Molecular Phylogenetics and Evolution 55 (2010) 168–177

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: Incomplete concerted evolution and the origin of pseudogenes

Long-Qian Xiao [a,b], Michael Möller [c], Hua Zhu [a,*]

[a] Xishuangbanna Tropical Botanical Garden, The Chinese Academy of Sciences, Kunming 650223, China
[b] Graduate University of the Chinese Academy of Sciences, Beijing 100049, China
[c] Royal Botanical Garden Edinburgh, Edinburg EH3 5LR, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 30 June 2009
Revised 19 November 2009
Accepted 20 November 2009
Available online 27 November 2009

Keywords:
Internal transcribed spacer (ITS)
Concerted evolution
Cycas
Pseudogenes
Phylogeny

ABSTRACT

Molecular studies of six species from the ancient extant seed plant *Cycas*, covering a wide range of its morphological diversity and all major areas of distribution, revealed a high level of intra-individual polymorphism of the internal transcribed spacer (ITS1, 5.8S, and ITS2) region, indicative of incomplete nrDNA concerted evolution. Through a range of comparisons of sequence characteristics to functional cDNA ITS copies, including sequence length and substitution variation, GC content, secondary structure stability, the presence of a conserved motif in the 5.8S gene, and evolutionary rates, the PCR amplified divergent genomic DNA ITS paralogs were identified as either putative pseudogenes, recombinants or functional paralogs. This incomplete ITS concerted evolution may be linked to the high number of nucleolar organizer regions in the *Cycas* genome, and the incomplete lineage sorting due to recent species divergence in the genus. Based on the distribution of a 14 bp deletion, an early evolutionary origin of the pseudogenes is indicated, possibly predating the diversification of *Cycas*. Due to their early origin combined with the unconstraint evolution of the ITS region in pseudogenes, they accumulate high levels of homoplastic mutations. This leads to random relationships among the pseudogenes due to long-branch attractions, whereas the phylogenetic relationships inferred from the functional ITS paralogs grouped the sequences in species specific clades (except for *C. circinalis* and *C. rumphii*). The findings of our extensive study will have a wide significance, for the evolution of these molecular sequences, and their utilization as a major marker for reconstructing phylogenies.

1. Introduction

The nuclear ribosomal-DNA (nrDNA) cistron consists of 18S, ITS1, 5.8S, ITS2, and 26S in all land plants except bryophytes (Sone et al., 1999), and is tandemly repeated in several hundred to thousand copies. Each paralog is separated by an inter-genic spacer (IGS) and the tandem repeats located at one or more loci on one or several non-homologous chromosomes in a genome. One of the remarkable properties of nrDNA (including ITS) genes is that their paralogs within individuals are quite homogenous, resulting from concerted evolution. The underlying molecular processes are presumed to involve unequal crossing over (Smith, 1976) and gene conversion (Nagylaki, 1984). nrDNA paralogs will, however, display polymorphisms in individuals where concerted evolution is incomplete, for example in cases where hybridization is involved (Muir et al., 2001), or where concerted evolution cannot act between paralogs effectively when they are dispersed on non-homologous chromosomes in the genome (Wei and Wang, 2004).

In recent years, multiple divergent ITS paralogs within individuals have been observed in several plant groups (e.g. Harpke and Peterson, 2006; Grimm and Denk, 2007; Ochieng et al., 2007; Zheng et al., 2008), which suggest incomplete concerted evolution across the repeats. In particular, studies in non-flowering plants, though limited to Coniferales (Pinaceae, e.g. Wei et al., 2003; Campbell et al., 2005; Kan et al., 2007) and Gnetales (*Gnetum*, Won and Renner, 2005), showed that intra-genomic ITS paralogs are often considerably divergent. These studies also indicated different underlying evolutionary drives that may explain observed differences in the polymorphism patterns of ITS paralogs between Coniferales and Gnetales. In both lineages, ITS paralogs displayed extensive length variations, especially in the ITS1 region. In Pinaceae at least, non-homologous recombination and/or unequal crossing over between tandem repeats were proposed as the main factors responsible for these length variations (Kan et al., 2007). A different mechanism may explain the length variations in *Gnetum*, because only a few tandem repeats were detected and all were limited to individual species (Won and Renner, 2005).

