Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 44 (2007) 765–777
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Structural evolution of nrDNA ITS in Pinaceae and its phylogenetic implications

Xian-Zhao Kan [a], Shan-Shan Wang [a], Xin Ding [a], Xiao-Quan Wang [a,b,*]

[a] *State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, The Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, China*
[b] *Graduate School of the Chinese Academy of Sciences, Beijing 100039, China*

Received 11 September 2006; revised 24 April 2007; accepted 7 May 2007
Available online 25 May 2007

Abstract

Nuclear ribosomal DNA (nrDNA) has been considered as an important tool for inferring phylogenetic relationships at many taxonomic levels. In comparison with its fast concerted evolution in angiosperms, nrDNA is symbolized by slow concerted evolution and substantial ITS region length variation in gymnosperms, particularly in Pinaceae. Here we studied structure characteristics, including subrepeat composition, size, GC content and secondary structure, of nrDNA ITS regions of all Pinaceae genera. The results showed that the ITS regions of all taxa studied contained subrepeat units, ranging from 2 to 9 in number, and these units could be divided into two types, longer subrepeat (LSR) without the motif (5′-GGCCACCCTAGTC) and shorter subrepeat (SSR) with the motif. Phylogenetic analyses indicate that the homology of some SSRs still can be recognized, providing important informations for the evolutionary history of nrDNA ITS and phylogeny of Pinaceae. In particular, the adjacent tandem SSRs are not more closely related to one another than they are to remote SSRs in some genera, which may imply that multiple structure variations such as recombination have occurred in the ITS1 region of these groups. This study also found that GC content in the ITS1 region is relevant to its sequence length and subrepeat number, and could provide some phylogenetic information, especially supporting the close relationships among *Picea*, *Pinus*, and *Cathaya*. Moreover, several characteristics of the secondary structure of Pinaceae ITS1 were found as follows: (1) the structure is dominated by several extended hairpins; (2) the configuration complexity is positively correlated with subrepeat number; (3) paired subrepeats often partially overlap at the conserved motif (5′-GGCCACCCTAGTC), and form a long stem, while other subrepeats fold onto itself, leaving part of the conserved motif exposed in hairpin loops.
© 2007 Elsevier Inc. All rights reserved.

Keywords: nrDNA ITS; Pinaceae; Length variation; Subrepeat; Secondary structure; Concerted evolution; Molecular phylogeny

1. Introduction

Eukaryotic ribosomal RNA multigene family are present in hundreds or thousands of repeats tandemly arranged at one or several locations (Nei and Rooney, 2005; Eickbush and Eickbush, 2007). Each rDNA repeat unit consists of three rRNA gene regions (18S, 5.8S, and 26S), two internal transcribed spacers (ITS1 and ITS2), and an intergenic spacer (IGS) in plants. The spacer regions act in pre-RNA processing, and may serve to organize specific soluble factors, possibly acting in a manner which is analogous with that of the free small nucleolar ribonucleo protein particles (snoRNPs) (Lalev and Nazar, 1998; Michot et al., 1999; Côté and Peculis, 2001). Sequence of the ITS region, including ITS1, 5.8S, and ITS2, can diagnose organismal origins and phylogenetic relationships at many taxonomic levels and provide a useful paradigm for molecular evolutionary studies (Baldwin et al., 1995; Wendel et al., 1995; Buckler and Holtsford, 1996a,b; Markos and Baldwin, 2002; Razafimandimbison et al., 2004), although there

* Corresponding author. Address: State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, The Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, China. Fax: +86 10 62590843.
E-mail address: xiaoq_wang@ibcas.ac.cn (X.-Q. Wang).

1055-7903/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2007.05.004

are some controversies (Alvarez and Wendel, 2003; Bailey et al., 2003). In comparison with its high concerted evolution in angiosperms through unequal crossing-over (Smith, 1976) and gene conversion (Nagylaki and Petes, 1982; Nagylaki, 1984; Hillis et al., 1991), nrDNA of gymnosperms, particularly in Pinaceae, is symbolized by slow concerted evolution and exhibits marked length variation of ITS1 (944–3271 bp) (Karvonen and Savolainen, 1993; Karvonen et al., 1994; Gernandt and Liston, 1999; Maggini et al., 2000; Wei et al., 2003; Campbell et al., 2005). Liston et al. (1996) also observed that Pinaceae ITS1 is much longer than that of other seed plants and, recognizing the difficulty of assessing the homology of the character "ITS1 length", presented that the patterns of length variation might have some systematic utility.

Marrocco et al. (1996) reported the first complete ITS region nucleotide sequence of a non-flowering seed plant (*Pinus pinea*) and the occurrence of ITS1 subrepeats in the species. Gernandt and Liston (1999) found that subrepeats also occurred in ITS1 of *Larix* and*Pseudotsuga*, and all subrepeats had a highly conserved central core of GGCCACCCTAGTC. Campbell et al. (2005) found that large variation in subrepeat number existed among pinaceous genera, and the conserved motif (GGCCACCCT AGTC) occurred in all Pinaceae ITS1, synapomorphic for the family. Interestingly, Maggini et al. (2000) revealed that ITS1 subrepeats of the pine family could be divided, according to occurrence of the motif, into two types, namely longer subrepeat (LSR) without the motif and shorter subrepeat (SSR) with the motif. In particular, Campbell et al. (2005) noted that the correlation between similarity and proximity of *Picea* ITS1 subrepeats may be the result of subrepeat duplication or concerted evolution within rDNA repeats. So far LSR has only been reported from *Picea* (Maggini et al., 2000). Does it also exist in other genera of Pinaceae? More sampling is obviously needed to investigate the origin and evolution of LSR, which may further shed light on the evolutionary dynamics of subrepeats and structure evolution of the ITS region. On the other hand, Campbell et al. (2005) found that number of subrepeats did not appear to be phylogenetically informative in Pinaceae, based on a sampling of seven genera. How many subrepeats are there in the other genera? What is their distribution? In order to reveal the mechanisms underlying length variation of the ITS region and genealogical relationships among ITS1 subrepeats in Pinaceae, it is necessary to sample all genera of the family and sequence the ITS region completely.

The family Pinaceae, including 11 widely accepted genera, could be divided into two major groups, the clade comprising *Cathaya, Larix, Picea, Pinus*, and *Pseudotsuga* and the clade consisting of *Abies*, *Keteleeria*, *Nothotsuga*, *Pseudolarix*, and *Tsuga*, besides *Cedrus* with an unique position (Wang et al., 2000). To date, the entire ITS1 region sequence can be obtained from *Pinus* (Marrocco et al., 1996; Gernandt et al., 2001), *Picea* (Maggini et al., 2000; Campbell et al., 2005), *Larix, Pseudotsuga* (Gernandt and Liston, 1999; Wei et al., 2003; Wei and Wang, 2004;), *Tsuga*, *Pseudolarix*, and *Abies* (Vining and Campbell, 1997; Campbell et al., 2005), but has not been determined in other four genera of Pinaceae (*Cathaya*, *Cedrus*, *Keteleeria*, and *Nothotsuga*). Moreover, sequences of 5.8S rDNA and the ITS2 region have not been reported from *Abies*, *Pseudolarix*, and *Tsuga*. In the present study, we determined the entire nucleotide sequences of the ITS region from *Abies, Cathaya*, *Cedrus*, *Keteleeria*, *Nothotsuga*, *Pseudolarix*, and *Tsuga*. Based on the new data combined with previous reports from other genera, we tried to address (1) composition, distribution and evolution of nrDNA ITS1 subrepeats in Pinaceae, (2) the relationship between length variation and subrepeat number in ITS1 and its phylogenetic implications, and (3) characteristics of ITS1 secondary structure of the pine family.

2. Materials and methods

2.1. Plant materials

Seven Pinaceae genera without complete ITS sequence data up to now were sampled, including *Abies*, *Cathaya*, *Cedrus*, *Keteleeria*, *Nothotsuga*, *Pseudolarix*, and *Tsuga*. Sources of materials used in the present study are shown in Table 1. Voucher specimens have been deposited in the herbarium of Institute of Botany, the Chinese Academy of Sciences (PE). The other ITS sequences were retrieved from GenBank (Table 1).

