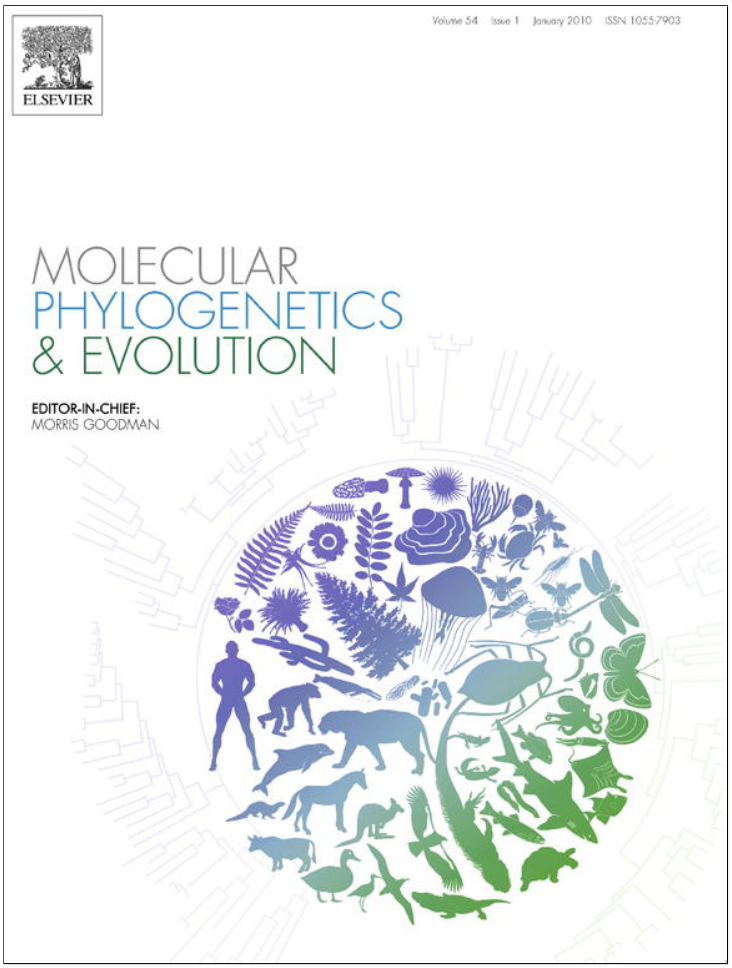

Molecular Phylogenetics and Evolution 54 (2010) 136–149

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Fast evolution of the retroprocessed mitochondrial *rps*3 gene in Conifer II and further evidence for the phylogeny of gymnosperms

Jin-Hua Ran, Hui Gao, Xiao-Quan Wang *

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

ARTICLE INFO

Article history:
Received 26 March 2009
Revised 8 September 2009
Accepted 9 September 2009
Available online 15 September 2009

Keywords:
Mitochondrial gene
*rps*3
Accelerated evolution
Retroprocessing
Intron loss
RNA editing
Conifer II

ABSTRACT

The popular view that plant mitochondrial genome evolves slowly in sequence has been recently challenged by the extraordinarily high substitution rates of mtDNA documented mainly from several angiosperm genera, but high substitution rate acceleration accompanied with great length variation has been very rarely reported in plant mitochondrial genes. Here, we studied evolution of the mitochondrial *rps*3 gene that encodes the ribosomal small subunit protein 3 and found a dramatically high variation in both length and sequence of an exon region of it in Conifer II. A sequence comparison between cDNA and genomic DNA showed that there are no RNA editing sites in the Conifer II *rps*3 gene. Southern blotting analyses of the total DNA and mtDNA, together with the real-time PCR analysis, showed that *rps*3 exists as a single mitochondrial locus in gymnosperms. It is very likely that the Conifer II *rps*3 gene has experienced retroprocessing, i.e., the re-integration of its cDNA into the mitochondrial genome, followed by an evolutionary acceleration due to the intron loss. In addition, the phylogenetic analysis of *rps*3 supports the sister relationship between conifers and Gnetales. In particular, the monophyly of conifer II is strongly supported by the shared loss of two *rps*3 introns. Our results also indicate that the mitochondrial gene tree would be affected in topology when the "edited" paralogs are analyzed together with their genomic sequences.

1. Introduction

A number of unique features are exhibited in the plant mitochondrial genome, such as frequent structural rearrangements (Clifton et al., 2004; Ogihara et al., 2005), incorporation of foreign DNA and gene transfer to the nuclear genome (Adams et al., 2000; Bergthorsson et al., 2004; Knoop, 2004; Wang et al., 2007), widespread RNA editing (Freyer et al., 1997; Handa, 2003; Chaw et al., 2008) and *trans*-splicing (Malek et al., 1997; Groth-Malonek et al., 2005). These have complicated the evolutionary analysis of plant mitochondrial genome, in contrast to its simple, circular and compact counterpart of about 16 kb in most animals (Boore, 1999). It has been widely accepted that plant mitochondrial DNA (mtDNA) evolves rapidly in structure, but slowly in sequence (Palmer and Herbon, 1988), with silent substitution rates less than one-third that of chloroplast DNA (cpDNA), one-sixth that of nuclear DNA (nDNA), and 100 times slower than that of animal mtDNA (Wolfe et al., 1987; Gaut, 1997; Muse, 2000; Handa, 2003; Drouin et al., 2008). However, in recent years, extraordinarily high substitution rates have been documented from mtDNA of several angiosperm genera, such as *Pelargonium* (Palmer et al., 2000; Parkinson et al., 2005; Bakker et al., 2006), *Plantago* (Cho et al., 2004; Bakker et al., 2006), *Silene* (Städler and Delph, 2002; Houliston and Olson, 2006; Barr et al., 2007; Mower et al., 2007; McCauley and Ellis, 2008; Sloan et al., 2008), and *Acorus* (Mower et al., 2007), intensifying the debate on the evolution of plant mtDNA. These studies indicate that substitution rates are highly accelerated in a mitochondrial gene or throughout a mitochondrial genome, although the underlying mechanisms could be variable, including horizontal gene transfer (e.g., Adams et al., 2002), mutation rate acceleration (Barr et al., 2007; Sloan et al., 2008) or absence of the DNA repair mechanisms (Cho et al., 2004; Parkinson et al., 2005; Bakker et al., 2006; Mower et al., 2007), ancient coalescence, and occasional paternal leakage (Städler and Delph, 2002; Houliston and Olson, 2006; McCauley and Ellis, 2008). To date, in plants, greatly accelerated sequence evolution of mtDNA has been reported mostly from herbaceous angiosperms and the accelerations have only been detected at relatively lower taxonomic levels, primarily within a genus or species, although Mower et al. (2007) found great increase in synonymous substitution rate of mtDNA in a couple of gymnosperms. According to our knowledge, high substitution rate acceleration accompanied with great length variation has been very rarely reported in plant mitochondrial genes.

Compared to the simple uniparental inheritance of cytoplasmic genes in angiosperms, i.e., predominantly maternal inheritance of

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

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.09.011

cpDNA and strictly maternal inheritance of mtDNA (Birky, 2008) except for rare paternal transmission of mtDNA in *Silene vulgaris* (McCauley et al., 2005), the inheritance pattern of organelle genomes is quite diverse in gymnosperms. For example, the chloroplast and mitochondrial genomes are paternally and maternally inherited in Pinaceae and Taxaceae, respectively, but both genomes are paternally inherited in Araucariaceae, Cupressaceae and Podocarpaceae and maternally inherited in Cycadaceae, Ephedraceae, Ginkgoaceae and Gnetaceae (Mogensen, 1996). Hence, the cytoplasmic DNA of gymnosperms may be tremendously different in variation pattern between families or main clades, providing us good materials for the study of plant mtDNA evolution. It is interesting that a higher silent substitution rate in both mt- and cp- DNA has been observed in gymnosperms except *Ephedra* when organelle is inherited paternally than inherited maternally, suggesting that the inheritance pathway could be another factor responsible for the evolutionary rate variation of cytoplasmic DNA (Whittle and Johnston, 2002). Moreover, Mower et al. (2007) found a wide range of variation in absolute substitution rates of mitochondrial genes of gymnosperms and, in particular, the absolute synonymous substitution rates (R_S) are very low in *Cycas*, *Ginkgo*, *Pinus* and *Zamia* but much higher in *Araucaria*, *Ephedra*, *Gentum*, *Juniperus* and *Podocarpus*. However, the two studies only sampled a few genera (12 and 9, respectively) of gymnosperms, although nine families were covered. Further investigation on mtDNA variation in gymnosperms with extensive sampling is obviously needed.

The plant mitochondrial genome contains three classes of genes, ribosomal RNA (rRNA), transfer RNA (tRNA), and protein-coding genes. Transfer of some protein-coding genes to the nuclear genome has been frequently reported (Nugent and Palmer, 1991; Adams et al., 2001, 2002), which leads to the difficulty in the exploration of their evolutionary rates. The *rps*3 gene encodes ribosomal small subunit protein 3, which plays important roles in the assembly of the small subunit at the last step and the assembly of other ribosomal proteins (Mizushima and Nomura, 1970; Ramakrishnan et al., 1986), and in repairing the oxidative DNA damage (Ahne et al., 1988; Neu et al., 1998). This gene is widely distributed in fungi, Rickettsia, Bradyrhizobium and plants with a C-terminal sequence motif conserved among all studied groups (Bullerwell et al., 2000; Chaw et al., 2008; Wang et al., 2008). According to Regina et al. (2005), length and nucleotide sequence of the *rps*3 gene are very conserved between angiosperms (e.g., *Helianthus annuus*, *Magnolia liliiflora*) and the gymnosperm *Cycas revoluta*, although it has been reported that the protein of this gene is highly variable in length and amino acid sequence in non-seed plants and some other organisms (e.g., Bullerwell et al., 2000). Interestingly, a novel additional group II intron was found in the *rps*3 gene of *Cycas revoluta*, and was regarded as a distinctive intron signature in gymnosperms (Regina et al., 2005). Moreover, the recent complete sequencing of the mitochondrial genome of *Cycas taitungensis* has showed a number of unique features of the gymnosperm mtDNA, such as abundant RNA editing sites and an exceptionally elevated nonsynonymous substitution rate for the protein-coding genes (Chaw et al., 2008). It would be of great interest to explore the evolution of protein-coding mitochondrial genes in the diverse gymnosperms.

In the present study, variation patterns of the *rps*3 gene were investigated in all gymnosperm families except Phyllocladaceae and Stangeriaceae, and its phylogenetic implications were discussed.

2. Materials and methods

2.1. Plant materials

We sampled 53 genera from all families of gymnosperms except Phyllocladaceae and Stangeriaceae, and two angiosperms *Helianthus annuus* and *Magnolia liliiflora* (Table 1). Voucher specimens are deposited in the herbarium of Institute of Botany, Chinese Academy of Sciences (PE). The DNA sequences obtained in this study are deposited in GenBank under Accession Numbers (Table 1). Sequences retrieved from GenBank include *rps*19, *rps*3 and *rpl*16 of *Cycas revoluta* (AY345867), *H. annuus* (AF319170) and *M. liliiflora* (AF319171), as well as *mat*K of gymnosperms (Table S1, Supplementary data).

2.2. DNA and RNA extraction, PCR amplification, cloning and sequencing

Total DNAs were extracted from fresh or dried leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987; Rogers and Bendich, 1988). The *rps*3 gene and its flanking regions, including partial sequences of the two genes *rps*19 and *rpl*16, were amplified with primers shown in Fig. 1 and Table S2 (Table S2, Supplementary data). PCR strategies were also described in Table S2. PCR products were separated by 1.5% agarose gel electrophoresis, cleaned with the TIANgel Midi Purification Kit (TIANGEN Biotech Co., Ltd., Beijing, China), and directly sequenced.