Divergent paralogs can also result from recent or historic inter- and/or intra-genomic duplication events (e.g. Möller et al., 2008),

* Corresponding author. Fax: +86 871 5160916.
E-mail address: zhuh@xtbg.ac.cn (H. Zhu).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.11.020

and phylogenetic trees reconstructed from such paralogs may reflect the history of such gene duplication events, as well as speciation events (Nei, 1987; Sanderson and Doyle, 1992), though not necessarily (Denduangboripant and Cronk, 2000; Möller et al., 2008). Among divergent rDNA paralogs, non-functional pseudogenes are prominent, and many studies have demonstrated the existence of pseudogenes in plant genomes, where concerted evolution of nrDNA is incomplete. Pseudogenes are characterized by a higher relative substitution rate, an increased AT content, and lower secondary structure stability (reviewed by Álvarez and Wendel, 2003).

An important and often debated issue is the effect of ITS pseudogenes on phylogenetic inferences. On the one hand, pseudogenes, assumed to have escaped from functional constraints, have accumulated many mutations and can cluster randomly across phylogenetic trees due to long-branch attraction (LBA), which confounds attempts to recover correct phylogenetic species relationships (e.g. Kita and Ito, 2000; Mayol and Rosselló, 2001). On the other hand, ITS pseudogenes can potentially be useful for phylogenetic analyses of closely related species, when the functional paralogs provide too low variation (e.g. Ochieng et al., 2007). In non-flowering plants, such as *Larix* and *Gnetum*, a few highly divergent ITS paralogs have been identified that degenerated to pseudogenes, but still clustered with conspecific functional paralogs in gene trees (Wei et al., 2003; Won and Renner, 2005). These may have a recent origin and have little detrimental impact on phylogenetic analyses.

It is well known that extant gymnosperms include four morphologically highly divergent orders, i.e. Coniferales, Cycadales, Ginkgoales, and Gnetales. Although the evolutionary relationships among these are debated, cycads are likely the earliest diverged gymnosperm lineage (e.g. Chaw et al., 2000), because their mature pollen has multi-ciliate sperms and their ovules are borne on the margins of leaf-like megasporophylls (Stevenson, 1990). Among the cycads, the genus *Cycas* is the most widespread and diverse lineage, with a geographic range from Africa eastwards to the Pacific islands, and from China and southern Japan southwards to Australia. It forms the sister lineage to all other extant cycads (Treutlein and Wink, 2004; Chaw et al., 2005), and is possibly the most ancient extant seed plant.

In view of the differences in the evolution of nrDNA paralogs between Coniferales and Gnetales, and the few clones per species examined in the latter (2–6 clones per species, Won and Renner, 2005), a more comprehensive sampling within individuals is desirable to obtain a more accurate estimate of the level of ITS paralog polymorphism. This will also lead to a better understanding of the dynamics of ITS evolution in non-flowering seed plants.

In the present study, we thus obtained a total of 103 clones of nrDNA ITS regions of six plants representing six species of the Cycadales genus *Cycas*. These included also functional cDNA copies isolated from transcribed RNA. The specific aims of the study were as follows: (1) to document the level and pattern of intra- and inter-species nrDNA variation, (2) to demonstrate the existence of nrDNA pseudogenes, released from concerted evolution, (3) to estimate their effect on phylogenetic analyses, and (4) to explore the dynamics of evolution of nrDNA copies in cycads.

Until now, detailed studies on a larger sample of intra-individual nrDNA copies were rarely undertaken, and this is the first such study in Cycadales. However, our findings will be of much wider significance, for the evolution of these molecular sequences, and their utilization as a major marker for inferring phylogenetic relationships.

2. Materials and methods

2.1. Plant materials

In *Cycas*, the limited morphological evolution resulted in taxonomic confusion (Hill, 1994; Yang and Meerow, 1996), and several alternative classifications have been proposed. For example, Wang (2000) divided the genus into four subgenera according to their seed types, while Hill (2004) defined six sections in *Cycas* almost entirely depending on reproductive characters. In the present study, six (one individual per species) out of the about 90 described species in the genus were examined, covering a wide range of the morphological variation, and all the major areas of distribution of *Cycas* (Table 1).

2.2. Nucleic acid isolation, PCR and RT-PCR, cloning, and sequencing

Total DNA was extracted from silica gel dried leaves following the CTAB protocol (Doyle, 1991). One plant of *C. revoluta* was selected as reference for the extraction of putative functional ITS copies, and RNA isolation was carried out using a TRIZOL Kit (Invitrogen, Carlsbad, CA), following the manufacture's protocol. The total RNA extract was treated with DNase to exclude DNA contamination, before first-strand cDNA synthesis was carried out using PrimeScript RTase (TaKaRa, Dalian, China) according to the manufacturer's protocol.