2.2. DNA extraction and PCR amplification, cloning and sequencing

Total DNA was extracted from fresh or silica gel dried needles using the CTAB method following the protocol of Rogers and Bendich (1988) and used as template in polymerase chain reaction. The ITS region was amplified with primers ITS1N of Wei et al. (2003) or ITS5* of Liston et al. (1996) located on the 18S rDNA and ITS4 of White et al. (1990) or 26S25R of Nickrent et al. (1994) (Fig. 1). The PCR was carried out in a volume of 25 μl containing 5–50 ng of DNA template, 6.25 pmol of each primer, 0.2 mM of each dNTP, 2 mM $MgCl_2$, and 0.75 U of Ex *Taq* DNA polymerase (TaKaRa, Dalian, China). Amplification was conducted in a Peltier Thermal Cycler (PTC-100, MJ Research) or a Tgradient 96 U thermocycler (Biometre, Göttingen, Germany). PCR cycles were as follows: one cycle of 4 min at 70 °C, 4 cycles of 40 s at 94 °C, 20 s at 52 °C, and 2 min 30 s at 72 °C, followed by 36 cycles of 20 s at 94 °C, 20 s at 48–55 °C, and 2 min 30 s (3 min 30 s for *Cathaya*) at 72 °C, with a final extension step of 10 min at 72 °C. PCR products were separated by 1.5% agarose gel electrophoresis. The band with the right size was cut out and purified using GFX™ PCR DNA and Gel Band Purification Kit (Pharmacia), and then cloned with pGEM®-T Easy Vector System II (Promega).

Table 1
Sources of plant materials and nrDNA ITS sequences analyzed in the present study

Taxon (abbreviation)	Source/voucher	Clone Nos.	GenBank Accession Nos.
Abies concolor (Gord.) Lindl. ex Hildebr. (Aco)	Campus of Michigan State University, USA/9902M	6, 7, 8, 15, 45, 21, 29	DQ975359–DQ975365
Abies fabri (Mast.) Craib (Afa)	Gongga Mts., Sichuan, China/Wang2003-1	1, 2	DQ975352, DQ975353
Cathaya argyrophylla Chun et Kuang (Car)	Dayao Mountain, Guangxi, China/DY12	445, 675	DQ975347, DQ975348
Cedrus deodara (Roxb.) G. Don (Cde)	Botanic Garden, Institute of Botany, Beijing, China/ Wang2006-1	14	DQ975357
Keteleeria evelyniana Mast. (Kev)	Botanic Garden, Kunming Institute of Botany, China/9631	2	DQ975354
Larix potaninni Batalin (Lpo)	Wei et al. (2003)		AY188552
Nothotsuga longibracteata Hu ex Page (Nlo)	Xinning, Hunan, China/LuoZC-002	9, 10, 16	DQ975349–DQ975351
Picea abies (L.) Karst. (Pab)	Maggini et al. (2000)		AJ243165–AJ243167
Pinus pinea L. (Ppi)	Marrocco et al. (1996)		X87936
Pinus quadrifolia Parl. ex Sudw. (Pqu)	Gernandt et al. (2001)		AF343990
Pseudolarix amabilis (Nelson) Rehd. (Pam)	Lushan Botanical Garden, China/04092	1, 10	DQ975355, DQ975356
Pseudotsuga menziesii (Mirbel) Franco (Pme)	Gernandt and Liston (1999)		AF041353
Tsuga chinensis (Franch.) Pritz (Tch)	Lushan Botanical Garden, China/9919	4	DQ975358
Tsuga mertensiana (Bong.) Carrière (Tme)	Campbell et al. (2005)		AY570231

For each taxon, at least 30 clones with correct insertion (determined by digestion with EcoRI) were screened by comparing restriction fragments of Msp I or both HaeIII and HinfI. All distinct clones were sequenced with the two PCR primers and several internal primers (Fig. 1), i.e., CarF1 (5′-ATGTGCAAGGGGCTTTGCTTC), CarF2 (5′-CTAATGC TTTGTGCGTCAGTC), CarF3 (5′-AGGGAGTGGTAGGCTAATACC), CarR1 (5′-TG GG ATAACAAGGAGAAGCAG), KevR (5′-AGGAA CACGCTGCTGCACAT), CdeR (5′-AG CAGAGCAC AAGATACAAG), NloR (5′-CAACCGCGACACACG TGCAAAAC), TchR (5′-CTCTGTGCCTTTGGTTC TCT), DURA1323R (5′-GCACAACCGACATAGGAGG T) and P2N (5′-GAGAGCCGAGATATCCGTTG), using the DYEnamic ET Terminator Kit (Amersham Pharmacia Biotech). After precipitation in 95% EtOH and 3M NaAc (pH 5.2), the sequencing products were separated on a MegaBACE 1000 automatic DNA sequencer (Amersham Biosciences, Buckinghamshire, UK).

2.3. Data analyses

Boundaries of the ITS1, 5.8S, and ITS2 regions were identified in comparison with other available sequences (Gernandt and Liston, 1999). SSRs were identified by search of the ITS region for the core sequence, 5′-GGCCACCCTAGTC, with a threshold of 20% mismatch (Gernandt et al., 2001), while LSRs were detected by Tandem Repeats Finder (ver. 4.00, http://tandem.bu.edu/trf/trf.html) (Benson, 1999) and manual alignment. Sequence alignments were made with CLUSTAL X v.1.81 (Thompson et al., 1997) and refined manually. MEGA v.3.1 (Kumar et al., 2004) was applied to estimate GC content and nucleotide substitutions (*d*) for ITS1, 5.8S, and ITS2 regions separately according to Kimura's two-parameter model (Kimura, 1980).

To investigate the evolutionary relationships of subrepeats and their phylogenetic implications, two analyses were performed with the maximum parsimony method using PAUP version 4.0b10 (Swofford, 2002): (1) only ITS1 subrepeat sequences of *Cathaya argyrophylla* were used, with *Picea* as the outgroup; (2) all ITS1 subrepeat sequences from the eleven genera of Pinaceae, excluding LSRs, were used, and both *Pinus* subgenera were included since the divergence between them is equal or greater than other sister genera in Pinaceae. All character states were specified as unordered and equally weighted with indels as missing data. Heuristic search was implemented with 100 random addition sequence replicates, tree bisection-reconnection (TBR) branch swapping, the MULTREES option, and a maximum of 1000 trees saved per round. To evaluate relative robustness of the clades found in the most parsimonious trees, the bootstrap analysis (Felsenstein, 1985) employed 500 replicates using the same heuristic search settings except that a maximum of 100 trees were saved per round. We also constructed a phylogeny of SSRs of the Pinaceae ITS1 with the maximum-likelihood (ML) method. Modeltest 3.7 (Posada and Crandall, 1998) was used to determine the best model of sequence evolution for the SSR sequences, and the HKY + I + G model was suggested as the best using Akaike Information Criterion (AIC). Likelihood analysis was performed in PHYML version 2.4.3 (Guindon and Gascuel, 2003) using the HKY model. ML parameters were then optimized, with a BIONJ tree as a starting point (Gascuel, 1997). Support values for nodes on the ML tree were estimated with 500 bootstrap replicates (Felsenstein, 1985).

Secondary structure of the ITS1 region of all 11 Pinaceae genera was predicted with Mfold 3.2 for Linux (Math-