The same individuals of the six conifer species including *Abies holophylla*, *Araucaria heterophylla*, *Larix gmelinii* var. *principis-rupprechtii*, *Platycladus orientalis*, *Podocarpus macrophyllus* and *Taxus chinensis*, and *Ginkgo biloba* were also used in the RT-PCR analysis. Total RNAs were isolated from young leaves using the Trizol reagent (Invitrogen, Carlsbad, CA, USA). Poly (A) mRNA was purified from the total RNA using the Oligotex mRNA Mini Kit (Qiagen, Hilden, Germany) after the digestion by DNase I (RNase Free) (TakaRa Biotech Co., Dalian, China). The first strand cDNA was synthesized using SuperScript™II RNase H-free reverse transcriptase (Invitrogen, Carlsbad, CA, USA) with a Poly (T) primer (PTA) according to the manufacturer's instruction. The cDNAs were amplified by 3′-RACE with an adapter primer (AP) and primers *rps*3-E3F2 and *rps*3-E3F3, and by RT-PCR (Hemi-nested PCR) with primers *rps*19-F, *rps*3-E2F, *rps*3-E3R4 and *rpl*16-R (Fig. 1, Table S2, Supplementary data). The 3′-RACE and RT-PCR products were cloned with the pGEM-T Easy Vector System II (Promega, Madison, USA). For each taxon, at least 24 clones with correct insertion (determined by digestion with EcoRI) were picked and screened by comparing their restriction patterns of HinfI. All distinct clones were sequenced. After precipitation, the sequencing products were separated on a MegaBACE 1000 automatic sequencer (Amersham Biosciences, Buckinghamshire, UK) or a 96-capillary 3730XL DNA analyzer (Applied Biosystems, Foster City, CA, USA).

Real-time PCR was used to quantify the copy number of the *rps*3 gene. 0.5 μl and 0.05 μl total DNAs of *Picea smithiana* and *Taxus chinensis* were used in quantitative PCR reactions using the Rotor-Gene 2000 Real-Time Amplification System (Corbett Research, Eight Mile Plains, New South Wales, Australia). Serial dilutions of extracted plasmid DNA (from *Escherichia coli*) containing the target genes were used as the standards for real-time PCR. The target genes included a low copy nuclear gene *cad* encoding cinnamyl alcohol dehydrogenase from *P. smithiana*, a single copy nuclear gene *needly* (a homolog of *LEAFY*) from *T. chinensis*, and mitochondrial genes *nad*5 and *rps*3 from *P. smithiana* and *T. chinensis*, respectively. Primer pairs needly-E3Ff/needly-E3Rr, nad5-iF1/nad5-iR1, CAD-E6F/CAD-P2R, *rps*3i2-probF/*rps*3i2-probR and *rps*3-Taxus-F/*rps*3-Taxus-R used for the amplification of above genes were listed in Table S2 (Table S2, Supplementary data). PCRs were conducted using a Qiagen Quantitect™ SYBR® Green RT-PCR Kit (Qiagen, CA, USA) and melting analysis was routinely performed to check for the identity of PCR product. The PCR program included 40 cycles of denaturation at 95 °C for 20 s, annealing at 60 °C for 30 s, and elongation at 72 °C for 30 s. Melt curves, gel electrophoresis, and

Table 1
Sources of materials.

Family	Species	Sources/vouchers or DNA nos.	Length of *rps*3 (bp)						GenBank Accession No.
			*rps*3i1	*rps*3i2	E1	E2	E3	Total exon	
Pinaceae	*Abies holophylla* Maxim.	Botanic Garden, Institute of Botany, Beijing, China/Ran07001	1897	1672	74	193	1407	1674	FJ824837
	Cathaya argyrophylla Chun et Kuang	Dayao Mountain, Guangxi, China/Wang, X.-Q. DY08	2086	1471	74	193	1407	1674	FJ824833
	Cedrus deodara (Roxb.) G. Don	Botanic Garden, Institute of Botany, Beijing, China/Ran07002	2135	2051	74	193	1407	1674	FJ824835
	Keteleeria fortunei (A. Murray bis) Carrière	Hangzhou Botanical Garden, Zhejiang, China/9960	1071	1484	74	193	1404	1671	FJ824832
	Larix gmelinii (Rupr.) Kuzen. var. *principis-rupprechtii* (Mayr) Pilg.	Botanic Garden, Institute of Botany, Beijing, China/Ran07003	0	0	74	193	1428	1695	FJ824828
	Nothotsuga longibracteata (W.C. Cheng) Hu ex C.N. Page	Xinning, Hunan, China/LuoZC-002	981	1046	74	193	1410	1677	FJ824836
	Picea smithiana (Wall.) Boiss.	Botanic Garden, Institute of Botany, Beijing, China/Ran07004	2090	2443	74	196	1440	1710	FJ824838
	Pinus monophylla Torr. & Frém.	Califolia, USA/PineW14	2037	2282	74	193	1407	1674	FJ824834
	Pseudolarix amabilis (J. Nelson) Rehder	Lushan Botanical Garden, Jiangxi, China/Ran07015	1152	0	74	193	1404	1671	FJ824830
	Pseudotsuga sinensis Dode var. *gaussenii* (Flous) Silba	The Yellow Mountain, Anhui, China /025	0	0	74	193	1443	1710	FJ824829
	Tsuga dumosa (D. Don) Eichler	Yaojiaping, Sichuan, China/YJP30	868	1062	74	193	1410	1677	FJ824831
Conifer II									
Araucariaceae	*Agathis dammara* (Lamb.) Rich. & A. Rich.	South China Botanical Garden, Guangzhou, China (cultivated) /Ran07027	0	0	74	196	1887	2157	FJ843626
	Araucaria heterophylla (Salisb.) Franco	Botanic Garden, Institute of Botany, Beijing, China/Ran07010	0	0	74	196	1926	2196	FJ843625
Cephalotaxaceae	*Cephalotaxus sinensis* (Rehder & E.H. Wilson) H.L. Li	Botanic Garden, Institute of Botany, Beijing, China/Ran07011	0	0	74	205	2013	2292	FJ843623
Cupressaceae	*Actinostrobus pyramidalis* Miq.	Mount Annan Botanic Garden, NSW, Australia/865921	0	0	74	211	1656	1941	FJ843614
	Austrocedrus chilensis (D. Don) Pic. Serm. & Bizzarri	Royal Tasmanian Botanical Gardens, Tasmania, Australia (cultivated) /Ran08001	0	0	74	211	1680	>1965	FJ843615
	Callitris columellaris F. Muell.	Mt Tomah Botanic Gardens, NSW, Australia/15917	0	0	74	211	1686	1971	FJ843636
	Callitris macleayana (F. Muell.) F. Muell.	Mt Tomah Botanic Gardens, NSW, Australia/20020717	0	0	74	211	1614	1899	FJ843637
	Callitris rhomboidea R. Br. ex Rich. & A. Rich.	Mt Tomah Botanic Gardens, NSW, Australia/866983	0	0	74	211	1659	1944	FJ843635
	Calocedrus decurrens (Torr.) Florin	America/Chen weilie (2002.10.10)	0	0	74	205	1563	1842	FJ843639
	Chamaecyparis pisifera (Siebold & Zucc.) Endel.	Botanic Garden, Institute of Botany, Beijing, China/Ran07031	0	0	74	208	1626	1908	FJ843640
	Cryptomeria japonica (Thunb. ex L.f.) D. Don	Lushan Botanical Garden, Jiangxi, China/Ran07018	0	0	74	205	1536	1815	FJ843646
	Cunninghamia lanceolata (Lamb.) Hook.	Luding, Sichuan, China/W2003-2	0	0	74	205	1485	1764	FJ843647
	Cupressus funebris Endl.	Jinfo Mountains, Sichuan, China /CU001	0	0	74	217	1611	1902	FJ843632
	Diselma archeri Hook.f.	Mt Tomah Botanic Gardens, NSW, Australia/850340	0	0	74	208	1623	1905	FJ843638
	Fitzroya cupressoides (Molina) I.M. Johnst.	Royal Tasmanian Botanical Gardens, Tasmania, Australia (cultivated) /Ran08002	0	0	74	208	1638	1920	FJ843617
	Fokienia hodginsii (Dunn) A. Henry & H.H. Thomas	Lushan Botanical Garden, Jiangxi, China(cultivated)/Ran07028	0	0	74	208	1710	1992	FJ843616
	Glyptostrobus pensilis (Staunton ex D. Don) K. Koch	Hangzhou Botanical Garden, Hangzhou, China/Ran07019	0	0	74	205	1545	1824	FJ843642
	Juniperus squamata Buch.-Ham. Ex D. Don	Botanic Garden, Institute of Botany, Beijing, China/Ran07006	0	0	74	217	1611	1902	FJ843648
	Libocedrus bidwillii Hook. f.	Mt Tomah Botanic Gardens, NSW, Australia/876573	0	0	74	208	1599	1881	FJ843619

	Libocedrus yateensis Guillaumin	Mt Tomah Botanic Gardens, NSW, Australia/871397	0	0	74	208	1599	1881	FJ843629
	Metasequoia glyptostroboides Hu & W.C. Cheng	Botanic Garden, Institute of Botany, Beijing, China/Ran07007	0	0	74	205	1584	1863	FJ843633
	Pilgerodendron uviferum (D. Don) Florin	Royal Tasmanian Botanical Gardens, Tasmania, Australia (cultivated) /Ran08003	0	0	74	208	1599	1881	FJ843618
	Platycladus orientalis (L.) Franco	Botanic Garden, Institute of Botany, Beijing, China/Ran07014	0	0	74	205	1605	1884	FJ843630
	Sequoia sempervirens (D. Don) Endl.	Zhejiang University, Hangzhou, China/Ran07020	0	0	74	205	1620	1899	FJ843627
	Sequoiadendron giganteum (Lindl.) J. Buchholz	California, America/H-1996	0	0	74	208	1518	1800	FJ843628
	Taiwania cryptomerioides Hayata	Lushan Botanical Garden, Jiangxi, China/Ran07022	0	0	74	205	1536	1815	FJ843643
	Taxodium distichum (L.) Rich.	South China Botanical Garden, Guangdong, China/Ran07021	0	0	74	205	1545	1824	FJ843645
	Taxodium distichum (L.) Rich. var. *imbricatum* (Nutt.) Croom	Botanic Garden, Institute of Botany, Beijing, China/Ran07008	0	0	74	205	1545	1824	FJ843644
	Thuja koraiensis Nakai	Lushan Botanical Garden, Jiangxi, China/Ran07016	0	0	74	205	1659	1938	FJ843631
	Thujopsis dolabrata (Thunb. ex L.f.) Siebold et Zucc.	Lushan Botanical Garden, Jiangxi, China/Ran07017	0	0	74	205	1563	1842	FJ843641
	Widdringtonia nodiflora (L.) Powrie	Mt Tomah Botanic Gardens, NSW, Australia/961485	0	0	74	208	1644	1926	FJ843634
Podocarpaceae	*Dacrydium elatum* (Roxb.) Wall. ex Hook.	Hainan, China/WZS04	0	0	191	211	1443	1845	FJ843610
	Dacrycarpus imbricatus (Blume) de Laub.	Hainan, China/WZS02	0	0	191	211	1452	1854	FJ843611
	Nageia formosensis (Dummer) C.N. Page	Hangzhou Botanical Garden, Hangzhou, China(cultivated)/Ran07029	0	0	185	211	1455	1851	FJ843609
	Podocarpus macrophyllus (Thunb.) Sweet	Botanic Garden, Institute of Botany, Beijing, China/Ran07012	0	0	191	211	1455	1857	FJ843612
Sciadopityaceae	*Sciadopitys verticillata* (Thunb.) Siebold & Zucc.	Lushan Botanical Garden, Jiangxi, China/Ran07024	0	0	74	262	2025	2361	FJ843620
Taxaceae	*Amentotaxus argotaenia* (Hance) Pilg.	Xingshan, Huibei, China/11282a	0	0	74	193	1875	2142	FJ843622
	Pseudotaxus chienii (W.C. Cheng) W.C. Cheng	Lushan Botanical Garden, Jiangxi, Chian/Ran07023	0	0	74	193	1359	1626	FJ843613
	Taxus chinensis (Pilg.) Rehder	Botanic Garden, Institute of Botany, Beijing, China/Ran07009	0	0	74	199	1404	1677	FJ843624
	Torreya californica Torr.	Jordan Botanic Garden, Geneva, Switzerland/WXQ2142	0	0	74	193	2142	2409	FJ843621
Gnetales									
Ephedraceae	*Ephedra equisetina* Bunge	Botanic Garden, Institute of Botany, Beijing, China/Ran07025	–	–	–	–	–	–	
Gnetaceae	*Gnetum montanum* Markgr.	Napo, Guangxi, China/1849	3795	1261	?	196	1851	>2057	FJ824839
Welwitschiaceae	*Welwitschia mirabilis* Hook. fil.	Shenzhen Fairy Lake Botanical Garden, Shenzhen, China (cultivated)/Ran07030	?	0	?	>89	2160	>2249	FJ843649
Ginkgoaceae	*Ginkgo biloba* Linn.	Botanic Garden, Institute of Botany, Beijing, China/Ran07013	2278	983	74	193	1473	1740	FJ824840
Cycadales									
Cycadaceae	*Cycas revoluta* Thunb.	GenBank	2981	1985	74	193	1473	1740	AY345867
Zamiaceae	*Zamia furfuracea* L.	Beijing Botanical Garden, Beijing, China/Ran07026	2758	1774	74	193	1455	1722	FJ824841
Outgroups									
	Helianthus annuus L.	GenBank	976	0	74	193	1416	1683	AF319170
	Magnolia liliiflora Desr.	GenBank	1866	0	74	193	1305	1572	AF319171