The entire ITS region, comprising ITS1, 5.8S, and ITS2, was amplified from genomic DNA from all species with primers ITS5* (5′-GGAAGGAGAAGTCGTAACAAGG-3′) (Liston et al., 1996), and 26S–25R (5′-TATGCTTAAACTCAGCGGGT-3′) (Nickrent et al., 1994). PCR was carried out in 25 μl volumes, 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). PCRs were performed with a GeneAmp PCR System 9700 (Perkin–Elmer, Waltham, USA), using a profile of 4 min at 96 °C, followed by 30 cycles of 30 s at 96 °C, 45 s at 55 °C, and 1 min 20 s at 72 °C, with a final extension step for 7 min at 72 °C. RT-PCR on cDNA of *C. revoluta* was performed using the PrimeScript™ RT-PCR Kit and PCR conditions as described above.

PCR and RT-PCR products were separated by electrophoresis in 1.0% agarose gels, and the one bright band present was cut out for purification using a Gel Band Purification Kit (Amersham Pharma-

Table 1
Information on affiliation, voucher number, and origin of six species of *Cycas* analysed.

Species	Affiliation		Voucher number	Origin
	System (Wang, 2000)	System (Hill, 2004)		
C. revoluta	Subgen. Panzhihuaenses	Sect. Asiorientalis	Xiao, 06010	Ryukyu Islands in Japan
C. debaoensis	Subgen. Panzhihuaenses	Sect. Stangerioides	Xiao, 06014	Western Guangxi and southwest Yunnan in China
C. siamensis	Subgen. Cycas	Sect. Indosinensis	Xiao, 06021	Central Thailand, and the central plateau region in Vietnam
C. rumphii	Subgen. Cycas	Sect. Cycas	Xiao, 06023	Moluccan islands in Indonesia
C. circinalis	Subgen. Truncata	Sect. Cycas	Xiao, 06024	Western Ghats in Indian
C. platyphylla	Subgen. Media	Sect. Cycas	Xiao, 06035	Petford district in Australia

Voucher specimens are deposited in Herbarium of Xishuangbanna Tropical Botanical Garden, CAS.

cia Biotech, Inc., Piscataway, USA). The fragments were then cloned with the pGEM–T Easy Vector System I (Promega,Madison, USA). Fifteen to 26 genomic clones were obtained from each individual, and additionally six cDNA clones from *C. revoluta*. All clones were sequenced with the two PCR primers, using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit on an automatic DNA Sequencer (ABI PRISM 377, Perkin Elmer, Foster City, CA), through the Shanghai Sangon Biological Engineering Technology & Services Co., Ltd.

2.3. Data analyses

Boundaries of the ITS1, 5.8S, and ITS2 regions were identified in comparison to sequences from Chaw et al. (2005) (GenBank accession numbers FJ907968–FJ908070). Sequences were aligned initially with CLUSTAL X (Thompson et al., 1997). BioEdit v 5.0.6. (Hall, 1999) was used to refine the alignments manually and to determine the lengths and GC contents of ITS1, 5.8S, and ITS2 separately. The Tandem Repeats Finder (Benson, 1999) was used to detect repeats in the ITS sequences. Differences in substitution rate can discriminate functional from pseudogenes (Buckler and Holtsford, 1996a,b). The distribution and pattern of nucleotide substitution in all sequences was investigated with the cDNA clone *C. revoluta*_C2 as the reference sequence, using HYPERMUT (Rose and Korber, 2000). This program was originally designed to study the sequence evolution of HIV, showing excessive levels of G ⇒ A mutations. It assumes that all differences arose from a single substitution, and all substitutions observed in each sequence, compared to the reference sequence, are summarized and their physical locations along the sequences graphically illustrated.

The 5.8S secondary structure is very similar, even between plants and animals, and thus, its sequences highly conserved (Suh et al., 1992). In this study, the presence of a 14-bp motif in 5.8S, 5′-GAATTGCAGAATCC-3′, which is seed–plant specific (Jobes and Thien, 1997; Won and Renner, 2005), was located in the sequences of *Cycas* obtained here. Minimum free energy (ΔG at 37 °C) of the 5.8S secondary structure of the clones was estimated with mFold (Zuker, 2003). Functional copies posses lower free energy levels and thus higher stability, compared to pseudogenes (Buckler and Holtsford, 1996a,b).

