet 114:223–235

Jaramillo-Correa JP, Bousquet J (2005) Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. Genetics 171:1951–1962

Jaramillo-Correa JP, Bousquet J, Beaulieu J et al (2003) Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. Theor Appl Genet 106:1353–1367

Jaramillo-Correa JP, Aguirre-Planter E, Khasa DP et al (2008) Ancestry and divergence of subtropical montane forest isolates: molecular biogeography of the genus *Abies* (Pinaceae) in southern México and Guatemala. Mol Ecol 17:2476–2490

Jaramillo-Correa JP, Grivet D, Terrab A et al (2011) The Strait of Gibraltar as a major phylogeographic barrier in Mediterranean conifers: a comparative phylogeographic survey. Mol Ecol 19:5452–5468

Jiang Z-Y, Peng Y-L, Hu X-X et al (2011) Cytoplasmic DNA variation and genetic delimitation of *Abies nephrolepis* and *Abies holophylla* in northeastern China. Can J For Res 41:1555–1561

Karhu A, Dieterich J-H, Savolainen O (2000) Rapid expansion of microsatellite sequences in pines. Mol Biol Evol 17:259–265

Kelchner SA (2002) Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. Am J Bot 89:1651–1669

Laroche J, Bousquet (1999) Evolution of the mitochondrial *rps*3 intron in perennial and annual angiosperms and homology with *nad*5 intron I. Mol Biol Evol 16:441–452

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

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial gene flow among refugia. Proc Natl Acad Sci USA 99:14590–14594

Liepelt S, Mayland-Quellhorst E, Lehme M, Ziegenhangen B (2010) Contrasting geographical patterns of ancient and modern genetic lineages in Mediterranean *Abies* species. Plant Syst Evol 284:141–151

Lilly JW, Harvey MJ (2001) Small, repetitive DNAs contribute significantly to the expanded mitochondrial genome of cucumber. Genetics 159:317–328

Lin C-P, Huang J-P, Wu C-S et al (2010) Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biol Evol 2:504–517

Liu T-S (1971) A monograph of the genus *Abies*. Dissertation, National Taiwan University, Taipei

Löhne C, Borsch T (2005) Molecular evolution and phylogenetic utility of the *pet*D group II intron: a case study in basal angiosperms. Mol Biol Evol 22:317–332

Lynch M, Conery JS (2003) The origins of genome complexity. Science 302:1401–1404

Maddison WP, Maddison DR (2011) Mesquite: a modular system for evolutionary analysis, version 2.75. http://mesquiteproject.org

McDonald MJ, Wand W-C, Huang H-D, Leu J-Y (2011) Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. PLoS Biol 9:e1000622. doi:10.1371/journal.pbio.1000622

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 Scotts pine, *P. sylvestris*. Theor Appl Genet 101:1269–1272

Morgante M (2006) Plant genome organisation and diversity: the year of the junk! Curr Opin Biotech 17:168–173

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

Morse AM, Peterson DG, Islam-Faridi MN et al (2009) Evolution of genome size and complexity in *Pinus*. PLoS ONE 4:e4332

Mouhamadou B, Férandon C, Chazoule S, Barroso G (2007) Unusual accumulation of polymorphic microsatellite loci in a specific region of the mitochondrial genome of two mushroom-forming *Agrocybe* species. FEMS Microbiol Lett 272:276–281

Mower JP, Touzet P, Gummow JS et al (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol 7:135

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

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155

Ogihara Y, Yamazaki Y, Murai K et al (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Genetics 33:6235–6250

Palmer JD, Adams KL, Cho Y et al (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci USA 97:6960–6966

Peng Y, Yi S, Wang J et al (2012) Phylogeographic analysis of the fir species in southern China suggests complex origin and genetic admixture. Ann For Sci 69:409–416

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol Syst 37:187–214

Polezhaeva MA, Lascoux M, Semerikov VL (2010) Cytoplasmic DNA variation and biogeography of *Larix* Mill. in Northeast Asia. Mol Ecol 19:1239–1252

Primer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. Mol Biol Evol 15:997–1008

Ran J-H, Gao H, Wang X-Q (2010) Fast evolution of the retroprocessed mitochondrial *rps*3 gene in conifer II and further evidence for the phylogeny of gymnosperms. Mol Phyl Evol 54:136–149

Rushforth KD (1989) Two new species of *Abies* (Pinaceae) from western México. Notes Roy Bot Gard Edinburgh 46:101–109

Semerikova SA, Semerikov VL, Lascoux M (2011) Post-glacial history and introgression in *Abies* (Pinaceae) species of the Russian Far East inferred from both nuclear and cytoplasmic markers. J Biogeo 38:326–340

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

Swigoňová Z, Lai J, Ma J et al (2004) Close split of Sorghum and maize genome progenitors. Genome Res 14:1916–1923

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460

Thuiller A-C, Bru D, David J et al (2002) Direct estimation of mutation rate for 10 microsatellite loci in Durum wheat, *Triticum turgidum* (L.) Thell. ssp. *durum* desf. Mol Biol Evol 19:122–125

Vázquez-Lobo A (1996) Evolución de hongos endófitos del género Pinus. Tesis de licenciatura, Facultad de Ciencias, Universidad Nacional Autónoma de México, México, DF

Vergnaud G, Denoeud F (2000) Minisatellites: mutability and genome architecture. Genome Res 10:899–907

Victoria FC, da Maia LC, Costa de Oliveira A (2011) *In silico* comparative analysis of SSR markers in plants. BMC Plant Biol 11:15

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

Wang J, Abbot RJ, Peng YL et al (2011) Species delimitation and biogeography of two fir species (*Abies*) in central China: cytoplasmic DNA variation. Heredity 107:362–370

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Pop Biol 7:256–276

Wilkinson GS, Mayer F, Kerth G, Petri B (1997) Evolution of repeated sequence arrays in the *D-loop* region of bat mitochondrial DNA. Genetics 146:1035–1048

Woloszynska M, Kieleczawa J, Ornatowska M et al (2001) The origin and maintenance of small repeats in the bean mitochondrial genome. Mol Genet Genomics 265:865–872

Xiang Q-P, Xiang Q-Y, Guo Y-Y, Zhnag X-C (2009) Phylogeny of Abies (Pinaceae) inferred from nrITS sequence data. Taxon 58:141–152

Zar JH (2010) Biostatistical analysis, 5th edn. Prentice Hall, Upper Saddle River

Zhu Y, Queller DC, Strassmann JE (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. J Mol Evol 50:324–338

Ziegenhagen B, Fady B, Kuhlenkamp V, Liepelt S (2005) Differentiating groups of *Abies* species with a simple molecular marker. Silvae Genet 54:123–126