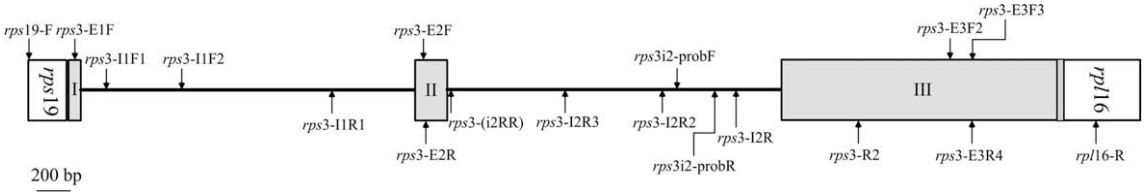

Fig. 1. Schematic organization of the mitochondrial *rps*19, *rps*3 and *rpl*16 gene cluster in *Cycas revoluta* (data from Regina et al., 2005), and the positions of the primers used in the study. Lines and boxes represent introns and exons, respectively. The grey boxes represent the exons of *rps*3, and the box with slants indicates the overlap between *rps*3 and *rpl*16.

sequencing were used to examine each sample for purity and specificity. Statistical analysis was performed using an unpaired one-tail T-test assuming unequal variance, and the level of significance was set at a *P* value of 0.05.

2.3. Southern blot analysis

Southern blotting was carried out to determine whether the *rps*3 gene exists as a single mitochondrial locus in gymnosperms. The genomic DNAs of *Ephedra equisetina*, *Ginkgo biloba*, *Picea smithiana*, and *Taxus chinensis* were isolated using the DNAsecure Plant Kit (TIANGEN, Beijing, China) and quantified by an Eppendorf spectrophotometer (Eppendorf, Hamburg, Germany). Meanwhile, the mitochondria of *T. chinensis* was extracted following the protocol of Li et al. (2000), and treated with DNase I to eliminate genomic DNA contamination. Over 15 μg total DNA of each species and about 0.1 μg mtDNA of *T. chinensis* were digested by the restriction enzymes EcoRV and HindIII (Takara Biotech Co., Dalian, China), respectively, separated on a 0.8% agarose gel and transferred to Hybond-N^+ membrane (Amersham Biosciences, Buckinghamshire, UK). Two probes, i.e., *rps*3 universal (the partial exon regions of *rps*3 and *rpl*16 conserved among gymnosperms) and *rps*3i2-specific (the conserved region of *rps*3 intron 2 in gymnosperms), were obtained by PCR amplification using the genomic DNA of *Ginkgo biloba*, and an additional *Taxus*-specific probe (an exon region of *rps*3 in *Taxus* that can not be aligned with other gymnosperms) was amplified from *T. chinensis* genomic DNA. The primer pairs used to amplify the *rps*3 universal (466 bp), rps3i2-specific (326 bp) and *Taxus*-specific (615 bp) probes are *rps*3-E3F3/*rpl*16-R, *rps*3-probF/*rps*3-probR and *rps*3-E2F/*rps*3-R2, respectively (Table S2, Supplementary data). In the *rps*3 universal probe region, EcoRV has a recognition site (338 bp in *Taxus*, 339 bp in *Ginkgo*, and 343 bp in *Picea*). Probes were labeled with alkaline phosphatase using a Gene Images AlkPhos Direct labeling and detection system (Amersham Biosciences, Buckinghamshire, UK). All hybridization processes were as described by the manufacturer. After being washed twice for 10 min each with primary wash buffer at 55 °C and then twice for 5 min each with secondary wash buffer at room temperature, the blots were drained and developed by the addition of CDP-Star detection reagent (30–40 μl/cm^2; Amersham Biosciences, Buckinghamshire, UK) for 2–5 min at room temperature. Then, the blots were wrapped and exposed to Hyperfilm (Amersham Pharmacia Biotech, Little Chalfont, UK).

2.4. Data analysis

Sequence alignments were made with CLUSTAL X (Thompson et al., 1997) and refined manually. Genomic DNA and cDNA sequences of the same species were compared with the BioEdit software (Hall, 1999). Phylogenetic trees were constructed based on nucleotide sequences of the conserved regions in *rps*19, *rps*3 and *rpl*16, using three methods, i.e., maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI). The MP analysis was performed in PAUP*4.0b10 (Swofford, 2002) using a heuristic search with 500 random addition sequence replicates, tree-bisection-reconnection (TBR) branch-swapping, MULTREES and a maximum of 1000 trees saved per round. All character states were equally weighted. Internal branch support was estimated with 500 bootstrap replicates (Felsenstein, 1985) using the same heuristic search settings except that a maximum of 100 trees were saved per round. The ML and BI analyses were performed in PHYML version 2.4.3 (Guindon and Gascuel, 2003) and MrBayes version 3.1.2 (Huelsenbeck et al., 2001; Ronquist and Huelsenbeck, 2003), respectively. Modeltest 3.7 (Posada and Crandall, 1998) and MrModeltest 2.2 (Nylander, 2004) were used to determine the best models of sequence evolution for ML and BI analyses, respectively. Based on Akaike Information Criterion (AIC), GTR + G + I was suggested as the best model for ML, and GTR + G for BI. Parameters for the ML analysis were optimized with a BIONJ tree as a starting point (Gascuel, 1997), and support values for nodes on the tree were estimated with 500 bootstrap replicates (Felsenstein, 1985). In the BI analysis, the MCMC algorithm was run for 1,000,000 generations with 4 incrementally heated chains, starting from random trees and sampling one out of every 100 generations. The first 300 samples for each run were discarded as burn-in. Phylogenetic inference was based on those trees sampled after generation 30,000. In addition, MacClade 4.08 (http://macclade.org) was used to estimate numbers of intron gains and losses of the *rps*3 gene in different gymnosperm lineages.

Sequence identities were calculated using the software BioEdit (Hall, 1999). Mega 4.1 (Tamura et al., 2007; Kumar et al., 2008) (http://www.megasoftware.net/) was applied to estimate nucleotide substitutions (*d*) for the *rps*3 gene using the maximum composite likelihood (MCL) method, and the standard errors of the estimates were calculated based on 500 bootstrap replicates. To test whether Conifer II (Bowe et al., 2000), also referred as *Cupressophyta* including all conifers except the pine family (Cantino et al., 2007; Rai et al., 2008), differs from other gymnosperms (excluding Conifer II and Gnetales) in evolutionary rate, Tajima's nonparametric relative rate test (Tajima, 1993) was applied as implemented in MEGA, using *Helianthus annuus* as the outgroup.

To detect individual sites under positive selection in Cupressaceae and Pinaceae, the two largest families that could represent the two main lineages of conifers, the likelihood ratio test (LRT) was carried out between each of the pairs of models M0/M3, M1a/M2a, and M7/M8 by the codeml program in PAML version 4 (Yang, 2007). Convergence was checked by performing the analysis with different starting parameters, particularly with d_N/d_S ratios of 0.5, 1.5 and 2.5 as recommended. The highest ln*L* among 3 trials was used for the computation of LRT. Codon frequencies were calculated by using the F3 × 4 model. An empirical Bayes approach was used to determine sites subjected to positive selection.

Absolute rates of synonymous (R_S) and nonsynonymous (R_N) substitutions in the conservative regions of the gymnosperm *rps*3 gene (excluding RNA editing sites) were calculated for a subset of the taxa that include all families we sampled and main clades in each family, following the method of Cho et al. (2004) and Parkinson et al. (2005). The ML topology of *mat*K was obtained by the

same method as that used for the *rps*3 gene mentioned previously, and TVM + I + G was suggested as the best model by Modeltest. Since the TVM model is not included in the PhyML program, the most sophisticated model GTR was used. Divergence times within gymnosperms were estimated based on the *mat*K gene topology. Rate constancy across lineages was examined using a likelihood ratio test (LRT) (Felsenstein, 1988). When a LRT compared likelihood scores with and without the clock assumption on the ML tree, the molecular clock was rejected in favor of the null hypothesis of no clock (δ = 1894.42, df = 25, $P < 0.001$). Thus, divergence times were finally estimated using penalized likelihood in r8s ver. 1.71 (Sanderson, 2003). A cross validation analysis was performed to obtain the most likely smoothing parameter. Standard errors of the divergence times were calculated with a nonparametric bootstrap procedure, in which 500 datasets were simulated with the SEQBOOT program in PHYLIP version 3.6a2 (Felsenstein, 1993) and were used to generate new ML trees with PAUP for divergence time estimation. Branch lengths, representing the number of substitutions per synonymous site (d_S) or number of substitutions per nonsynonymous site (d_N), were calculated for the *rps*3 gene using codeml in the PAML. The absolute rates of synonymous substitution per branch (R_S) and nonsynonymous substitution per branch (R_N) as well as standard errors for the d_S, d_N, R_S and R_N values were determined according to the formulas described by Parkinson et al. (2005).

3. Results

3.1. Sequence characterization, selection test and phylogenetic analysis

In this study, we obtained nucleotide sequence of a region comprising 3′ *rps*19 (198–236 bp), *rps*3, and 5′ *rpl*16 (186–191 bp) from all sampled gymnosperm genera except Gnetales (Table 1). For the *rps*3 gene, no intron was found in *Larix*, *Pseudotsuga* and all Conifer II families, only the first intron was found in *Pseudolarix*, and the second intron was absent in *Welwitschia* (no information about the first intron); all of the other samples analyzed had two introns (rps3i1 and rps3i2), showing great variation in length (Table 1). We did not get PCR product from *Ephedra equisetina*, although more than 10 primer pairs were tried. In addition, after trying more than five primer pairs, we still failed to get the first and second exons of *rps*3 and the *rps*19 gene from *Welwitschia mirabilis*. In *Gnetum montanum*, any primer located on the *rps*19 gene did not work, and thus this gene and the 5′ end of the *rps*3 exon 1 were not obtained from this species. A BLAST search (Altschul et al., 1997) showed that all rps3i1 sequences obtained in our study are similar to the group IIA introns found in the *rps*3 gene of *Cycas revoluta* and other land plants (Laroche and Bousquet, 1999; Dai et al., 2003), and all rps3i2 sequences are very similar to the rps3i2 of *C. revoluta* which was classified as a group II intron (Regina et al., 2005).

The length of *rps*3 exons is highly variable in gymnosperms, especially in Conifer II, ranging from 1626 bp (*Pseudotaxus chienii*) to 2409 bp (*Torreya californica*) (Table 1; Fig. S1, Supplementary data). The length variation was much lower in Araucariaceae (2157–2196 bp), Cupressaceae (1764–1992 bp), Pinaceae (1674–1710 bp) and Podocarpaceae (1842–1854 bp) than in Taxaceae (1626–2409 bp). For all *rps*3 sequences analyzed here, no length variation was found in the first exon (74 bp, except 185 bp or 191 bp in Podocarpaceae) and a narrow size range of 193–217 bp occurred in the second exon, whereas the third exon varied from 1359 bp to 2160 bp, showing an extraordinarily high variation in length and sequence (Fig. S1, Supplementary data). The lowest exon sequence similarity between genera in a single family was 0.890 (*Agathis-Araucaria*), 0.802 (*Cunninghamia-Fokienia*), 0.910 (*Picea-Pseudotsuga*), 0.969 (*Dacrydium-Nageia*), and 0.490 (*Taxus-Torreya*) for Araucariaceae, Cupressaceae, Pinaceae, Podocarpaceae and Taxaceae, respectively (Table S3, Supplementary data). The highest genetic distance between confamilial genera of these five families was 0.066 ± 0.006 (*Agathis-Araucaria*), 0.060 ± 0.006 (*Cunninghamia-Widdringtonia*), 0.054 ± 0.006 (*Cedrus-Pseudolarix*), 0.021 ± 0.004 (*Dacrydium-Nageia*), and 0.194 ± 0.012 (*Taxus-Torreya*), respectively (Table S4, Supplementary data). For the total length of *rps*3 exons, 1617 bp were alignable among Cycadaceae, Ginkgoaceae, Pinaceae and Zamiaceae that represent three out of the four major gymnosperm groups. However, only 885 bp were alignable among families of Conifer II. No sequence homologous or similar to the regions that can not be aligned among Conifer II families was found in GenBank. In addition, there were more indels in Cupressaceae than in Pinaceae in the sequence alignment.