MEGA 4 (Kumar et al., 1994) was used to estimate pairwise distances under the Kimura's (1980) two-parameter model, and DnaSP (Rozas and Rozas, 1999) to calculate the average number of nucleotide difference (K), the nucleotide diversity (π), and Tajima's D test for presumed functional and putative pseudogenes (excluding *C. platyphylla*_3 and *C. platyphylla*_16, both of which had a >100-bp deletion). GENECONV tests were performed using the substitution model for putative recombinant detection, as implemented in the Recombination Detection Program package3b27 (Padidam et al., 1999).

PAUP* 4.0 (Swofford, 2002) was used to perform phylogenetic analyses of the whole ITS region (ITS1, 5.8S, ITS2), implementing Neighbor-Joining (NJ) under the HKY85 model with 1000 bootstrap replicates. A maximum parsimony (MP) analysis was carried out with all characters equally weighted and treated as unordered, and a heuristic search implemented with gaps treated as missing, 1000 random addition sequence replicates, TBR branch swapping and the MULTREES option on. Though Zamiaceae is the sister clade of *Cycas* (Treutlein and Wink, 2004; Chaw et al., 2005), their ITS sequences cannot be aligned unambiguously, due to too high sequence divergences and substantial length variations. Thus, in the present study no outgroups were included, and the tree displayed unrooted. Clade support was assessed through bootstrap analyses with 1000 replicates using heuristic search options as described above.

3. Results

Five pairs of ITS sequences (including one pair of cDNA copies) isolated from an individual were identical, and only one of each pair remained in the following analyses. In total 98 distinct genomic ITS paralogs were obtained from the six species and five functional cDNA ITS paralogs obtained through RT-PCR from RNA of *C. revoluta*.

3.1. Length and repeats of the ITS region in Cycas

The entire ITS region varied in length from 735 (*C. platyphylla*_16) to 1091 bp (*C. platyphylla*_20), and their alignment was 1133 bp long. As in other gymnosperms, the 5.8S gene and the ITS2 region were relatively length conserved, their lengths ranging from 136 to 161 bp and 230 to 248 bp, respectively (Appendix). Exceptions were sequence *C. circinalis*_16, which had a 61-bp deletion in 5.8S, and *C. platyphylla*_16 with an ITS2 length of 208 bp. In contrast, ITS1 was extremely length variable with a range from 366 to 684 bp. *C. platyphylla*_16, and *C. platyphylla*_3 had, with 366 and 419 bp, respectively, markedly short ITS1 regions. The functional *C. revoluta* cDNA ITS paralogs had a relatively uniform length of 677 to 681, 246, and 161 bp for ITS1, ITS2, and 5.8S, respectively. The only exception was clone *C. revoluta*_C5 which had a 16 bp deletion in ITS2 (Appendix). The average lengths of the ITS regions did not vary greatly among the species, or compared to the functional reference ITS paralogs of *C. revoluta* (Table 3).

Using the Tandem Repeats Finder, no repeat was detected using the default search options (alignment parameters 2, 7, 7, and minimum alignment score 50). Six putative repeats (Table 2) were detected among the ITS sequences under the most relaxed search options (alignment parameters 2, 3, 5, and minimum alignment score 30). The putative repeats were 2–23 bp in length, had 2.1–3.0 copies, and 68–100 percent matches. Among these, three repeats were located in ITS1, two in ITS2, and one in 5.8S.

3.2. GC content of the ITS region

Distinct clusters of clones emerged when plotting the GC content of the ITS spacers against each other (Fig. 1A). In most cases the GC content correlated well between the two ITS spacers. The functional cDNA ITS paralogs of *C. revoluta* possessed high GC values in the spacers (64.1% in ITS1, 65% in ITS2), and a markedly lower value for 5.8S (55.8%) (Table 3).

Some paralogs from each species clustered tightly around the functional *C. revoluta* paralogs, indicating they share very similar GC values. The remaining paralogs scattered separately in a loose cluster, indicating that their GC values varied greatly (Fig. 1A). This allowed the separation into two distinct classes of sequences, designated as 'presumed functional' when clustering with the functional cDNA paralogs of *C. revoluta*, or as 'putative pseudogene' when clustering scattered (Table 3).

The average GC content of the tight cluster was 64.4%, 65.2%, and 55.4% in ITS1, ITS2, and 5.8S, respectively. The remaining clones had an average GC values of 52.2%, 51.6%, and 44.8% for ITS1, ITS2, and 5.8S, respectively, about 10% to 13% lower than functional paralogs, and with high standard deviations (Table 3).