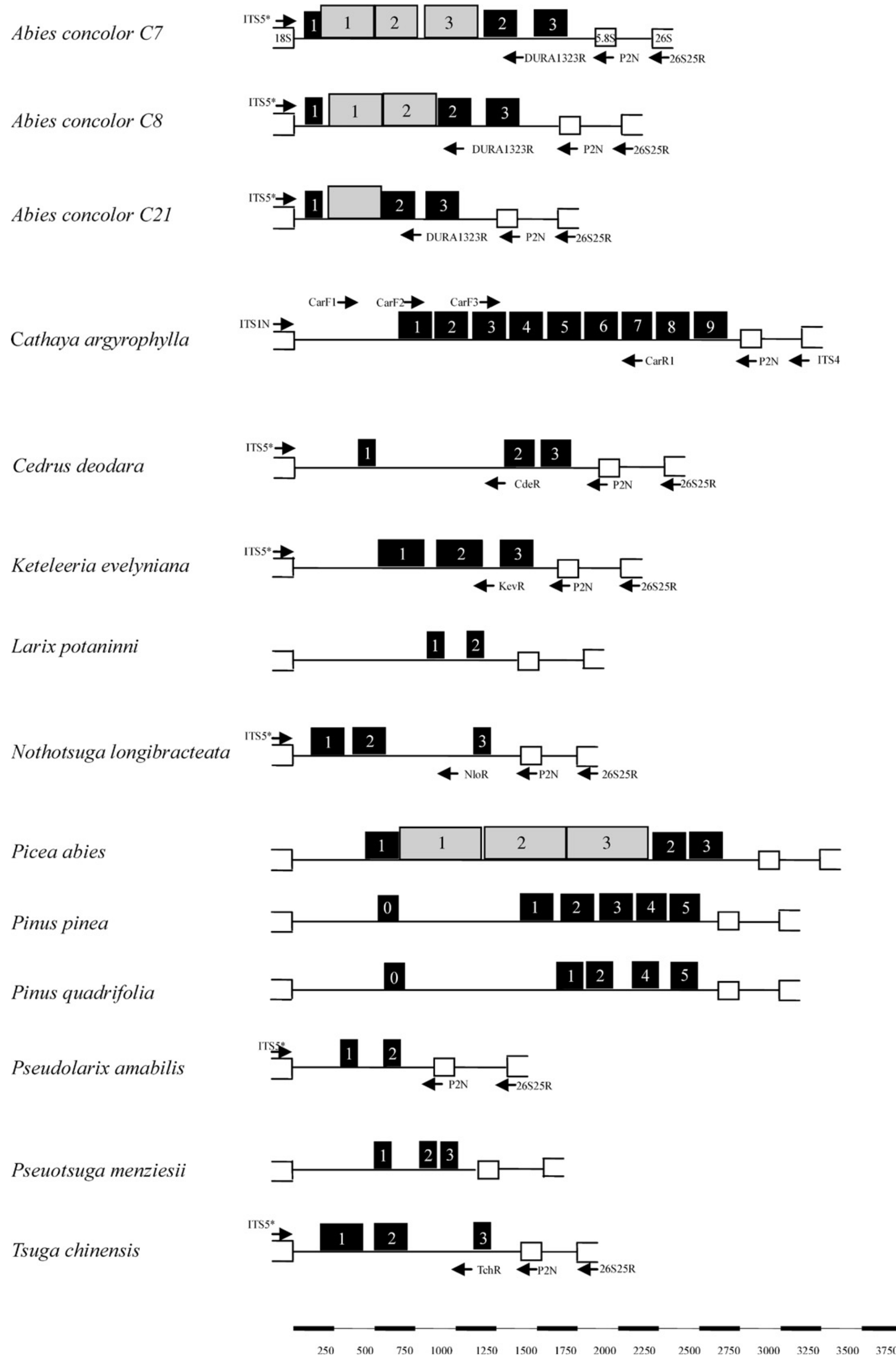

Fig. 1. Schematic representation and primer position for the ITS region in Pinaceae. Shadowed boxes represent the long subrepeats (LSRs), while solid black boxes represent the short subrepeats (SSRs). The SSRs and LSRs are numbered in order from the 5′-end of the sequence, respectively. Horizontal arrows indicate the locations and directions of primers used in the present study. The scale (in base pairs) at the bottom indicates relative lengths. Total length of the ITS region for *Picea* is based on Maggini et al. (2000) (GenBank Accession Nos. AJ243165, AJ243167), and Germano and Klein (1999) (GenBank Accession No. AF136621).

ews et al., 1999; Zuker, 2003). We used default values for RNA sequence (linear), ionic conditions (1 M NaCl, no divalent ions), percent suboptimality (5), upper bound on the number of computed foldings (50), window parameter (15 for *Pseudolarix*, 25 for *Cathaya*, *Pinus*, and *Picea*, 20 for the rest species), maximum size of a bulge/interior loop (30), maximum asymmetry of a bulge/interior loop (30), and maximum distance between paired bases (no limit). The folding temperature was set at 25 °C, in considering of the habitats of the plants we studied. We examined Mfold output for the location of subrepeats and conserved motifs, the patterns of pairing among subrepeats, and overall structure of ITS1. Options of Mfold were selected following Gernandt et al. (2001) and Campbell et al. (2005).

3. Results

3.1. Length variation and GC content of the ITS region

Nineteen distinct clones were recognized by restriction analysis and completely sequenced (Table 1). The ITS region (ITS1 + 5.8S + ITS2) ranges from 1348 bp (*Pseudolarix*) to 3189 bp (*Cathaya*) in size for the seven Pinaceae genera we studied. The 5.8S coding region has a conserved length of 162 bp except a 1-bp insertion found in *Cathaya*. In addition, only slight length variation was found in the ITS2 region (Table 2). For example, *Cedrus* and *Keteleeria* have shorter ITS2, i.e., 233 and 239 bp, respectively, than the others (242–249 bp). In contrast, extensive length variation occurred in the ITS1 region. For instance, *Cathaya* has the longest ITS1 of 2731 bp, which is 1.9 times longer than the shortest of 944 bp in *Pseudolarix*. Moreover, the three clones from the same individual of *Abies concolor*, *Aco-21*, *Aco-8*, and *Aco-7*, differ greatly in the ITS1 length, being 1648, 1933, and 2216 bp, respectively. However, this intragenomic length variation was not detected from the congeneric species *Abies fabri* by restriction analysis of nearly 20 clones (Table 2 and Fig. 1).

Average GC content is slightly higher in ITS2 (58.8-65.3%) than in ITS1 (51.3–62.5%), and is significantly lower in the 5.8S region (50.6–52.5%) than in the two spacers. In addition, a marked low GC content (52.1–52.3%) was found in ITS1 of *Cathaya* (Table 2). It is interesting that the GC content is higher in SSRs than in the other region of ITS1. Especially, GC% of SSRs is 10.8% higher than that of the non-repeat region in *Cathaya* (Table 3).

3.2. Sequence characterization of the ITS-1 region

In the ITS1 sequence, we found two types of subrepeats, SSR and LSR, as designated by Maggini et al. (2000). SSR occurs in each genus, and its number varies among genera, ranging from two in *Pseudolarix* to nine in *Cathaya*. Every SSR has a conserved motif of 5′-GGCCACCCTAGTC3-′ as reported in Gernandt and Liston (1999), except that two T/C transitions occurred in the conserved motif of two SSRs from *Keteleeria*, giving rise to GGCCACCC-CAGTC and GGCCACCCTAGCC, respectively. All SSRs of *Cathaya* and *Keteleeria* are adjacently arrayed, but in the other genera one of the SSRs is remotely located at the 3′- or 5′-end of ITS-1. In contrast, LSR was only found in *Abies* except the previous report in *Picea* by Maggini et al. (2000), and its number has intra- and intergenomic variation. For example, two clones (*Afa-1* and *Afa-2*) from *A. fabri* only have one LSR, while 1–3 LSRs occur in *Aco-21*, *Aco-8*, and *Aco-7* from *A. concolor*, respectively (Fig. 1 and Table 3). Actually, if we had not aligned ITS1 sequences of *A. fabri* and *A. concolor*, it would be difficult to find the only one copy of LSR in *A. fabri*.

It is of great interest that marked length variation of subrepeats exists among genera and within genome, even the same clone. For example, SSR1 is much shorter in

Table 2
Summary of length variation (bp) and G + C content (%) of the ITS region in Pinaceae

Taxa/clones (abbreviation)	ITS1		5.8S		ITS2		Total length
	Length	GC%	Length	GC%	Length	GC%	
Abies concolor C7 (Aco-7)	1811	58.6	162	51.2	243	61.7	2216
Abies concolor C8 (Aco-8)	1528	59.0	162	51.2	243	60.7	1933
Abies concolor C21 (Aco-21)	1243	60.0	162	51.2	243	60.5	1648
Abies fabri	1254–1255	60.3	162	50.6–51.2	243–244	61.9–62.0	1661
Cathaya argyrophylla	2730–2731	52.1–52.3	163	50.9	245	58.8–59.2	3138–3189
Cedrus deodara	1813	59.7	162	52.5	233	61.8	2208
Keteleeria evelyniana	1689	62.5	162	51.9	239	65.3	2090
Larix potaninni	1373	57.0	162	51.2	232	59.5	1676
Nothotsuga longibracteata	1345–1347	59.4–59.9	162	51.2–51.9	249	61.8–62.2	1756–1758
Picea abies	2784–3271	53.1–53.2	162[a]	51.9	238	56.3	3184–3671
Pinus pinea	2631	52.8	161	52.2	245	60.0	3037
Pinus quadrifolia	2601	51.3	162	49.4	242	59.1	3010
Pseudolarix amabilis	944	60.7	162	52.5	242–243	62.4–63.0	1348–1349
Pseudotsuga menziesii	1177	59.1	162	52.5	232	64.9	1571
Tsuga chinensis	1357	59.1	162	51.2	247	61.5	1766

[a] 5.8S length is from Germano and Klein (1999) (GenBank Accession No. AF136621).