The relative rate test was performed for a total of 795 bp of the *rps*3 exons that were reliably aligned among the selected representatives of gymnosperms (listed in Table S1, Supplementary data), and the pairwise comparisons found that the rate of nucleotide substitution was significantly higher in Conifer II than in the other gymnosperms comprising Pinaceae, Ginkgoaceae and Cycadales ($p < 0.01$ in all cases). The same pattern of nucleotide substitution rate difference was also found when using the 1st + 2nd and the 3rd codon positions, respectively (Fig. S2, Supplementary data).

The LRT test showed that all the three models (M2a, M3 and M8) which allow for selection were significantly favored over the other models ($P < 0.001$) in Cupressaceae, whereas only M3 gave significant support to the hypothesis of positive selection in Pinaceae. In addition, much more positively selected sites were found in Cupressaceae than in Pinaceae in all cases (Table 3).

The alignment of the combined *rps*19, *rps*3 and *rpl*16 exons had 4797 nucleotide sites, of which 1301 could be reliably aligned and were used in the phylogenetic analysis. The phylogenetic trees generated by PAUP, PHYML and MrBayes were topologically identical except that Gnetales was sister to Conifer II in the MP tree with a bootstrap support of 60%, which could be caused by long branch attraction (LBA) (Felsenstein, 1978; Hendy and Penny, 1989; Huelsenbeck, 1995). The MrBayes tree was shown in Fig. 2. The MacClade analysis based on the topology of the MrBayes tree showed that four independent *rps*3 intron losses could have occurred in gymnosperms, i.e., one in Conifer II, one in *Larix* and *Pseudotsuga*, one in *Pseudolarix*, and one in *Welwitschia* (Fig. 2). The former two lost both introns.

3.2. RT-PCR, real-time PCR and Southern blotting

As reported by Regina et al. (2005), our 3′-RACE and RT-PCR analyses also showed that the three genes, *rps*19, *rps*3 and *rpl*16, were transcribed together as polycistronic mRNAs. No RNA editing site was found in *Araucaria heterophylla*, *Larix gmelinii* var. *principis-rupprechtii*, *Platycladus orientalis*, *Podocarpus macrophyllus*, or *Taxus chinensis*. That is, all sequences of RT-PCR products from the above taxa were identical to the sequences obtained from genomic DNA. Correspondingly, the *rps*3 gene of these species does not have intron, as mentioned earlier. In contrast, about 30 and 36 RNA editing sites were detected from the direct sequences of RT-PCR products of *Abies holophylla* and *Ginkgo biloba*, respectively. These two species have introns in *rps*3. It is interesting that only nine editing sites were shared by *A. holophylla*, *G. biloba* and *Cycas revoluta*. In addition, we found that the number of RNA editing sites varied among different cDNA clones.

Southern blotting of genomic DNA with the *rps*3 universal probe showed that *Ginkgo biloba*, *Picea smithiana*, and *Taxus chinensis* had one *rps*3 locus. One recognition site of EcoRV in the target sequence gave rise to two signals in the sample digested by this restriction enzyme (1 weak signal in *Picea* cannot be seen). *Ephedra*

Table 2
Copies of the mitochondrial and nuclear genes in *Picea smithiana* and *Taxus chinensis* simulated by real-time PCR.

		Picea smithiana		*Taxus chinensis*	
		copy/ 0.5 μl	copy/0.05 μl	copy/0.5 μl	copy/0.05 μl
Low copy nuclear gene (Nc)	*needly*			3.59E + 3	4.65E + 2
	cad	2.55E + 3	3.65E + 2		
Mitochondrial gene (Mt)	*nad*5	5.52E + 4	7.89E + 3	1.17E + 6	1.11E + 5
	*rps*3	5.94E + 4	6.68E + 3	2.44E + 5	2.12E + 4
Mt/Nc	*nad*5/nuclear gene	22	21	326	238
	*rps*3/nuclear gene	23	18	68	68

equisetina, however, had no hybridization signal (Fig. S3, Supplementary data). When the *Taxus*-specific *rps*3 probe was used, hybridization signal was only detected from *Taxus chinensis*. As expected, no signal was detected in *T. chinensis* and *E. equisetina* by the rps3i2-specific probe (Figures not shown). When the above three probes were hybridized with the *T. chinensis* mitochondrial DNA, we obtained the same results as that of genomic DNA (Fig. 3, no signal detected by the rps3i2-specific probe), indicating that the *rps*3 gene we amplified is located in the mitochondrial genome.

The copy numbers of each gene in *Picea smithiana* and *Taxus chinensis* simulated by real-time PCR were shown in Table 2. Like the mitochondrial gene *nad*5, the *rps*3 gene has many more copies than the low copy nuclear gene *cad* and the single copy nuclear gene *needly*, suggesting that *rps*3 belongs to the mitochondrial genome. In *T. chinensis*, the estimated copy number of *nad*5 is much higher than that of *rps*3. This could have resulted from different specificity and efficiency of the primers used in this species.

3.3. Absolute substitution rates in the rps3 gene of gymnosperms

A chronogram of gymnosperms based on the topology of the *mat*K gene tree is shown in Fig. 4, with estimated ages of all branches listed in Table 4. The R_S value for the conservative regions of the *rps*3 gene is highly variable in gymnosperms. Especially, this value was nearly equal to zero synonymous substitution per site per billion years (SSB) in the most recent common ancestor (MRCA) of Pinaceae genera, in sharp contrast to the high value of 3.0712 SSB in MRCA of Conifer II. In addition, in Conifer II, the R_S value was much lower in MRCA of the two families, Podocarpaceae and Araucariaceae, mainly distributed in the Southern Hemisphere than in MRCA of Sciadopitys, Cephalotaxaceae, Taxaceae and Cupressaceae (Table 4; Fig. 4). Compared to the R_S value, the R_N value varied in a much narrower range (Table 4; Fig. 4).

4. Discussion

4.1. Unexpected fast evolution of the rps3 exon III in Conifer II

Recently, some studies suggest that plant mitochondrial genes could have variable and sometimes high substitution rates within a species or genus, such as *Pelargonium* (Parkinson et al., 2005; Bakker et al., 2006), *Plantago* (Cho et al., 2004; Bakker et al., 2006), *Silene* (Städler and Delph, 2002; Houliston and Olson, 2006; Barr et al., 2007; Mower et al., 2007; McCauley and Ellis, 2008; Sloan et al., 2008) and *Acorus* of angiosperm, and *Ephedra* of gymnosperm (Mower et al., 2007). In these previous studies, rapid mitochondrial gene evolution always shows acceleration in base substitution, especially in synonymous substitution, rather than in length or structure of coding regions. Surprisingly, our present study found that the *rps*3 gene of Conifer II has great variation in both exon length and nucleotide sequence, except in the

Table 3
Positively selected sites of the *rps*3 gene in Cupressaceae and Pinaceae inferred at P_b = 95%, with those reaching 99% shown in bold.

Model	Cupressaceae			Pinaceae		
	ln*L*	Estimates of parameters	Positively selected sites	ln*L*	Estimates of parameters	Positively selected sites
M0	−4283.6	ω = 0.63117	None	−3909.5	ω = 0.44667	None
M1a	−4217.1	p_0 = 0.67763 p_1 = 0.32237 ω_0 = 0.00000 ω_1 = 1.00000	Not allowed	−3885	p_0 = 0.60466 p_1 = 0.39534 ω_0 = 0.00000 ω_1 = 1.00000	Not allowed
M2a	−4181.6	p_0 = 0.75333 p_1 = 0.13724 p_2 = 0.10943 ω_0 = 0.06653 ω_1 = 1.00000 ω_2 = 4.90963	90 I, 316 F, 347 L, 488 E, 41 **I**, **203 Y**, **206 E**, **523 S**, **654 K**, **690 S**, **706 L**, **717 F**, **718 F**, **731 S**, **736 K**	−3880.6	p_0 = 0.70491 p_1 = 0.25329 p_2 = 0.04179 ω_0 = 0.09518 ω_1 = 1.00000 ω_2 = 3.74821	430 L
M3	−4179.7	p_0 = 0.85219 p_1 = 0.14554 p_2 = 0.00227 ω_0 = 0.12354 ω_1 = 3.99692 ω_2 = 19.02831	**41 I**, **82 A**, **90 I**, **100 I**, **103 K**, **151 G**, **168 Y**, **176 K**, **203 Y**, **206 E**, **220 V**, **224 I**, **316 F**, **347 L**, **348 G**, **488 E**, **508 K**, **523 S**, **552 L**, **654 K**, **690 S**, **706 L**, **710 T**, **717 F**, **718 F**, **731 S**, **736 K**, 42 V, 59 T, 79 H, 147 S, 182 V, 270 E, 398 G, 435 V, 447 L, 524 A, 536 L, 539 L, 546 L, 613 Q, 768 P	−3880.5	p_0 = 0.81157 p_1 = 0.17753 p_2 = 0.01090 ω_0 = 0.14540 ω_1 = 1.67652 ω_2 = 5.81925	**170 S**, **200 N**, **222 F**, **240 N**, **289 P**, **332 F**, **337 N**, **373 P**, **385 P**, **430 L**, **475 Q**, **543 S**, 179 S, 389 L
M7	−4219.4	p = 0.00555 q = 0.00839	Not allowed	−3885	p = 0.00500 q = 0.00746	Not allowed
M8	−4181.7	p_0 = 0.87114 p = 0.92470 q = 4.97240 (p_1 = 0.12886) ω = 4.58104	41 I, 90 I, 203 Y, 206 E, 316 F, 347 L, 488 E, 523 S, 654 K, 706 L, 717 F, 718 F, 731 S, 736 K, 82 A, 100 I, 103 K, 151 G, 168 Y, 176 K, 220 V, 224 I, 348 G, 508 K, 552 L, 690 S, 710 T	−3880.7	p_0 = 0.94799 p = 0.23235 q = 0.51178 (p_1 = 0.05201) ω = 3.49442	430 L, 543 S