3.3. Secondary structure minimum free energy of 5.8S rDNA

The minimum free energy (ΔG at 37 °C) of the 5.8S secondary structures ranged from –19.4 to –6.5 (excluding *C. circinalis*_16, which showed a large deletion in 5.8S) (Table 3). The functional cDNA paralogs of *C. revoluta* had a low average value of -16.4. Plotting the minimum free energy values *versus* the GC content of 5.8S,

Table 2
Tandem repeats found between ITS sequences in *Cycas*.

Sequence label	Consensus pattern	Consensus size	Copy number	Percent matches	Location
*C. revoluta*_C1, C2, C4, 1, 4, 5, 9, 12	CTCTACACGG	10	2.2	83	ITS1
*C. debaoensis*_2, 3, 6	GGGCAGG	7	2.1	100	ITS1
*C. siamensis*_2, 3, 6, 7, 10, 12	GCGGGCAGCGTTGTCTCGGCGAG	23	2.6	68	ITS1
*C. rumphii*_3, 11, 13, 8, 2, 4, 5, 12, 14, 17, 10, 6; *C. circinalis*_14, 3, 12, 15, 2, 9, 1, 10, 13, 8	TCGGTCGT	8	3.0	75	ITS2
*C. platyphylla*_2, 4, 19, 12, 7, 9, 11, 13, 28, 5	CGGAAATGCACGACCCTCGCCAA	23	2.0	72	5.8S
	TCGGTCGT	8	3.0	75	ITS2

resulted in a similar distribution of the paralogs compared to the GC values in the spacers (Fig. 1A and B). The average value for presumed functional paralogs was significantly lower than that for the putative pseudogenes (−16.8 kcal/mol with $\sigma = 1.1$ for functional copies vs −11 kcal/mol with $\sigma = 2.2$ for pseudogenes). As with the GC content the presumed functional copies showed far less variation (Table 3, Fig. 1B).

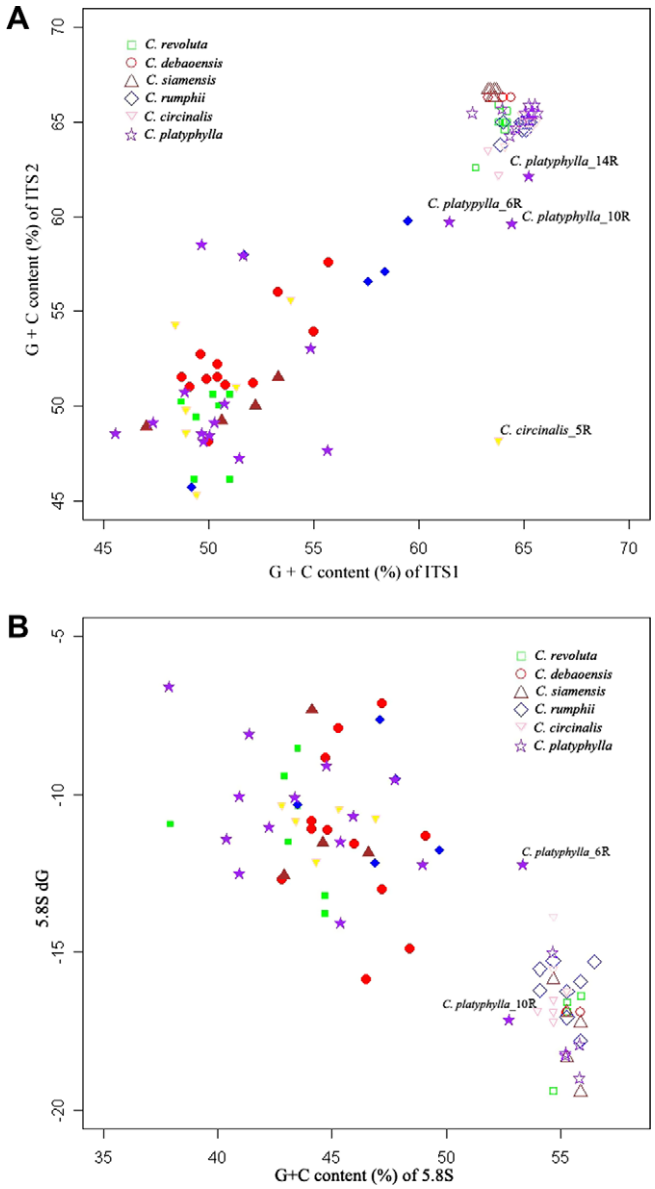

Fig. 1. Plot of the G + C content of all paralogs obtained from six *Cycas* species. (A). ITS1 *versus* ITS2. (B). 5.8S *versus* the minimum free energy (ΔG at 37 °C) of the secondary structure of 5.8S, excluding *C. circinalis*_16 with a 61-bp deletion in 5.8S. Shaded symbols indicate putative pseudogenes.