Table 3
Subrepeat (SR) number, length, position, and GC content in the ITS1 region of Pinaceae

Taxa/clones	SR No.	Length of SRs (nt positions)	GC% of each SR	GC% of SR regions	GC% of non-SR regions	SR% of ITS1
A. concolor C7	3+3[a]	88 (61–148), 242 (1078–1319), 246 (1371–1616), 283 (229–511), 283 (512–794), 283 (795–1077)	62.5, 64.5, 65.4, 55.5, 55.5, 55.8	59.2	57.0	31.9, 78.7[b]
A. concolor C8	3+2	88 (63–150), 241 (797–1037),247(1089–1335), 283 (231–513), 283 (514–796)	61.4, 65.2, 66.0, 55.1, 56.3	60.4	54.8	37.8, 74.9
A. concolor C21	3+1	88 (61–148), 241 (512–752), 247 (804–1050), 283 (229–511)	64.8, 64.0, 66.1,55.1	61.7	56.7	46.3, 69.1
A. fabri C1	3+1	88 (67–154), 242(517–758), 246(815–1060), 282 (235–516)	63.6, 64.1, 66.7, 57.1	62.5	55.6	45.9, 68.4
C. argyrophylla C445	9	221 (672–892), 216 (897–1112), 217 (1116–1332), 216 (1336–1551), 215 (1555–1769), 216(1771–1986), 210 (1991–2200), 220 (2202–2421), 218 (2423–2640)	54.8, 56.0, 54.8, 58.3, 55.8, 54.6, 54.3, 55.0, 53.2	55.2	44.4	71.4
C. deodara	3	64 (373–436), 114 (1305–1418), 117 (1423–1539)	59.4, 60.5, 62.4	61.0	59.5	16.3
K. evelyniana	3	289 (550–838), 291 (852–1142), 214 (1285–1498)	61.9, 63.6, 70.0	64.7	60.6	47.0
L. potaninni	2	64 (802–865), 78 (1092–1169)	64.1, 59.0	62.3	56.4	10.3
N. longibracteata C9	3	202 (230–431), 203 (475–677), 99 (1087–1185)	59.9, 61.6, 64.6	61.5	58.9, 61.8–62.2	37.5
P. abies	3+3	210 (423–632), 229 (2229–2457), 228 (2458–2685), 480 (633–1112), 480 (1126–1605), 586 (1606–2191)	57.1, 48.9, 55.7, 51.7, 53.0, 52.4	53.0	53.0	24.0, 79.5
P. amabilis C1	2	144 (208–351), 143 (603–745)	63.9, 62.9	63.4	59.5	30.4
P. menziesii	3	68 (524–591), 79 (823–901), 77 (902–978)	64.7, 60.8, 59.7	61.6	58.6	19.0
P. pinea	6	174 (496–669), 231 (1375–1605), 237 (1619–1855), 237 (1856–2092), 225 (2096–2320), 215 (2321–2535)	58.0, 54.1, 51.5, 55.5, 54.5, 55.6	55.0	50.9	51.0
P. quadrifolia	5	140 (570–646), 187 (1534–1720), 212 (1720–1932), 203 (2003–2205), 193 (2278–2470)	56.4, 57.2, 55.2, 46.8, 60.1	55.0	49.2	35.9
T. chinensis	3	263 (196–458), 258 (462–719), 117 (1065–1181)	59.7, 61.2, 59.8	60.3	60.0	47.0

[a] Numbers underlined are values of LSRs.
[b] The first number is for SSRs only, and the second number is for all SRs.

Cedrus (64 bp), *Larix* (64 bp), *Pseudotsuga* (68 bp) and *Abies* (88 bp) than in the other genera (140–289 bp), and in the clone *Tch-4* SSR1 (263 bp) and SSR2 (258 bp) are more than twice the size of SSR3 (117 bp) (Table 3). In addition, SSRs located in the same position of different clones (corresponding SSRs) are much less divergent than SSRs tandemly arranged in the same clone. For an example in *Cathaya*, sequence similarity between corresponding SSRs in the two clones *Car-445* and *Car-675* ranges from 94.7% to 100%, while that between the nine SSRs (SSR1–SSR9) ranges from 69.9% to 92.6% for *Car-675*, and from 74.6% to 92.6% for *Car*-445. Among the three SSRs found in each of the six genera *Abies*, *Cedrus*, *Nothotsuga*, *Picea*, *Pseudotsuga*, and *Tsuga*, two are adjacent, remote from the third, and more similar to one another than either is to the third one (Fig. 1).

3.3. Phylogenetic analyses of ITS1 subrepeats

When only SSRs of *Cathaya* and the outgroup *Picea abies* were included, the sequence alignment was 197 bp long. Parsimony analysis of this data set yielded 10 most parsimonious trees with a tree length of 225 steps, a consistency index (CI) of 0.764 and a retention index (RI) of 0.810. In the tree, all *Cathaya* SSRs formed a strongly supported monophyletic group, which was further divided into two main clades, one including SSR1, SSR3, SSR4, SSR5, SSR7, and SSR9, and the other comprising SSR2, SSR6, and SSR8. Obviously, this clustering order does not correspond to position relationships among the SSRs in the DNA strand (Fig. 1).

The aligned sequences of all 51 SSRs from the 11 genera of Pinaceae, excluding some unalignable regions, had 113 characters, of which 83 were variable and 80 were parsimony informative. Parsimony analysis using heuristic search generated 355 most parsimonious (MP) trees (tree length = 472 steps, CI = 0.4089, RI = 0.6872), one of which is shown in Fig. 2. The tree is topologically similar to the maximum-likelihood (ML) tree (Fig. 3). Five common clades were found in the ML and MP trees, including the strict-consensus tree of the latter, although most clades had low bootstrap support and some within-clade topological differences existed between Figs. 2 and 3. Clade I consisted of SSR0 of *P. pinea* and *Pinus quadrifolia*, SSR1 of *Picea abies*, and all 9 SSRs of *Cathaya*; Clade II included SSR2 and SSR3 of *Picea*, and all SSRs, except SSR0, of the two pine species analyzed; Clade III was composed of SSR1 of *Nothotsuga*, *Tsuga chinensis*, *Tsuga mertensiana*, *Pseudolarix*, *Larix*, and *Pseudotsuga*, and SSR2 of the former three taxa; Clade IV harbored all SSRs of *Keteleeria*,