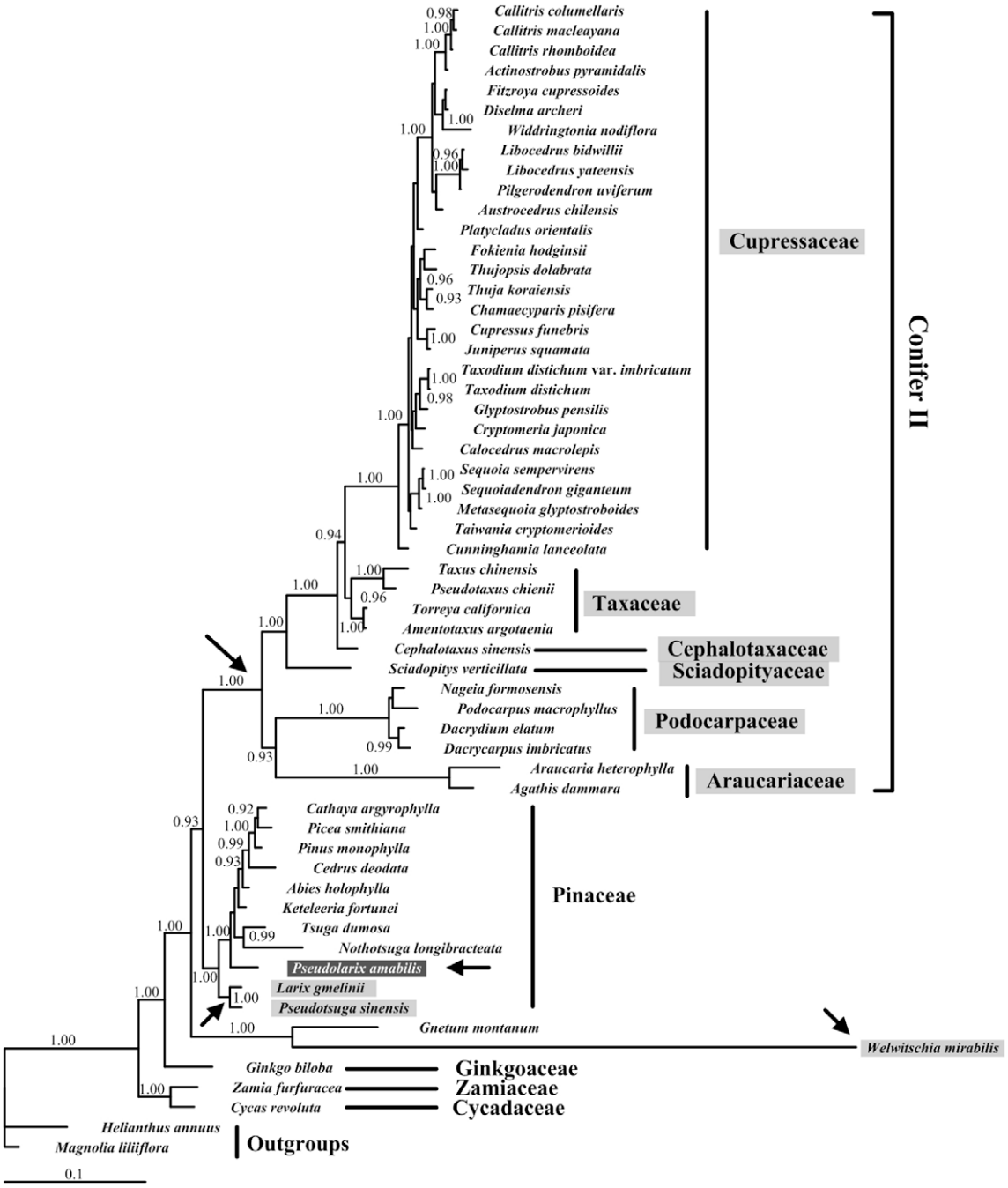

Fig. 2. Bayesian phylogram of gymnosperms constructed from sequence analysis of the *rps*3 gene with *Helianthus annuus* and *Magnolia liliiflora* serving as outgroups. Values on nodes represent the Bayesian posterior probabilities. Shaded taxa do not have intron and the taxon with black background has the first intron in the *rps*3 gene. The arrows indicate the events of intron loss.

functionally conserved regions at N- and C-terminal (Table 1; Fig. S1, Supplementary data). Most of its exon III sequence can not be aligned among families of Conifer II and the unalignable regions do not match any homologous sequences in GenBank. For example, the exon III of the shortest *rps*3 gene in Conifer II is 1359 bp (*Pseudotaxus chienii*), but only 632 bp are conserved and can be aligned. The highest genetic distance between genera of Conifer II reaches to 0.498 ± 0.019 (*Sciadopitys-Torreya*). However, *rps*3 is conserved between other gymnosperms (Cycadaceae, Zamiaceae, Ginkgoaceae and Pinaceae) and angiosperms (represented by *Helianthus annuus* and *Magnolia liliiflora*) in length and sequence. The highest genetic distance among genera of other gymnosperms and the sampled angiosperms is 0.218 ± 0.013 (*Cedrus-Helianthus*), a value much lower than that in Conifer II (Table S4, Supplementary data). Moreover, the relative rate test indicates that the rate of nucleotide substitution is significantly higher in Conifer II than in the other gymnosperms excluding Gnetales (Fig. S2, Supplementary data). It is of great interest that the R_S value in MRCA of Conifer II is much higher than in MRCA of its sister group Pinaceae (Fig. 4), and in particular the evolutionary rate acceleration in Conifer II could be much greater than the present estimate, since many unalignable regions in the *rps*3 exon III have been deleted in the calculation. The dramatically rapid divergence of *rps*3 in Conifer II in contrast to the conservativeness of this gene in most of seed

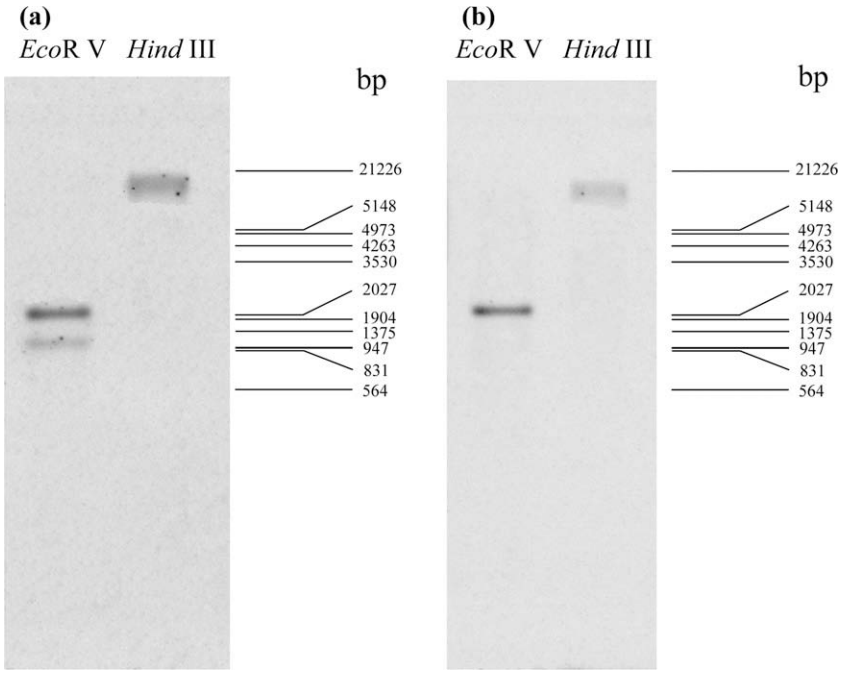

Fig. 3. Southern blot hybridization of *Taxus chinensis* mtDNA with the *rps*3 universal probe (a) and the *Taxus*-specific *rps*3 probe (b), using the two restriction enzymes EcoRV and HindIII, respectively. Numbers on the right indicate DNA molecular weight marker (Lamda DNA/EcoRI + HindIII).

plant lineages poses an interesting question concerning mitochondrial gene evolution.

It may be hypothesized that the *rps*3 gene of Conifer II that we obtained came from the nuclear rather than mitochondrial genome. That is, the *rps*3 gene could have been duplicated in the common ancestor of Conifer II, and then a duplicate copy was transferred to the nuclear genome or by mRNA-mediated transfer, where it evolved rapidly, as reported in some other genes (Blanchard and Schmidt, 1995; Laroche et al., 1997; Adams et al., 1999, 2002; Daley et al., 2002). However, this hypothesis is not supported by the Southern blotting analyses, which has proved that the *rps*3 gene has only one locus in gymnosperms and is located in mitochondria (Fig. 3; Fig. S3, Supplementary data). The real-time PCR analysis also indicates that *rps*3 comes from the mitochondrial genome (Table 2). In addition, it is also very unlikely that ancient coalescence or recent hybridization, two putative mechanisms responsible for the high variation rate in *Silene* (Städler and Delph, 2002; Houliston and Olson, 2006), have occurred between Conifer II families, because the divergence of these families dated back to at least the Late or Middle Jurassic (Cheng et al., 2000; Farjón, 2001).

Whittle and Johnston (2002) found that the organellar DNA of gymnosperms has a higher mutational rate when it is inherited paternally rather than maternally. However, this male-driven evolution hypothesis cannot explain the great length variation of *rps*3 in Conifer II, although the mitochondrial DNA of most Conifer II families is paternally inherited (Mogensen, 1996). The *rps*3 gene of the family Taxaceae has the largest length variation in Conifer II, ranging from 1626 bp to 2409 bp (Table 1), but this family has a maternally inherited mitochondria (Mogensen, 1996).

It is very likely that the Conifer II *rps*3 gene has experienced retroprocessing, i.e., the re-integration of its cDNA into the mitochondrial genome, followed by an evolutionary acceleration in functionally unimportant regions. Many previous studies showed that the retroprocessing process could result in the loss of introns (Bowe and dePamphilis, 1996; Bakker et al., 2006), and the retroprocessed mitochondrial genes were often characterized by the absence of RNA editing sites (Niu et al., 2005). The loss of two introns and absence of RNA editing in the Conifer II *rps*3 gene strongly suggest that this gene could have had a history of retroprocessing, given that RNA editing occurs in mitochondria of all major groups of vascular plants including Conifer II (Hiesel et al., 1994; Lu et al., 1998).

The mRNA-mediated intron loss might also have led to the evolutionary rate acceleration of *rps*3 in Conifer II. Firstly, a growing body of evidences suggests that introns play important roles in the maintenance of secondary structure of immature messenger RNAs (pre-mRNAs) and influence many stages of mRNA metabolism such as transcription and editing (Stephan and Kirby, 1993; Kirby et al., 1995; Rodríguez-Trelles et al., 2006). Secondly, some investigations showed that evolutionarily conserved genes preferentially accumulated introns whereas rapidly regulated genes were intron poor (Carmel et al., 2007; Jeffares et al., 2008). Thirdly, statistical analyses have proved a negative correlation between evolutionary rate at the protein level and intron size in *Drosophila* (Marais et al., 2005). The group II introns are the ancestor of spliceosomal intron (Lambowitz and Zimmerly, 2004; Lynch and Richardson, 2002), or share a common ancestor with the spliceosome (Toor et al., 2008). They can function not only as self-splicing ribozymes that catalyze their own excision from precursor-mRNAs (Pyle and Lambowitz, 2006) but also as retroelements (Robart et al., 2007; Toor et al., 2008). Hence, group II introns, like the introns of *rps*3, should have some similar functions with the spliceosomal intron, although there is no direct evidence that group II intron gain/loss affects the evolutionary rate of genes. In addition, the fact that the *rps*3 gene has more sites under positive selection in Cupressaceae (without intron) than in Pinaceae (with intron) also indicates that genes with introns are under stronger selective pressure (Table 3). Considering the dramatic difference in the R_S value between MRCAs of Conifer II and Pinaceae and the relatively consistent R_S values in other branches except Gnetales, the great *rps*3 evolutionary rate acceleration very likely occurred during the early divergence between Conifer II and Pinaceae.