3.4. Indel distribution in the ITS region

The alignment gap sizes in the ITS regions across all clones ranged markedly from 1 bp to 327 bp. Among these gaps, all indels were unique to only one paralog or several species specific sequences, except for a 14-bp indel, located at position 6–20 near the ITS1 5′-end. This was shared by all presumed functional ITS paralogs of *C. siamensis*, *C. rumphii*, *C. circinalis*, and *C. platyphylla* and the four recombinants (*C. platyphylla*_6R, 10R, and 22R, and *C. rumphii*_7R, see below). This indel divided the functional copies into two types (Fig. 2A and B).

4. 14 bp conserved motif in 5.8S rDNA

The 14 bp motif in 5.8S, conserved across seed plants, was detected in all presumed functional genomic and functional *C. revoluta* cDNA ITS paralogs. The only exception was *C. debaoensis*_6, where the motif showed one A ⇒ G substitution at position 11 (to 5′-GAATTGCAGAGTCC-3′). The motif was further found in nine putative pseudogenes (Appendix).

4.1. Nucleotide substitution patterns

The nucleotide variation profiles show the physical location of nucleotide substitutions in all functional cDNA and presumed functional genomic paralogs (Fig. 2A), and all putative pseudogenes (Fig. 2B), respectively. All putative pseudogenes accumulated visibly more substitutions, which were randomly scattered throughout the ITS region, when compared to the functional cDNA ITS paralog *C. revoluta*_2.

4.2. Sequence diversity and neutrality test

The pairwise distance within species ranged from 0.1% to 1.5% for the functional cDNA paralogs of *C. revoluta*, 0.2–4.6% within the presumed functional paralogs of the species, and 8.1–28.3% across the putative pseudogenes within the species (Table 3). Overall the pairwise distances between pseudogenes were very high with an average of 20.6%.

Similarly, the sequence diversity across the ITS region of all putative pseudogenes, estimated either by the number of nucleotide differences (K) or by nucleotide diversity (π), was remarkably higher than for functional cDNA paralogs and presumed functional sequences. Though, some differences between presumed functional sequences and functional cDNA paralogs were also observed. This was mainly caused by their different sample sizes (e.g. 5 vs 49) and different numbers of species (1 vs 6) (Table 4). For both parameters, the number of nucleotide differences (K) and nucleotide diversity (π) across the entire ITS region, putative pseudogenes had values more than four and six times higher, compared to presumed functional sequences.

Evolutionary neutrality was tested using Tajima's D test on the total number of segregating sites. Although the values of the cDNA paralogs did not show significant deviations because of their extremely small sample size, the presumed functional sequences, and

Table 3
Average lengths and GC content of ITS regions and free energy of 5.8S secondary structures of 103 paralogs of six species of *Cycas* analysed. SD – standard deviation.