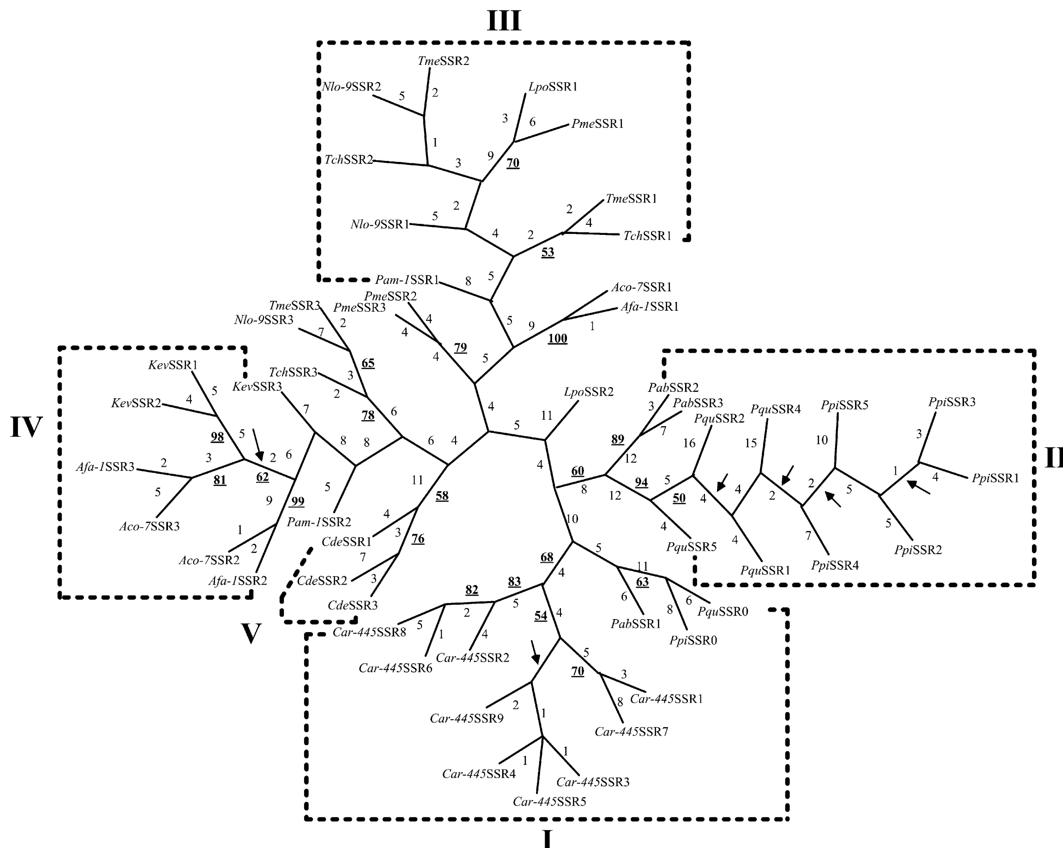

Fig. 2. One of the 355 most parsimomious trees (unrooted) constructed from sequence analysis of 51 SSRs from the 11 genera of Pinaceae (length = 472, CI = 0.4089, RI = 0.6872). Numbers on the branches denote branch lengths and bootstrap percentages (underlined) greater than 50%, respectively. The numbers following taxa indicate different clones. Arrows indicate clades collapsed in the strict consensus tree.

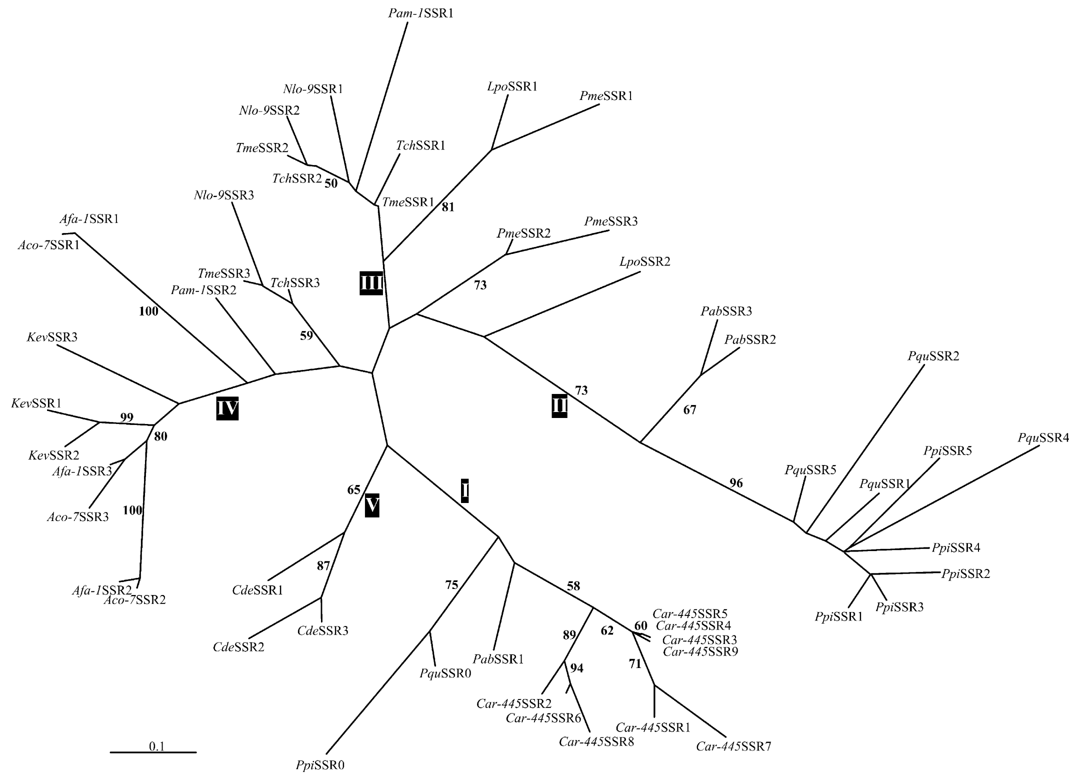

Fig. 3. The maximum-likelihood tree (unrooted) of the 51 SSRs from the 11 genera of Pinaceae. Numbers on the branches denote bootstrap percentages greater than 50%, and the numbers following taxa indicate different clones.

and SSR2 and SSR3 of *A. concolor* and *A. fabri*; Clade V was formed by all of the three SSRs of *Cedrus*. In addition, SSR1 of the two *Abies* species formed a strongly supported group, while SSR3 of *Nothotsuga* and *Tsuga* species clustered together. Considering the high intergeneric divergence of SSRs, we reset the data by reducing the sequences flanking the core motif (GGCCACCCTAGTC) to 71 bp, and conducted a further MP analysis using this data set. Consequently, this analysis still got very similar result.

3.4. Secondary structure of Pinaceae ITS1

We obtained 563 lowest free-energy folds from the 11 genera of Pinaceae (Table 4). In *A. concolor*, the three LSRs of *Aco-7* could pair each other, forming LSR1-2, LSR2-3, and LSR1–3 pairing (Fig. 4). Furthermore, SSR2 paired with SSR3, while SSR1 was alone, with a double hairpin at the tip, containing three loops. The first four bases (GGCC) of the conserved motif of SSR1 were exposed in the hairpin loops, while paired subrepeats partially overlapped at the conserved motif (5′-GGCCACCC TAGTC-3′) (Fig. 4 and Table 4). In *A. fabri*, the only one copy of LSR was not paired with the three SSRs, and formed a long hairpin by itself (data not shown).

For *C. argyrophylla*, all nine SSRs could result in complex predicted configurations. We analyzed all 41-folds of *Car-445*, of which 26 (63.4%) had three pairings among SSRs, respectively, with the remaining 15-folds having only two pairings. In all pairing arms, most bases of SSRs were paired. For example, in the first lowest free-energy fold of *Car-445*, SSR5–SSR6 arms were created by pairing 122 bases of each SSR, leaving 59 nucleotides of SSR5, and 52 nucleotides of SSR6 unpaired in bulges or small hairpins.

In *P. quadrifolia*, two types of configurations were found in the threefolds. In the first type ($\Delta G = -1118.0$ kcal/mol), SSR1 paired with SSR2, and SSR4 paired with SSR5, while SSR0 formed a large hairpin by itself. In the second type, SSR1 and SSR2 paired each other, while SSR0, SSR4, and SSR5 formed large hairpins respectively, with a part (GGCC) of the conserved motif of SSR4 and SSR5 exposed in a loop. We found only one type of configuration in the three folds of *P. pinea*, in which SSR1 paired SSR2, and SSR3 paired SSR4 (Table 4).

The three SSRs were unpaired in *T. chinensis*, with two configurations in the three folds. In contrast, SSR1 and SSR2 of the congeneric species *T. mertensiana* paired with each other. In the five genera *Cedurs*, *Keteleeria*, *Larix*, *Pseudolarix*, and *Picea*, SSRs were also unpaired. Notably, one double hairpin appeared in the tip of SSR1 and SSR3 of *Keteleeria evelyniana*, respectively, with -GGCC- and -CCACC- of the conserved motifs of the two SSRs exposed in the terminal loops correspondingly. Two types of configurations were found in the threefolds of *Larix potaninni*. In the first configuration, GGCCA of the conserved motif of SSR1 and GGCC of the conserved motif of SSR2 emerged in hairpin loops, while the conserved motif of SSR2 was