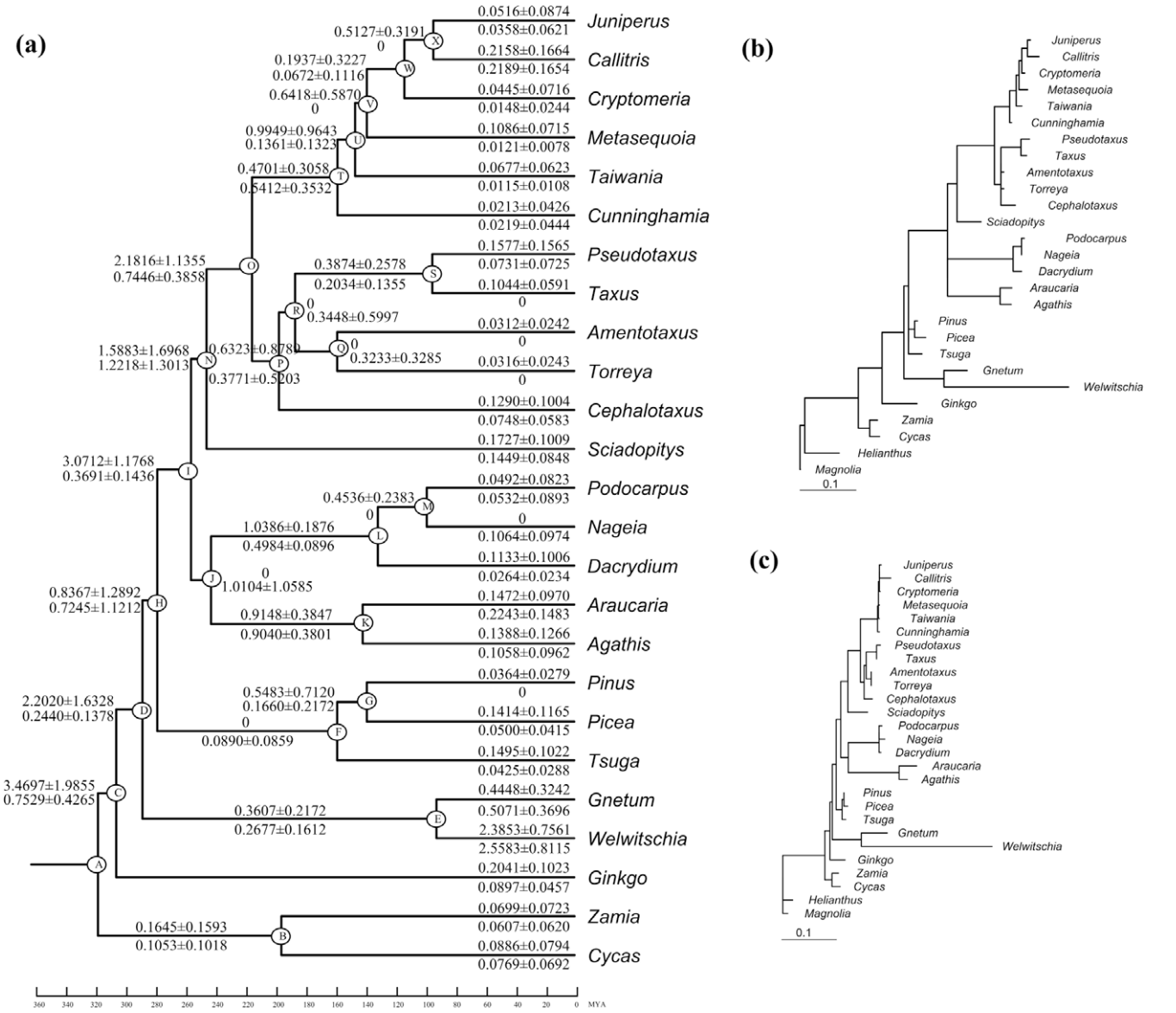

Fig. 4. Evolutionary rate variation (a) as well as synonymous (b) and nonsynonymous (c) sequence divergence in the conserved region of the gymnosperm *rps*3 gene. The chronogram in Fig. 4a was generated based on the ML topology of the *mat*K gene, with the absolute rates of synonymous (R_S) and nonsynonymous (R_N) substitutions of the *rps*3 gene plotted above and below branches, respectively. R_S and R_N values were calculated as in Table 4, using branch lengths from Fig. 4b and Fig. 4c, respectively. Divergence times for nodes A–X are listed in Table 4. MYA, million years ago. The most recent common ancestor (MRCA) of gymnosperms (node A) was fixed with an age of 321 Mya (Taylor and Taylor, 1993), and other nine nodes were constrained with minimal ages, including four (B, 174 Mya; C, 307 Mya; J, 245 Mya; N, 192 Mya) cited from Magallón and Sanderson (2005) as well as G at 140 Mya (Savard et al., 1994), M at 99.6 Mya (Krassilov, 1974; Kimura et al., 1988), Q at 160 Mya (Nicholson, 1990), V at 140 Mya (Ma et al., 2005; Farjón, 2005), and X at 95 Mya (McIver, 2001).

The highly variable region of the Conifer II *rps*3 genes is located in the third exon, which may suggest that this region is not so important for the secondary and crystal structure and functional expression. Although the negative correlation between the evolutionary rate of a gene and its functional importance has recently been debated, some extremely conserved genes (or DNA sequences) are very likely involved in important functions (Wang and Zhang, 2009). The intron loss in the Conifer II *rps*3 gene may have further resulted in the unexpected fast evolution of its third exon and given rise to the great length variation, due to the low functional constraints and structural stability of this region. This inference is supported by the fact that the exon length of *rps*3 has great variation in non-seed plants and some other organisms, except the conserved N- and C-terminal regions, such as 149 AA in *Monosiga brevicollis* (Bullerwell et al., 2000), 270 AA in *Chara vulgaris* (Turmel et al., 2003), and 499 AA in *Moniliophthora perniciosa* (E.F. Formighieri et al., unpublished).

4.2. Implications of the rps3 gene evolution for the phylogeny of gymnosperms

Rare genomic changes, such as intron gain/loss, large indels, gene order variation in the cytoplasm genome, retroposon insertion and coding frame shift, provide a number of complementary markers for molecular systematics (Rokas and Holland, 2000). Intron gain and loss are very common in plant mitochondrial genes (Adams et al., 2001; Joly et al., 2001; Knoop, 2004), and the events of mitochondrial gene (or its intron) loss have been successfully used to infer phylogenetic relationships at various taxonomic levels (Qiu et al., 1998; Adams and Palmer, 2003; Dombrovska and

Table 4
Divergence times, absolute synonymous and nonsynonymous substitution rates (R_S and R_N) of the *rps*3 gene for different branches of gymnosperms.

Branch	Time (MYA)	R_S (SSB)	R_N (SSB)
Pinaceae			
Picea	140 ± 5.74	0.1414 ± 0.1165	0.0500 ± 0.0415
Pinus	140 ± 5.74	0.0364 ± 0.0279	0
Tsuga	159.88 ± 7.92	0.1495 ± 0.1022	0.0425 ± 0.0288
Conifer II			
Callitris	95 ± 2.83	0.2158 ± 0.1664	0.2189 ± 0.1654
Juniperus	95 ± 2.83	0.0516 ± 0.0874	0.0358 ± 0.0621
Cryptomeria	114.70 ± 5.22	0.0445 ± 0.0716	0.0148 ± 0.0244
Metasequoia	140 ± 1.82	0.1086 ± 0.0715	0.0121 ± 0.0078
Taiwania	147.79 ± 3.25	0.0677 ± 0.0623	0.0115 ± 0.0108
Cunninghamia	159.55 ± 3.77	0.0213 ± 0.0426	0.0219 ± 0.0444
Taxus	95.74 ± 10.38	0.1044 ± 0.0591	0
Pseudotaxus	95.74 ± 10.38	0.1577 ± 0.1565	0.0731 ± 0.0725
Torreya	160 ± 0	0.0316 ± 0.0243	0
Amentotaxus	160 ± 0	0.0312 ± 0.0242	0
Cephalotaxus	199.17 ± 5.34	0.1290 ± 0.1004	0.0748 ± 0.0583
Sciadopitys	247.82 ± 3.43	0.1727 ± 0.1009	0.1449 ± 0.0848
Podocarpus	99.6 ± 0	0.0492 ± 0.0823	0.0532 ± 0.0893
Nageia	99.6 ± 0	0	0.1064 ± 0.0974
Dacrydium	132.45 ± 6.22	0.1133 ± 0.1006	0.0264 ± 0.0234
Agathis	142.68 ± 17.41	0.1388 ± 0.1266	0.1058 ± 0.0962
Araucaria	142.68 ± 17.41	0.1472 ± 0.0970	0.2243 ± 0.1483
Gnetales			
Gnetum	93.07 ± 7.27	0.4448 ± 0.3242	0.5071 ± 0.3696
Welwitschia	93.07 ± 7.27	2.3853 ± 0.7561	2.5583 ± 0.8115
Ginkgoaceae			
Ginkgo	308.78 ± 1.95	0.2041 ± 0.1023	0.0897 ± 0.0457
Cycadales			
Cycas	197.56 ± 19.18	0.0886 ± 0.0794	0.0769 ± 0.0692
Zamia	197.56 ± 19.18	0.0699 ± 0.0723	0.0607 ± 0.0620
A to B*	123.44 ± 19.18	0.1645 ± 0.1593	0.1053 ± 0.1018
A to C	12.22 ± 1.95	3.4697 ± 1.9855	0.7529 ± 0.4265
C to D	17.62 ± 3.19	2.2020 ± 1.6328	0.2440 ± 0.1378
D to E	197.95 ± 7.70	0.3607 ± 0.2172	0.2677 ± 0.1612
D to H	9.8 ± 3.45	0.8367 ± 1.2892	0.7245 ± 1.1212
H to F	121.34 ± 8.26	0	0.0890 ± 0.0859
F to G	19.88 ± 9.78	0.5483 ± 0.7120	0.1660 ± 0.2172
H to I	22.76 ± 3.52	3.0712 ± 1.1768	0.3691 ± 0.1436
I to J	13.46 ± 2.62	0	1.0104 ± 1.0585
J to K	102.32 ± 17.41	0.9148 ± 0.3847	0.9040 ± 0.3801
J to L	112.55 ± 6.22	1.0386 ± 0.1876	0.4984 ± 0.0896
L to M	32.85 ± 6.22	0.4536 ± 0.2383	0
I to N	10.64 ± 4.32	1.5883 ± 1.6968	1.2218 ± 1.3013
N to O	30.62 ± 5.78	2.1816 ± 1.1355	0.7446 ± 0.3858
O to P	18.03 ± 7.08	0.6323 ± 0.8789	0.3771 ± 0.5203
P to R	11.02 ± 7.38	0	0.3448 ± 0.5997
R to S	92.41 ± 11.56	0.3874 ± 0.2578	0.2034 ± 0.1355
R to Q	28.15 ± 5.09	0	0.3233 ± 0.3285
O to T	57.65 ± 5.99	0.4701 ± 0.3058	0.5412 ± 0.3532
T to U	11.76 ± 4.98	0.9949 ± 0.9643	0.1361 ± 0.1323
U to V	7.79 ± 3.72	0.6418 ± 0.5870	0
V to W	25.3 ± 5.23	0.1937 ± 0.3227	0.0672 ± 0.1116
W to X	19.70 ± 5.94	0.5127 ± 0.3191	0

* A–X correspond to the nodes in Fig. 4a.

Qiu, 2004; Regina et al., 2005). Previous studies have also suggested that the presence/absence and structural variation of Group II introns could be useful characters in the delimitation of taxa in land plants (De Benedetto et al., 1992; Rabbi and Wilson, 1993; Gugerli et al., 2001). For example, Gugerli et al. (2001) analyzed the second intron of the mitochondrial gene *nad*1 and found that the intron was absent in all non-Pinaceae conifers, *Welwitschia*, and seven angiosperms studied. They concluded that the intron loss was caused by a single ancient event in Conifer II but by several independent events in the other groups.

In the past 20 years, great efforts have been undertaken to improve our understanding of phylogenetic relationships within the gymnosperms (e.g., Donoghue and Doyle, 2000; Rydin and Källersjö, 2002; Rai et al., 2008). Much debate has focused on the position of Gnetales, which was considered as (1) the sister group of angiosperms (the anthophyte hypothesis, Doyle and Donoghue, 1986; Doyle, 1996), (2) the basal clade of seed plants (Hamby and Zimmer, 1992; Albert et al., 1994; Rai et al., 2003), (3) the basal clade of gymnosperms (Schmidt and Schneider-Poetsch, 2002), and (4) the sister group of conifers or Pinaceae (Goremykin et al., 1996; Chaw et al., 1997, 2000; Hansen et al., 1999; Samigullin et al., 1999; Soltis et al., 1999; Winter et al., 1999; Bowe et al., 2000; Gugerli et al., 2001; Magallón and Sanderson, 2002; Burleigh and Mathews, 2004, 2007; Mundry and Stützel, 2004; Hajibabaei et al., 2006; McCoy et al., 2008; Rai et al., 2008). So far, the last hypothesis, i.e., Gnetales as a sister group of Pinaceae or conifers, is supported by most of the molecular phylogenetic analyses, and thus is relatively widely accepted. However, the monophyly of conifers is still controversial due to the conflicting results among various molecular studies, especially with regard to the relationship between Pinaceae and Gnetales. For example, the monophyly of conifers is supported by the shared loss of a large inverted repeat in the chloroplast genome except the remnant of a short fragment less than 1 kb (Raubeson and Jansen, 1992; Wakasugi et al., 1994; Hirao et al., 2008; McCoy et al., 2008), whereas the hypothesis of Gnepines (Gnetales-Pinaceae) is consistent with the correlated loss of all *ndh* genes from the chloroplast genome of Gnetales and Pinaceae (Tsudzuki et al., 1992).