Taxon	Type	Number	Length bp(SD)				GC%(SD)			5.8S ΔG (SD) (kcal/mol)	Pairwise distance (%) min.–max. (mean)
			ITS1	ITS2	5.8S	Total	ITS1	ITS2	5.8S		
C. revoluta	C	5	678.6(1.4)	242.8(8.0)	161.0(0.0)	1082.4(6.7)	64.1(0.1)	65.0(0.3)	55.8(0.3)	−16.4(0.1)	0.1–1.5(0.7)
	F	7	676.6(1.2)	246.0(0.0)	161.0(0.0)	1083.6(1.2)	63.8(0.5)	64.8(1.1)	55.6(0.5)	−16.9(1.2)	0.2–2.5(1.1)
	P	7	661.1(30.2)	242.9(0.4)	160.9(0.4)	1064.9(30.5)	50.0(1.0)	49.0(2.2)	42.9(2.5)	−11.1(1.9)	21.8–26.5(23.9)
C. debaoensis	F	3	674.7(0.6)	240.0(0.0)	161.0(0.0)	1075.7(0.6)	63.9(0.6)	66.3(0.0)	55.5(0.3)	−16.9(0.0)	0.8–2.3 (1.7)
	P	12	666.5(17.6)	243.0(0.0)	160.3(2.6)	1069.8(17.5)	51.3(2.3)	52.4(2.5)	45.9(1.9)	−11.4(2.6)	8.1–26.7 (21.2)
C. siamensis	F	8	659.9(0.4)	246.0(0.0)	161.0(0.0)	1066.9(0.4)	63.5(0.1)	66.6(0.2)	55.5(0.4)	−17.3(1.2)	0.2–1.2(0.7)
	P	4	674.8(3.0)	242.5(0.6)	161.0(0.0)	1078.3(3.1)	50.8(2.8)	49.9(1.2)	44.6(1.5)	−10.8(2.4)	12.6–24.5(21.6)
C. rumphii	F	12	662.3(8.0)	246.0(0.0)	161.0(0.0)	1069.3(8.0)	64.8(0.5)	64.8(0.4)	55.3(0.7)	−16.2(0.7)	0.2–3.0 (1.2)
	P	5	670.5(8.0)	243.0(0.7)	155.8(11.1)	1070.2(10.0)	55.3(4.5)	55.4(5.6)	47.0(2.2)	−10.3(1.8)	15.5–27.0 (21.1)
C. circinalis	F	11	662.8(0.6)	246.5(1.3)	161.0(0.0)	1070.4(1.6)	64.8(0.7)	64.5(0.9)	54.9(0.4)	−16.2(0.9)	0.2–4.6 (1.6)
	P	6	667.8(14.4)	243.0(0.0)	150.2(24.6)	1061.0(38.9)	52.7(5.8)	49.8(3.4)	44.3(1.6)	−10.9(0.7)	18.7–25.7(23.5)
C. platyphylla	F	8	662.0(0.6)	244.3(0.8)	161.0(0.0)	1067.3(0.5)	65.0(0.6)	65.4(0.6)	55.6(0.5)	−18.1(1.4)	0.3–1.9(1.0)
	P	15	634.9(102.8)	240.1(9.7)	159.7(4.3)	1034.7(112.2)	53.0(5.9)	52.1(4.9)	44.5(3.8)	−11.0(2.6)	9.1–28.3(22.2)
All	F	49	664.8(7.1)	245.5(1.9)	161.0(0.0)	1071.2(6.9)	64.4(0.8)	65.2(1.0)	55.3(0.6)	−16.8(1.1)	0.2–7.5(3.0)
	P	49	657.4(58.1)	242.1(5.3)	158.5(9.7)	1058.0(64.2)	52.2(4.5)	51.6(4.2)	44.8(3.1)	−11.0(2.2)	0.9–30.9(20.6)

C – ITS nrDNA originating from cDNA; F – presumed functional ITS paralogs; P – putative ITS pseudogenes.

especially their 5.8S region, displayed significant deviations ($P < 0.05$ and $P < 0.01$) from expectations of the neutral model, whereas all pseudogenes were found to be under neutral selection (Table 4).

4.3. Test for recombination

Five putative pseudogenes (*C. siamensis*_15, *C. rumphii*_15, *C. circinalis*_5, *C. platyphylla*_14, and *C. platyphylla*_22) were found to be likely recombinants. All could be identified by their sharp discontinuities in the substitution patterns between the presumed functional and putative pseudogene ITS paralogs (Fig. 2A and B). While for *C. siamensis*_15, the major parent sequence was found to be *C. circinalis*_19 and minor parent sequence *C. siamensis*_14, the major parent sequence for the other recombinants could not be identified.

Three further putative pseudogenes (*C. platyphylla*_10, *C. platyphylla*_6, and *C. rumphii*_7) showed similar mixed substitution patterns as the above recombinants (Fig. 2A and B), though GENECONV did not identify these as such. This was likely because these showed a predominance of substitution pattern of functional paralogs, and only a very short region at the 3′end of ITS2 of a pseudogene substitution pattern, or reverse substitution pattern (*C. platyphylla*_22) (Fig. 2B) (c.f. Alhiyafi et al., 2007). All likely recombinants were labeled R and removed from the phylogenetic analysis since they introduced conflicting phylogenetic signals, those may result in the loss of clade resolution in the phylogenetic tree, and may obscure real phylogenetic relationships that may exist (McDade, 1992).