Table 4
ITS1 secondary structural features based on Mfold models in Pinaceae

Species/clones	Fold	ΔG[a]	SSR pairing[b]	Conserved motifs[c]									
				0	1	2	3	4	5	6	7	8	9
A. concolor C7[d]	39	−847.8	2–3, M	—	GGCC	Axial	Axial	—	—	—	—	—	—
A. concolor C8	39	−704.8	2–3, M	—	Axial	Axial	Axial	—	—	—	—	—	—
A. fabric C1	21	−608.4	2–3, M	—	GGCC	Axial	Axial	—	—	—	—	—	—
C. argyrophylla C 445	41	−1189.8	2–3, 5–6, 7–8, M	—	GGCC	Axial	Axial	Axial	Axial	Axial	Axial	Axial	Axial
		−1185.9	2–3, 5–6, 7–8, M	—	GGCC	Axial	Axial	Axial	Axial	Axial	Axial	Axial	CCACC
		−1184.7	2–3, 5–6, 7–9, M	—	GGCC	Axial	Axial	Axial	Axial	Axial	Axial	GG	Axial
C. deodara	43	−788.0	None	—	CCACC	Axial	Axial	—	—	—	—	—	—
K. evelyniana	47	−832.9	None	—	GGCC	CACCCUA	CCACC	—	—	—	—	—	—
L. potaninni	34	−623.1	None	—	GGCCA	GGCC	—	—	—	—	—	—	—
		−622.8	None	—	GGCCA	Axial	—	—	—	—	—	—	—
*N. longibracteata C*9	24	−672.7	1–2, M	—	Axial	Axial	GGCC	—	—	—	—	—	—
P. abies (AJ243165)	37	−1224.2	None	—	Axial	GGCCA	CCACC	—	—	—	—	—	—
P. abies (AJ243166)	37	−1601.8	None	—	Axial	GGCCA	CCACC	—	—	—	—	—	—
		−1599.3	None	—	CCACC	GGCCA	CCACC	—	—	—	—	—	—
P. pinea	32	−1106.9	1–2, 3–4, M	CCAC	Axial	Axial	Axial	Axial	CCUA	—	—	—	—
P. quadrifolia	29	−1118.0	1–2, 4–5, M	Axial	Axial	Axial	—	Axial	Axial	—	—	—	—
		−1115.5	1–2, M	Axial	Axial	Axial	—	GGCC	GGCC	—	—	—	—
P. amabilis C1	30	−457.9	None	—	GGCCA	GGCC	—	—	—	—	—	—	—
P. menziesii	50	−568.3	2–3, M	—	CCACC	Axial	Axial	—	—	—	—	—	—
T. chinensis	31	−642.7	None	—	GGCCA	Axial	GGCC	—	—	—	—	—	—
		−642.2	None	—	GGCCA	CCACC	GGCC	—	—	—	—	—	—
T. mertensiana	29	−682.9	1–2, M	—	Axial	Axial	GGCC	—	—	—	—	—	—

[a] Free energy value (kcal/mol), generated by Mfold software (Zuker, 2003).
[b] SSRs that pair with one another; M indicates that the SSRs pair for most of their lengths.
[c] Conserved motifs (CMs) are numbered as same as the SSR numbers; bases shown for a CM are those exposed in a hairpin.
[d] Numbers associated with taxa indicate different clones.

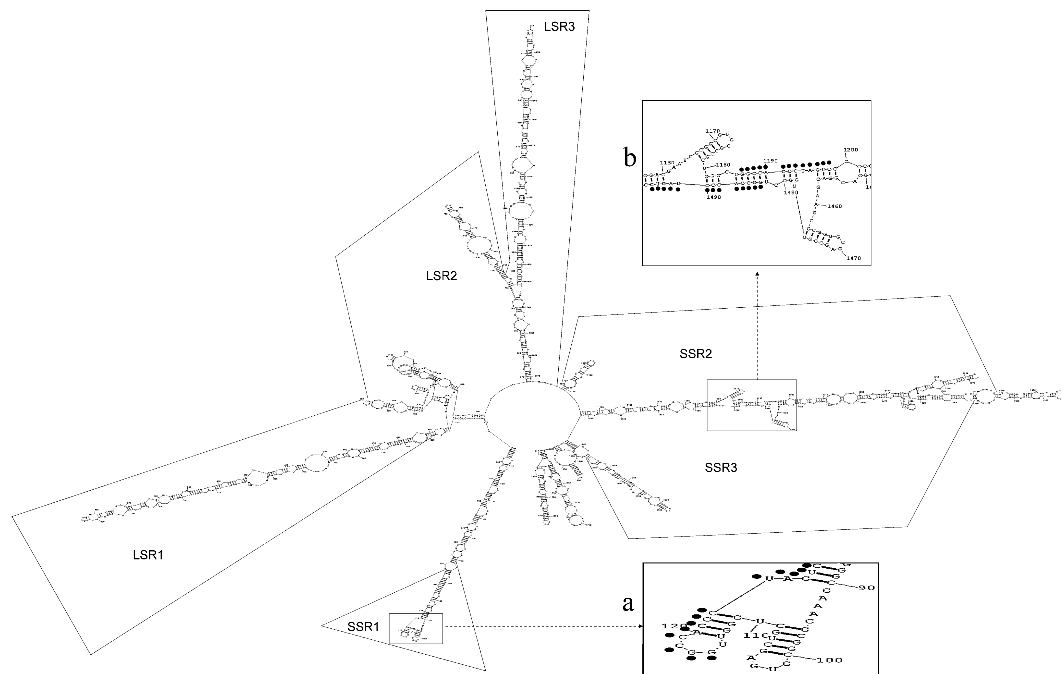

Fig. 4. Predicted secondary structure of RNA transcript of ITS1 of *Abies concolor* clone7 (*Aco-7*), showing LSRs and SSRs. (a) The double hairpin at the tip of SSR1 comprising three loops. (b) The first four bases (GGCC) of the conserved motif of SSR2 and SSR3, marked with the black dot, pair with one another.

not involved in the hairpin in the second configuration. The two adjacent SSRs, SSR1, and SSR2 in *Nothotsuga* and SSR2 and SSR3 in *Pseudotsuga*, paired with each other (Table 4).

4. Discussion

4.1. Length variation, subrepeat number, and composition in the ITS1 region of Pinaceae and their phylogenetic implications

Subrepeat arrays in nuclear ribosomal DNA (nrDNA) internal transcribed spacer 1 (ITS1) have been reported from trematodes (Kane and Rollinson, 1994; Kane et al., 1996; Dvorak et al., 2002; Warberg et al., 2005), insectas (Paskewitz et al., 1993; Tang et al., 1996; van Herwerden et al., 1998, 1999; von der Schulenburg et al., 2001) and gymnosperms (Gernandt and Liston, 1999; Maggini et al., 2000; Gernandt et al., 2001; Campbell et al., 2005). A better understanding of variation patterns of subrepeat number, composition and position in the ITS1 region will provide insights into the evolution dynamics of nrDNA, especially its structure and length evolution. Previous studies on Pinaceae ITS1 showed that LSR, longer subrepeat without the conserved motif (5′-GGCCACCCTAGTC-3′), occurred only in *Picea* (Maggini et al., 2000; Campbell et al., 2005), while SSR, shorter subrepeat with the conserved motif, existed in all seven pinaceous genera studied (Marrocco et al., 1996; Gernandt and Liston, 1999; Xiang et al., 2000; Gernandt et al., 2001; Wei et al., 2003; Campbell et al., 2005). The wide distribution of SSR is further corroborated by its occurrence in all 11 genera of the pine family found in the present study. In particular, our results indicate that LSR occurs not only in *Picea* but also in *Abies*, and its number has intragenomic variation (Fig. 1 and Table 3). *Picea* and *Abies* have origins in the Cretaceous and the Eocene respectively, based on estimation of the molecular clock of the *mat*K gene, and belong respectively to two main clades of Pinaceae (Wang et al., 2000), one comprising *Cathaya*, *Picea*, *Pinus*, *Larix*, and *Pseudotsuga* (hereafter referred to as Pinaceae-1), and the other comprising *Abies*, *Keteleeria*, *Nothotsuga*, *Pseudolarix*, and *Tsuga* (hereafter referred to as Pinaceae-2). The occurrence of LSR in the two genera is very likely to represent molecular convergence. However, it can not be completely ruled out that this type of subrepeat might descend from the common ancestor of Pinaceae, and then was lost in most genera of the pine family.