The evolutionary history of the *rps*3 gene has important implications for the phylogeny of gymnosperms. The *rps3* gene has only one intron (rps3i1), a typical and positionally conserved group II intron, in non-seed plants and angiosperms studied to date (Handa, 2003; Turmel et al., 2003), except that no intron was found in the liverwort *Marchantia polymorpha* (Oda et al., 1992) and the dicot *Beta vulgaris* (Kubo et al., 2000). An additional intron of *rps*3, rps3i2, was first reported from *Cycas revoluta* and *Ginkgo biloba* (Regina et al., 2005). In the present study, we found that both rps3i1 and rps3i2 are also present in *Zamia furfuracea*, *Gnetum montanum*, and all Pinaceae genera except for *Larix* and *Pseudotsuga* (two young genera with a sister relationship) without intron, and *Pseudolarix* with only rps3i1. In contrast, the two introns are absent in Conifer II (Table 1, Fig. 2). The wide occurrence of the two introns suggests that they represent symplesiomorphies, and four independent intron losses have occurred in gymnosperms, including two losses of both introns in the common ancestor of Conifer II and the *Larix-Pseudotsuga* clade, one loss of one or two introns in *Welwitschia*, and one loss of the second intron in *Pseudolarix*. That is, the shared loss of the two *rps*3 introns in Conifer II would be a synapomorphy supporting the monophyly of this conifer lineage as shown in Fig. 2. This inference is also consistent with the intron distribution pattern of *nad*1 (Gugerli et al., 2001), and results of some previous sequence-based phylogenetic studies (e.g., Goremykin et al., 1996; Chaw et al., 1997, 2000; Hansen et al., 1999; Samigullin et al., 1999; Burleigh and Mathews, 2007; Rai et al., 2008). However, as discussed above, the occurrence of the two *rps*3 introns in both *Gnetum montanum* and Pinaceae is very likely a symplesiomorphy, and thus should not be considered as evidence for supporting the sister group relationship between Gnetales and Pinaceae, the Gnepines hypothesis. In fact, the sister relationship between conifers and Gnetales is strongly supported by the *rps*3 gene tree (Fig. 2). In addition, the sister group relationships between *Larix* and *Pseudotsuga* and between *Cathaya*, *Picea* and *Pinus* (Wang et al., 2000) are also corroborated in this study.

It is interesting that *Larix* and *Pseudotsuga* form a clade sister to the other genera of Pinaceae in the phylogenetic tree of *rps*3. As mentioned earlier, no RNA editing site was found in the two sister genera whereas some editing sites occurred in the other genera of Pinaceae such as *Abies*. Sequences with and without RNA editing sites can be regarded as processed paralogs and will affect phylo-

genetic tree topology when they are analyzed together, although the editing sites are useful for phylogenetic reconstruction (Hiesel et al., 1994; Bowe and dePamphilis, 1995, 1996; Pesole et al., 1996; Lu et al., 1998; Szmidt et al., 2001). In the *rps*3 exon sequences, at least 14 T-C substitutions exist between the *Larix-Pseudotsuga* clade and the other Pinaceae genera, which should be responsible at least in part for the unexpected systematic position of *Larix* and *Pseudotsuga*. Transcripts are edited with various degrees of efficiency in the plant mitochondria (Schuster et al., 1990; Phreaner et al., 1996; Shikanai, 2006). In this study, more than four different cDNA clone sequences were detected in *A. holophylla* and *Ginkgo biloba* (data not shown), although most of RNA editing sites have been detected by the direct sequencing of cDNA. So we suggest that the number of RNA editing sites should be confirmed by more careful investigations.

Acknowledgments

The authors thank the handling editor and the two anonymous reviewers for their insightful comments and suggestions on the manuscript, Prof. Christopher Quinn from the Royal Botanic Gardens of Australia for his kind help in collecting leaf materials of some Cupressaceae species and Prof. Tao Sang from the Michigan State University for reading the manuscript. We also thank 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, 30730010 and 30870166).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.09.011.

References

Adams, K.L., Daley, D.O., Qiu, Y.-L., Whelan, J., Palmer, J.D., 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature 408, 354–357.

Adams, K.L., Rosenblueth, M., Qiu, Y.-L., Palmer, J.D., 2001. Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. Genetics 158, 1289–1300.

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol. Phylogenet. Evol. 29, 380–395.

Adams, K.L., Qiu, Y.-L., Stoutemeyer, M., Palmer, J.D., 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc. Natl. Acad. Sci. USA 99, 9905–9912.

Adams, K.L., Song, K.-M., Roessler, P.G., Nugent, J.M., Doyle, J.L., Doyle, J.J., Palmer, J.D., 1999. Intracellular transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox*2 genes in legumes. Proc. Natl. Acad. Sci. USA 96, 13863–13868.

Ahne, A., Müller-Derlich, J., Merlos-Lange, A.M., Kanbay, F., Wolf, K., Lang, B.F., 1988. Two distinct mechanisms for deletion in mitochondrial DNA of *Schizosaccharomyces pombe* mutator strains: Slipped mispairing mediated by direct repeats and erroneous intron splicing. J. Mol. Biol. 202, 725–734.

Albert, V.A., Backlund, A., Bremer, K., Chase, M.W., Manhart, J.R., Mishler, B.D., Nixon, K.C., 1994. Functional constraints and *rbc*L evidence for land plant phylogeny. Ann. Mo. Bot. Gard. 81, 534–568.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bakker, F.T., Breman, F., Merckx, V., 2006. DNA sequence evolution in fast evolving mitochondrial DNA *nad*1 exons in Geraniaceae and Plantaginaceae. Taxon 55, 887–896.

Barr, C.M., Keller, S.R., Ingvarsson, P.K., Sloan, D.B., Taylor, D.R., 2007. Variation in mutation rate and polymorphism among mitochondrial genes of *Silene vulgaris*. Mol. Biol. Evol. 24, 1783–1791.

Bergthorsson, U., Richardson, A.O., Young, G.J., Goertzen, L.R., Palmer, J.D., 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc. Natl. Acad. Sci. USA 101, 17747–17752.

Birky Jr., C.W., 2008. Uniparental inheritance of organelle genes. Curr. Biol. 18, R692–R695.

Blanchard, J.L., Schmidt, G.W., 1995. Pervasive migration of organellar DNA to the nucleus in plants. J. Mol. Biol. 41, 397–406.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.

Bowe, L.M., dePamphilis, C.W., 1995. Phylogenetic significance of RNA editing. Am. J. Bot. 82, 116.

Bowe, L.M., dePamphilis, C.W., 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. Mol. Biol. Evol. 13, 1159–1166.

Bowe, L.M., Coat, G., dePamphilis, C.W., 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc. Natl. Acad. Sci. USA 97, 4092–4097.

Bullerwell, C.E., Burger, G., Lang, B.F., 2000. A novel motif for identifying *rps*3 homologs in fungal mitochondrial genomes. Trends Biochem. Sci. 25, 363–365.

Burleigh, J.G., Mathews, S., 2004. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. Am. J. Bot. 91, 1599–1613.

Burleigh, J.G., Mathews, S., 2007. Assessing among-locus variation in the inference of seed plant phylogeny. Int. J. Plant Sci. 168, 111–124.

Cantino, P.D., Doyle, J.A., Graham, S.W., Judd, W.S., Olmstead, R.G., Soltis, D.E., Soltis, P.S., Donoghue, M.J., 2007. Towards a phylogenetic nomenclature of *Tracheophyta*. Taxon 56, 822–846.

Carmel, L., Rogozin, I.B., Wolf, Y.I., Koonin, E.V., 2007. Evolutionarily conserved genes preferentially accumulate introns. Genome Res. 17, 1045–1050.

Chaw, S.-M., Zharkikh, A., Sung, H.-M., Lau, T.-C., Li, W.-H., 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol. Biol. Evol. 14, 56–68.

Chaw, S.-M., Parkinson, C.L., Cheng, Y., Vincent, T.M., Palmer, J.D., 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc. Natl. Acad. Sci. USA 97, 4086–4091.

Chaw, S.-M., Shih, A.-C., Wang, D., Wu, Y.-W., Liu, S.-M., Chou, T.-Y., 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.

Cheng, Y., Nicolson, R.G., Tripp, K., Chaw, S.-M., 2000. Phylogeny of Taxaceae and Cephalotaxaceae genera inferred from chloroplast *mat*K gene and nuclear rDNA ITS region. Mol. Phylogenet. Evol. 14, 353–365.

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

Clifton, S.W., Minx, P., Fauron, C.M., Gibson, M., Allen, J.O., Sun, H., Thompson, M., Brad Barbazuk, W., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, R.K., Newton, K.J., 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol. 136, 3486–3503.

Dai, L., Toor, N., Olson, R., Keeping, A., Zimmerly, S., 2003. Database for mobile group II introns. Nucl. Acids Res. 31, 424–426.

Daley, D.O., Adams, K.L., Clifton, R., Qualmann, S., Millar, A.H., Palmer, J.D., Pratje, E., Whelan, J., 2002. Gene transfer from mitochondrion to nucleus: novel mechanisms for gene activation from *cox*2. Plant J. 30, 11–21.

De Benedetto, C., De Gara, L., Arrigoni, O., Albrizio, M., Gallerani, R., 1992. The structure of the cytochrome oxidase subunit II gene and its use as a new character in the construction of the phylogenetic tree of angiospermae. Plant Sci. 81, 75–82.

Dombrovska, O., Qiu, Y.-L., 2004. Distribution of introns in the mitochondrial gene *nad*1 in land plants: phylogenetic and molecular evolutionary implications. Mol. Phylogenet. Evol. 32, 246–263.

Donoghue, M.J., Doyle, J.A., 2000. Seed plant phylogeny: Demise of the anthophyte hypothesis? Curr. Biol. 10, R106–R109.

Doyle, J.A., Donoghue, M.J., 1986. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. Bot. Rev. 52, 321–431.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Doyle, J.A., 1996. Seed plant phylogeny and the relationships of Gnetales. Int. J. Plant Sci. 157, S3–S39.

Drouin, G., Daoud, H., Xia, J., 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol. Phylogenet. Evol. 49, 827–831.

Farjón, A., 2001. World checklist and bibliography of conifers, second ed. Royal Bot. Gard., Kew, England.

Farjón, A., 2005. A monograph of Cupressaceae and Sciadopitys. Royal Bot. Gard., Kew, England.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

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

Felsenstein, J., 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. 22, 521–565.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) Version 3.6a2. Distributed by the author. Seattle: Department of Genome Sciences, University of Washington. Available at http://evolution.gs.washington.edu/phylip.html.

Freyer, R., Kiefer-Meyer, M.C., Kössel, H., 1997. Occurrence of plastid RNA editing in all major lineages of land plants. Proc. Natl. Acad. Sci. USA 94, 6285–6290.

Gaut, B.S., 1997. Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht, M.K., MacIntyre, R.J., Clegg, M.T. (Eds.), Evolutionary biology, vol. 30. Plenum Press, NY, USA, pp. 93–120.

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.

Goremykin, V., Bobrova, V., Pahnke, J., Troitsky, A., Antonov, A., Martin, W., 1996. Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbc*L data do not support Gnetalean affinities of angiosperms. Mol. Biol. Evol. 13, 383–396.

Groth-Malonek, M., Pruchner, D., Grewe, F., Knoop, V., 2005. Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. Mol. Biol. Evol. 22, 117–125.

Gugerli, F., Sperisen, C., Büchler, U., Brunner, I., Brodbeck, S., Palmer, J.D., Qiu, Y.-L., 2001. The evolutionary split of Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. Mol. Phylogenet. Evol. 21, 167–175.

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

Hajibabaei, M., Xia, J., Drouin, G., 2006. Seed plant phylogeny: Gnetophytes are derived conifers and a sister group to Pinaceae. Mol. Phylogenet. Evol. 40, 208–217.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95–98.

Hamby, R.K., Zimmer, E.A., 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis, P.S., Soltis, D.E., Doyle, J.J. (Eds.), Molecular Systematics of Plants. Chapman and Hall, NY, USA, pp. 50–91.

Handa, H., 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acids Res. 31, 5907–5916.

Hansen, A., Hansmann, S., Samigullin, T., Antonov, A., Martin, W., 1999. *Gnetum* and the angiosperms: molecular evidence that their shared morphological characters are convergent, rather than homologous. Mol. Biol. Evol. 16, 1006–1009.

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Hiesel, R., Combettes, B., Brennicke, A., 1994. Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. Proc. Natl. Acad. Sci. USA 91, 629–633.

Hirao, T., Watanabe, A., Kurita, M., Kondo, T., Takata, K., 2008. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 8, 70.