Although most genera or species do not form a monophyletic group in the SSR phylogenies we obtained, the homology of some SSRs still can be recognized, providing important informations for the evolutionary history of nrDNA ITS and phylogeny of Pinaceae. In Pinaceae-1, *Cathaya*, *Picea*, and *Pinus* are closely related, and *Larix* has a sister relationship with *Pseudotsuga* (Wang et al., 2000). Correspondingly, in the MP and ML trees, clades I and II are composed of SSRs of *Cathaya*, *Picea* and *Pinus*, and SSR1 of *Larix* and *Pseudotsuga* forms a sister group in clade III (Figs. 2 and 3). In Pinaceae-2, *Abies* is sister to *Keteleeria* and *Pseudolarix* is sister to a subclade comprising *Nothotsuga* and *Tsuga* (Wang et al., 2000). These relationships are also supported by the SSR

phylogenies, in which clade IV only harbors SSRs of *Abies* and *Keteleeria*, and SSR3 of *Nothotsuga* and *Tsuga* species clusters together (Figs. 2 and 3). The congruence between the phylogeny of SSRs and that of Pinaceae indicates that the homology of some SSRs has been resolved. In particular, some SSRs with the same number but from different genera or species, such as SSR1 of *Larix* and *Pseudotsuga* in clade III and SSR2 of the two studied *Abies* species in clade IV, form sister groups and are much less divergent than SSRs tandemly arranged in the same clone. These sister SSRs are very likely homologous.

Surprisingly, *Cathaya* has much more SSRs than the other genera of Pinaceae, and its SSRs are obviously from within-species or within-genera duplication since all nine SSRs of the genus form a monophyletic group. It is interesting that the SSRs of *C. argyrophylla* are further divided into two subgroups, one including SSR1, SSR3, SSR4, SSR5, SSR7, and SSR9, and the other comprising SSR2, SSR6, and SSR8 (Figs. 2 and 3). This grouping order of SSRs does not conform with the finding of Campbell et al. (2005) that adjacent tandem SSRs are more closely related to one another than they are to remote SSRs in the pinaceous genera, and may imply that multiple structure variations such as recombination have occurred in the evolution of the ITS1 region of this species. This kind of structural mutation could also have happened in the other genera of Pinaceae such as *Abies* and *Keteleeria*, considering the clustering of SSRs in clades II and IV (Figs. 2 and 3). Two mechanisms of recombination are unequal crossing over (Smith, 1976) and gene conversion (Nagylaki and Petes, 1982; Nagylaki, 1984; Hillis et al., 1991), which are commonly accepted to be responsible for the concerted evolution of nrDNA (Nei and Rooney, 2005; Eickbush and Eickbush, 2007). Gernandt et al. (2001) suggested that the recombination between non-homologous subrepeats most likely occurred at the highly conserved core sequence, and that unequal crossing over resulted in the number variation of tandem subrepeats in Pinaceae. This kind of recombination may also be responsible for the deletion of the conserved motif sequence, which could give rise to the LSRs in the Pinaceae ITS1. As discussed above, multiple duplication of subrepeats and accompanying recombinations might explain the different arrangement of LSR and SSR in various genera of the pine family, population sampling will be very necessary to investigate the evolutionary dynamics of the ITS1 subrepeats.

Pinaceae ITS1 ranges in size from 944 to 3271 bp, exhibiting one of the largest intergeneric length variation in eukaryotes recorded to date. This extensive variation could be mainly attributed to the number variation of subrepeats. The three genera *Picea*, *Pinus*, and *Cathaya* have the longest ITS1, being 2747–3271, 2532–2751, and 2730–2731 bp, respectively (Maggini et al., 2000; Gernandt et al., 2001; Campbell et al., 2005; Table 2). Correspondingly, these genera have five to nine subrepeats, including both LSRs and SSRs, but the other pinaceous genera have only two to three subrepeats, excluding LSRs of *Abies* due to its intragenomic number variation (Table 3 and Fig. 1). Campbell et al. (2005) concluded that number of subrepeats did not appear to be phylogenetically informative in Pinaceae. In the present study, we analyzed the ITS1 sequence data of all 11 Pinaceae genera, and found that the three closely related genera *Picea*, *Pinus*, and *Cathaya*, which form a trichotomy in Pinaceae-1 (Wang et al., 2000), have much more subrepeats than the other genera, as mentioned above. In other words, the number of subrepeats could have some phylogenetic information.

4.2. GC content and its phylogenetic information

The GC content is very important for organism genomes (Bourgon et al., 2004; Arndt et al., 2005; Benovoy et al., 2005; Cohen et al., 2005; Ebersberger and Meyer, 2005). Oliver and Marin (1996) revealed a relationship between GC content and coding-sequence length, and Arndt et al. (2005) found that GC content of surrounding sequences is the best predictor of the rates of substitution. Furthermore, GC content is correlated with various genomic features such as repeat element distribution, methylation pattern (Jabbari and Bernardi, 1998), and, most remarkably, gene density. GC-rich regions include many genes with short introns while GC-poor regions are essentially deserts of genes (Galtier et al., 2001). The GC/AT content has spiked as genomic punctuation marks (Zhang et al., 2004). Allelic recombination does increase GC content (Fullerton et al., 2001; Galtier et al., 2001), and ectopic recombination between clustered, tandemly repeated genes can also increase their GC contents (Galtier, 2003; Benovoy et al., 2005). All these results suggest that GC content could have some functional relevance. In this study, we found that GC content is higher in the subrepeat region than in the non-repeat region for Pinaceae ITS1, with the exception of *Picea* (Table 3). Furthermore, it is obvious that ITS1 GC content is much lower in *Picea*, *Pinus*, and *Cathaya* than in the other genera (Table 2). As discussed formerly, the three genera are the most closely related in Pinaceae-1, and have the longest ITS1 and the highest number of subrepeats. Therefore, it could be inferred that GC content in the ITS1 region is relevant to its sequence length and subrepeat number, and may provide phylogenetic information for Pinaceae.

4.3. Secondary structure of ITS1

In most eukaryotes investigated so far, the secondary structure of ITS1, consisting of an open multibranch loop with several helices, is conserved at higher systematic levels, but may differ in number of helices and consistency of the loop (Gottschling et al., 2001; von der Schulenburg et al., 2001; Mayol and Rossello, 2001; Gottschling and Plotner, 2004; Campbell et al., 2005). In Pinaceae, the secondary structure of ITS1 has been reported from seven genera (Gernandt and Liston, 1999; Campbell et al., 2005). Our results combined with previous studies show several

characteristics of ITS1 secondary structure for the pine family as follows: (1) the structure is dominated by several extended hairpins (Fig. 4); (2) the configuration complexity is positively correlated with the number of subrepeats, e.g., *Cathaya*, with nine subrepeats, has 41-folds (Table 4); (3) Paired subrepeats often partially overlap at the conserved motif (5′-GGCCACCCTAGTC), and form a long stem, while other subrepeats fold onto itself, leaving part of the conserved motif exposed in hairpin loops (Fig. 4 and Table 4).

Acknowledgments

The authors thank the two anonymous reviewers for their insightful comments and suggestions on the manuscript. We also thank Prof. Zheng-Yu Liu for his help in the sample collection of *Cathaya*; Ms. Wan-Qing Jin for her assistance in the DNA sequencing. This study was supported by the National Natural Science Foundation of China (Grant Nos. 30425028, 30121003) and the Chinese Academy of Sciences (the special fund to Xiao-Quan Wang).

References

Alvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol 29, 417–434.

Arndt, P.F., Hwa, T., Petrov, D.A., 2005. Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. J. Mol. Evol. 60, 748–763.

Bailey, C.D., Carr, T.G., Harris, S.A., Hughes, C.E., 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol. Phylogenet. Evol. 29, 435–455.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J., 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. 82, 247–277.

Benovoy, D., Morris, R.T., Morin, A., Drouin, G., 2005. Ectopic gene conversions increase the G + C content of duplicated yeast and Arabidopsis genes. Mol. Biol. Evol. 22, 1865–1868.

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

Bourgon, R., Delorenzi, M., Sargeant, T., Hodder, A.N., Crabb, B.S., Speed, T.P., 2004. The serine repeat antigen (SERA) gene family phylogeny in *Plasmodium*: the impact of GC content and reconciliation of gene and species trees. Mol. Biol. Evol. 21, 2161–2171.

Buckler IV, E.S., Holtsford, T.P., 1996a. *Zea* ribosomal repeat evolution and substitution patterns. Mol. Biol. Evol. 13, 623–632.

Buckler IV, E.S., Holtsford, T.P., 1996b. *Zea* systematics: ribosomal ITS evidence. Mol. Biol. Evol. 13, 612–622.

Campbell, C.S., Wright, W.A., Cox, M., Vining, T.F., Major, C.S., Arsenault, M.P., 2005. Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. Mol. Phylogenet. Evol. 35, 165–185.