Houliston, G.J., Olson, M.S., 2006. Nonneutral evolution of organelle genes in *Silene vulgaris*. Genetics 174, 1983–1994.

Huelsenbeck, J.P., 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44, 17–48.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Jeffares, D.C., Penkett, C.J., Bähler, J., 2008. Rapidly regulated genes are intron poor. Trends Genet. 24, 375–378.

Joly, S., Brouillet, L., Bruneau, A., 2001. Phylogenetic implications of the multiple losses of the mitochondrial *coxII.i3* intron in the angiosperms. Int. J. Plant Sci. 162, 359–373.

Kimura, T., Ohana, T., Mimoto, K., 1988. Discovery of a Podocarpaceous plant from the lower Cretaceous of Kochi prefecture, in the outer zone of southwest Japan. Proc. Japan Acad. B 64, 213–216.

Kirby, D.A., Muse, S.V., Stephan, W., 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA 92, 9047–9051.

Knoop, V., 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr. Genet. 46, 123–139.

Krassilov, V.A., 1974. *Podocarpus* from the upper Cretaceous of eastern Asia and its bearing on the theory of conifer evolution. Palaeontology 17, 365–370.

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

Kumar, S., Dudley, J., Nei, M., Tamura, K., 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9, 299–306.

Lambowitz, A.M., Zimmerly, S., 2004. Mobile group II introns. Annu. Rev. Genet. 38, 1–35.

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

Laroche, J., Bousquet, J., 1999. Evolution of the mitochondrial *rps3* intron in perennial and annual angiosperms and homology to *nad*5 intron 1. Mol. Biol. Evol. 16, 441–452.

Li, J.-Y., Wen, Y.-D., Jiang, Ch.-Y., Li, F., Gao, C.-Ch., 2000. A simple and rapid method for isolating mitochondrial plasmid DNA from plant. Acta Sci. Nat. Univ. Nankai. 33, 49–52.

Lu, M.-Z., Szmidt, A.E., Wang, X.-R., 1998. RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *cox*I gene. Plant Mol. Biol. 37, 225–234.

Lynch, M., Richardson, A.O., 2002. The evolution of spliceosomal introns. Curr. Opin. Genet. Dev. 12, 701–710.

Ma, Q.-W., Li, F.-L., Li, Ch.-S., 2005. The coast redwoods (*Sequoia*, Taxodiaceae) from the Eocene of Heilongjiang and the Miocene of Yunnan. China. Rev. Palaeobot. Palyno. 135, 117–129.

Magallón, S., Sanderson, M.J., 2002. Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages. Am. J. Bot. 89, 1991–2006.

Magallón, S., Sanderson, M.J., 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. Evolution 59, 1653–1670.

Malek, O., Brennicke, A., Knoop, V., 1997. Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. Proc. Natl. Acad. Sci. USA 94, 553–558.

Marais, G., Nouvellet, P., Keightley, P.D., Charlesworth, B., 2005. Intron size and exon evolution in *Drosophila*. Genetics 170, 481–485.

McCauley, D.E., Bailey, M.F., Sherman, N.A., Damell, M.Z., 2005. Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. Heredity 95, 50–58.

McCauley, D.E., Ellis, J.R., 2008. Recombination and linkage disequilibrium among mitochondrial genes in structured populations of the gynodioecious plant *Silene vulgaris*. Evolution 62, 823–832.

McCoy, S.R., Kuehl, J.V., Boore, J.L., Raubeson, L.A., 2008. The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. BMC Evol. Biol. 8, 130.

McIver, E.E., 2001. Cretaceous *Widdringtonia* Endl. (Cupressaceae) from North America. Int. J. Plant Sci. 162, 937–961.

Mizushima, S., Nomura, M., 1970. Assembly mapping of 30S ribosomal proteins from *E. coli*. Nature 226, 1214–1218.

Mogensen, H.L., 1996. The hows and whys of cytoplasmic inheritance in seed plants. Am. J. Bot. 83, 383–404.

Mower, J.P., Touzet, P., Gummow, J.S., Delph, L.F., Palmer, J.D., 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol. Biol. 7, 135.

Mundry, M., Stützel, T., 2004. Morphogenesis of the reproductive shoots of *Welwitschia mirabilis* and *Ephedra distachya* (Gnetales), and its evolutionary implications. Org. Divers. Evol. 4, 91–108.

Muse, S.V., 2000. Examining rates and patterns of nucleotide substitution in plants. Plant Mol. Biol. 42, 25–43.

Neu, R., Goffart, S., Wolf, K., Schäfer, B., 1998. Relocation of *urf* a from the mitochondrion to the nucleus cures the mitochondrial mutator phenotype in the fission yeast *Schizosaccharomyces pombe*. Mol. Gen. Genet. 258, 389–396.

Nicholson, R., 1990. Chasing ghosts. Nat. Hist. 12, 10–13.

Niu, D.-K., Hou, W.-R., Li, S.-W., 2005. mRNA-mediated intron losses: evidence from extraordinarily large exons. Mol. Biol. Evol. 22, 1475–1481.

Nugent, J.M., Palmer, J.D., 1991. RNA-mediated transfer of the gene *cox*II from the mitochondrion to the nucleus during flowering plant evolution. Cell 66, 473–481.

Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala Univ., Uppsala, Sweden.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., Ohyama, K., 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. J. Mol. Biol. 223, 1–7.

Ogihara, Y., Yamazaki, Y., Murai, K., Akira, K., Toru, T., Takashi, S., Naohiko, M., Shuhei, N., Chiharu, N., Naoki, M., Shigeo, T., Minoru, M., Satoshi, F., Koichiro, T., 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 33, 6235–6250.

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.-L., Song, K.-M., 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc. Natl. Acad. Sci. USA 97, 6960–6966.

Palmer, J.D., Herbon, L.A., 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J. Mol. Evol. 28, 87–97.

Parkinson, C.L., Mower, J.P., Qiu, Y.-L., Shirk, A.J., Song, K.-M., Young, N.D., DePamphilis, C.W., Palmer, J.D., 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol. Biol. 5, 73.

Pesole, G., Ceci, L.R., Gissi, C., Saccone, C., Quagliariello, C., 1996. Evolution of the *nad* 3-*rps12* gene cluster in angiosperm mitochondria: comparison of edited and unedited sequences. J. Mol. Evol. 43, 447–452.

Phreaner, C.G., Williams, M.A., Mulligan, R.M., 1996. Incomplete editing of *rps*12 transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. Plant Cell 8, 107–117.

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

Pyle, A.M., Lambowitz, A.M., 2006. The RNA World third ed.. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), Cold Spring Harbor monograph series, vol. 43. Cold Spring Harbor Laboratory Press, NY, USA, pp. 469–505.

Qiu, Y.-L., Cho, Y., Cox, J.C., Palmer, J.D., 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394, 671–674.

Rabbi, M.F., Wilson, K.G., 1993. The mitochondrial *cox*II intron has been lost in two different lineages of dicots and altered in others. Am. J. Bot. 80, 1216–1223.

Rai, H.S., O'Brien, H.E., Reeves, P.A., Olmstead, R.G., Graham, S.W., 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. Mol. Phylogenet. Evol. 29, 350–359.

Rai, H.S., Reeves, P.A., Peakal, R., Olmstead, R.G., Graham, S.W., 2008. Inference of higher-order conifer relationships from a multi-locus plastid data set. Botany 86, 658–669.

Ramakrishnan, V., Graziano, V., Capel, M.S., 1986. A role for proteins S3 and S14 in the 30S ribosomal subunit. J. Biol. Chem. 261, 15049–15052.

Raubeson, L.A., Jansen, R.K., 1992. A rare chloroplast-DNA structural mutation is shared by all conifers. Biochem. Syst. Ecol. 20, 17–24.

Regina, T.M.R., Picardi, E., Lopez, L., Pesole, G., Quagliariello, C., 2005. A novel additional group II intron distinguishes the mitochondrial *rps*3 gene in gymnosperms. J. Mol. Evol. 60, 196–206.

Robart, A.R., Seo, W., Zimmerly, S., 2007. Insertion of group II intron retroelements after intrinsic transcriptional terminators. Proc. Natl. Acad. Sci. USA 104, 6620–6625.

Rodríguez-Trelles, F., Tarrío, R., Ayala, F.J., 2006. Origins and evolution of spliceosomal introns. Annu. Rev. Genet. 40, 47–76.

Rogers, S.O., Bendich, A.J., 1988. Extraction of DNA from plant tissues. Plant Mol. Biol. (Manual). A6, 1–10.

Rokas, A., Holland, P.W.H., 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15, 454–459.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rydin, C., Källersjö, M., 2002. Taxon sampling and seed plant phylogeny. Cladistics 18, 485–513.

Samigullin, T.K., Martin, W.F., Troitsky, A.V., Antonov, A.S., 1999. Molecular data from the chloroplast *rpo*C1 gene suggest a deep and distinct dichotomy of contemporary spermatophytes into two monophyla: gymnosperms (including Gnetales) and angiosperms. J. Mol. Evol. 49, 310–315.

Sanderson, M.J., 2003. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.

Savard, L., Li, P., Strauss, S.H., Chase, M.W., Michaud, M., Bousquet, J., 1994. Chloroplast and nuclear gene sequences indicated late Pennsylvanian time for the last common ancestor of extant seed plants. Proc. Natl. Acad. Sci. USA. 91, 5163–5167.

Schmidt, M., Schneider-Poetsch, H.A., 2002. The evolution of gymnosperms redrawn by phytochrome genes: the Gnetatae appear at the base of the gymnosperms. J. Mol. Evol. 54, 715–724.

Schuster, W., Wissinger, B., Unseld, M., Brennicke, A., 1990. Transcripts of the NADH-dehydrogenase subunit 3 gene are differentially edited in *Oenothera* mitochondria. EMBO J. 9, 263–269.

Shikanai, T., 2006. RNA editing in plant organelles: machinery, physiological function and evolution. Cell. Mol. Life Sci. 63, 698–708.

Sloan, D.B., Barr, C.M., Olson, M.S., Keller, S.R., Taylor, D.R., 2008. Evolutionary rate variation at multiple levels of biological organization in plant mitochondrial DNA. Mol. Biol. Evol. 25, 243–246.

Soltis, P.S., Soltis, D.E., Wolf, P.G., Nickrent, D.L., Chaw, S.-M., Chapman, R.L., 1999. The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? Mol. Biol. Evol. 16, 1774–1784.

Städler, T., Delph, L.F., 2002. Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. Proc. Natl. Acad. Sci. USA 99, 11730–11735.

Stephan, W., Kirby, D.A., 1993. RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. Genetics 135, 97–103.

Swofford, D.L., 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, MA, England.

Szmidt, A.E., Lu, M.-Zh., Wang, X.-R., 2001. Effects of RNA editing on the *cox*I evolution and phylogeny reconstruction. Euphytica 118, 9–18.

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599–607.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0.. Mol. Biol. Evol. 24, 1596–1599.

Taylor, T.N., Taylor, E.L., 1993. The Biology and Evolution of Fossil Plants. Prentice Hall, NJ, USA.

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.

Toor, N., Keating, K.S., Taylor, S.D., Pyle, A.M., 2008. Crystal structure of a self-spliced group II intron. Science 320, 77–82.

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

Turmel, M., Otis, C., Lemieux, C., 2003. The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. Plant Cell 15, 1888–1903.

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

Wang, D., Wu, Y.-W., Shih, A.C., Wu, C.-S., Wang, Y.-N., Chaw, S.-M., 2007. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. Mol. Biol. Evol. 24, 2040–2048.

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.

Wang, Y., Zeng, F., Hon, C.C., Zhang, Y., Leung, F.C., 2008. The mitochondrial genome of the Basidiomycete fungus *Pleurotus ostreatus* (oyster mushroom). FEMS Microbiol. Lett. 280, 34–41.

Wang, Z., Zhang, J.-Z., 2009. Why is the correlation between gene importance and gene evolutionary rate so weak? PloS Genet. 5, e1000329.

Whittle, C.A., Johnston, M.O., 2002. Male-driven evolution of mitochondrial and chloroplastidial DNA sequences in plants. Mol. Biol. Evol. 19, 938–949.

Winter, K.U., Becker, A., Münster, T., Kim, J.T., Saedler, H., Theissen, G., 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc. Natl. Acad. Sci. USA 96, 7342–7347.

Wolfe, K.H., Li, W.-H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.

Yang, Z.-H., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.