Cohen, N., Dagan, T., Stone, L., Graur, D., 2005. GC composition of the human genome: in search of isochores. Mol. Biol. Evol. 22, 1260–1272.

Côté, C.A., Peculis, B.A., 2001. Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast. Nucleic Acids Res. 29, 2106–2116.

Dvorak, J., Vanacova, S., Hampl, V., Flegr, J., Horak, P., 2002. Comparison of european *Trichobilharzia* species based on ITS1 and ITS2 sequences. Parasitology 124, 307–313.

Ebersberger, I., Meyer, M., 2005. A genomic region evolving toward different GC contents in humans and chimpanzees indicates a recent and regionally limited shift in the mutation pattern. Mol. Biol. Evol. 22, 1240–1245.

Eickbush, T.H., Eickbush, D.G., 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175, 477–485.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Fullerton, S.M., Bernardo Carvalho, A., Clark, A.G., 2001. Local rates of recombination are positively correlated with GC content in the human genome. Mol. Biol. Evol. 18, 1139–1142.

Galtier, N., 2003. Gene conversion drives GC content evolution in mammalian histones. Trends Genet. 19, 65–68.

Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159, 907–911.

Gascuel, O., 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14, 685–695.

Germano, J., Klein, A.S., 1999. Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana* and *P. rubens*. Theor. Appl. Genet. 99, 37–49.

Gernandt, D.S., Liston, A., 1999. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). Am. J. Bot. 86, 711–723.

Gernandt, D.S., Liston, A., Pinero, D., 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications for molecular systematic studies of pine species complexes. Mol. Phylogenet. Evol. 21, 449–467.

Gottschling, M., Hilger, H.H., Wolf, M., Diane, N., 2001. Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of boraginales. Plant Biol. 3, 629–636.

Gottschling, M., Plotner, J., 2004. Secondary structure models of the nuclear internal transcribed spacer regions and 5.8S rRNA in *Calciodinelloideae* (Peridiniaceae) and other dinoflagellates. Nucleic Acids Res. 32, 307–315.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hillis, D.M., Moritz, C., Porter, C.A., Baker, R.J., 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251, 308–310.

Jabbari, K., Bernardi, G., 1998. CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families. Gene 224, 123–127.

Kane, R.A., Ridgers, I.L., Johnston, D.A., Rollinson, D., 1996. Repetitive sequences within the first internal transcribed spacer of ribosomal DNA in schistosomes contain a Chi-like site. Mol. Biochem. Parasitol. 75, 265–269.

Kane, R.A., Rollinson, D., 1994. Repetitive sequences in the ribosomal DNA internal transcribed spacer of *Schistosoma haematobium*, *Schistosoma intercalatum* and *Schistosoma mattheei*. Mol. Biochem. Parasitol. 63, 153–156.

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

Karvonen, P., Szmidt, A.E., Savolainen, O., 1994. Length variation in the internal transcribed spacers of ribosomal DNA in *Picea abies* and related species. Theor. Appl. Genet. 89, 969–974.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5, 150–163.

Lalev, A.I., Nazar, R.N., 1998. Conserved core structure in the internal transcribed spacer 1 of the *Schizosaccharomyces pombe* precursor ribosomal RNA. J. Mol. Biol. 284, 1341–1351.

Liston, A., Robinson, W.A., Oliphant, J.M., Alvarez-Buylla, E.R., 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. Syst. Bot. 21, 109–120.

Maggini, F., Frediani, M., Gelati, M.T., 2000. Nucleotide sequence of the internal transcribed spacers of ribosomal DNA in *Picea abies* Karst. DNA Seq. 11, 87–89.

Markos, S., Baldwin, B.G., 2002. Structure, molecular evolution, and phylogenetic utility of the 5′ region of the external transcribed spacer of 18S–26S rDNA in *Lessingia* (Compositae, Astereae). Mol. Phylogenet. Evol. 23, 214–228.

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

Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911–940.

Mayol, M., Rossello, J.A., 2001. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. Mol. Phylogenet. Evol. 19, 167–176.

Michot, B., Joseph, N., Mazan, S., Bachellerie, J.P., 1999. Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences. Nucleic Acids Res. 27, 2271–2282.

Nagylaki, T., 1984. Evolution of multigene families under interchromosomal gene conversion. Proc. Natl. Acad. Sci. USA 81, 3796–3800.

Nagylaki, T., Petes, T.D., 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100, 315–337.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152.

Nickrent, D.L., Schuette, K.P., Starr, E.M., 1994. A molecular phylogeny of *Arceuthobium* (Viscaceae) based on nuclear ribosomal DNA internal transcribed spacer sequences. Am. J. Bot. 81, 1149–1160.

Oliver, J.L., Marin, A., 1996. A relationship between GC content and coding-sequence length. J. Mol. Evol. 43, 216–223.

Paskewitz, S.M., Wesson, D.M., Collins, F.H., 1993. The internal transcribed spacers of ribosomal DNA in five members of the *Anopheles gambiae* species complex. Insect Mol. Biol. 2, 247–257.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Razafimandimbison, S.G., Kellogg, E.A., Bremer, B., 2004. Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from Naucleeae (Rubiaceae). Syst. Biol. 53, 177–192.

Rogers, S.O., Bendich, A.J., 1988. Extraction of DNA from plant tissues. Plant Mol. Biol. Manual A 6, 1–10.

Smith, G.P., 1976. Evolution of repeated DNA sequences by unequal crossover. Science 191, 528–535.

Swofford, D., 2002. PAUP*. Phylogenetic Analysis using Parsimony (and other methods). Version 4. Sinauer, SunderlandMass.

Tang, J., Toe, L., Back, C., Unnasch, T.R., 1996. Intra-specific heterogeneity of the rDNA internal transcribed spacer in the *Simulium damnosum* (Diptera: Simuliidae) complex. Mol. Biol. Evol. 13, 244–252.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

van Herwerden, L., Blair, D., Agatsuma, T., 1998. Intra- and inter-specific variation in nuclear ribosomal internal transcribed spacer 1 of the *Schistosoma japonicum* species complex. Parasitology 116 (Pt. 4), 311–317.

van Herwerden, L., Blair, D., Agatsuma, T., 1999. Intra- and interindividual variation in ITS1 of *Paragonimus westermani* (Trematoda: digenea) and related species: implications for phylogenetic studies. Mol. Phylogenet. Evol. 12, 67–73.

Vining, T.F., Campbell, C.S., 1997. Phylogenetic signal in sequence repeats within nuclear ribosomal DNA internal transcribed spacer 1 in *Tsuga*. Am. J. Bot. 84 (Suppl.), 702.

von der Schulenburg, J.H., Hancock, J.M., Pagnamenta, A., Sloggett, J.J., Majerus, M.E., Hurst, G.D., 2001. Extreme length and length variation in the first ribosomal internal transcribed spacer of ladybird beetles (Coleoptera: Coccinellidae). Mol. Biol. Evol. 18, 648–660.

Wang, X.-Q., Tank, D.C., Sang, T., 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol. Biol. Evol. 17, 773–781.

Warberg, R., Jensen, K., Frydenberg, J., 2005. Repetitive sequences in the ITS1 region of ribosomal DNA in congeneric microphallid species (Trematoda: Digenea). Parasitol. Res. 97, 420–423.

Wei, X.-X., Wang, X.-Q., 2004. Recolonization and radiation in *Larix* (Pinaceae): evidence from nuclear ribosomal DNA paralogues. Mol. Ecol. 13, 3115–3123.

Wei, X.-X., Wang, X.-Q., Hong, D.-Y., 2003. Marked intragenomic heterogeneity and geographical differentiation of nrDNA ITS in *Larix potaninii* (Pinaceae). J. Mol. Evol. 57, 623–635.

Wendel, J.F., Schnabel, A., Seelanan, T., 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl. Acad. Sci. USA 92, 280–284.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, pp. 315–322.

Xiang, Q.-P., Xiang, Q.-Y., Liston, A., Fu, L.-G., Fu, D.-Z., 2000. Length variation of the nuclear ribosomal DNA internal transcribed spacer in the genus *Abies*, with reference to its systematic utility in Pinaceae. Acta. Bot. Sin. 42, 946–951.

Zhang, L., Kasif, S., Cantor, C.R., Broude, N.E., 2004. GC/AT-content spikes as genomic punctuation marks. Proc. Natl. Acad. Sci. USA 101, 16855–16860.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.