4.4. Neighbor joining tree

The topology of the NJ tree reconstructed from the entire ITS region of all sequences, minus the eight putative recombinants, showed the same major clades as the strict consensus MP tree (data not shown). In the unrooted NJ tree (Fig. 3), all functional cDNA and presumed functional genomic ITS copies grouped together in a monophyletic clade with high bootstrap support (NJ-BS = 86%; MP-BS = 83%). The putative pseudogenes of each species were found to be polyphyletic, clustering at random in the tree, and had short internal and long terminal branches. On the contrary, the functional ITS paralogs of most species formed monophyletic clades each with 82% or more bootstrap support. The *C. revoluta* clade included all five cDNA ITS copies, as well as the presumed functional paralogs. Only the paralogs of *C. rumphii* and *C. circinalis* were mixing together in one clade.

5. Discussion

5.1. Characterization of ITS pseudogenes

In the present study, only one pair of primers was used to amplify the ITS sequences, thus only pseudogenes where the primer sites were greatly preserved were obtained. Older and/or more diverged copies, where the primer sites have been lost will likely exist, but are not PCR amplifiable. Among the 98 distinct genomic ITS paralogs amplified here, half have very likely lost function and were categorized as pseudogenes. This was inferred from several sequence characteristics, mainly in comparison to the *C. revoluta* functional reference cDNA ITS paralogs. The putative pseudogenes had a significantly lower GC content, possessed lower secondary structure minimum free energy of 5.8S nrDNA, lacked a conserved seed plant specific 14-bp motif in 5.8S, showed increased nucleotide substitutions and sequence diversity, including pairwise distance, number of nucleotide differences (K), and nucleotide diversity (π). The criteria we used to identify pseudogenes were very powerful and even showed that some sequences used in previous phylogenetic studies of cycads (e.g. Chaw et al., 2005) were in fact pseudogenes (data not shown).

Pseudogenes apparently evolved without functional constraints and mutate neutrally, as indicated by the Tajima's D test here. Conversely, functional ITS paralogs were found to be subject to evolutionary constraints. This will be related to the maintenance of a specific secondary structure for the processing of mature RNAs (Mai and Coleman, 1997).

The 5.8S gene is the most reliable indicator of the functionality of ITS paralogs within the ITS region (Hershkovitz et al., 1999), and Yokota et al. (1989) suggested that the GC content of 5.8S in functional copies in plants ranges from 50.6% to 59.3%. Functional *Cycas* 5.8S copies showed an average of 55.3% GC content, which is well in line with the values of Yokota et al. (1989), while the values for the putative pesudogenes were with an average of 44.8%, significantly lower. Their low GC content, resulting from methylation related substitutions, can lead to the loss of gene function and may be the origin of the pseudogenes (reviewed by Bailey et al., 2003).

Evolutionary constraints can be detected by inspecting low-energy secondary structure models (Wolf et al., 2005). The

secondary structure stability of 5.8S is important for the function of post-transcriptional processes. The secondary structure modeling here demonstrated that the integrity of this structure is greatly compromised in the putative pseudogenes as seen from their increased free energy values, indicative of their lowered stability.

5.2. Marked intra-genomic ITS length variation

The length variation of the entire ITS region in *Cycas* was similar to other non-flowering plants (e.g. Xiang et al., 2000; Campbell et al., 2005; Won and Renner, 2005). Surprisingly, we found a 61-bp deletion in 5.8S in *C. circinalis*_16, which, to our knowledge, is the largest one observed in seed plants. Though based on the sequence characteristics it was categorized as a pseudogene (see below). In non-flowering plants, many large ITS length variations (>100 bp) have been observed. Among those, inter-specific ones represented a large proportion, and intra-genomic ones were only reported in *Cedrus deodara* (Liston et al., 1996), *Pinus sylvestris* (Karvonen and Savolainen, 1993) and *Larix potaninii* (Wei et al., 2003). In *Cycas*, the shortest paralog, *C. platyphylla*_16 with 735 bp, and the longest one, *C. platyphylla*_20 with 1091 bp, came from the same individual, and both were identified as putative pseudogenes (see above). In contrast, the average ITS lengths among species showed no significant variation (Table 3), indicating that similar length variations occurred in all *Cycas* species.

As in *Gnetum* (Won and Renner, 2005), no repeats were detected in *Cycas* using the Tandem Repeats Finder (Benson, 1999) under the strict options between ITS sequences, and even under the most relaxed search options, only a few short repeats (7–23 bp) were detected and were limited to paralogs of single species (except for one repeat shared by ITS2 of *C. cicirnalis* and *C. rumphii*). Those paralogs did not show large length variations.

These findings suggest that the presence of indels, rather than tandem repeats, as in Pinaceae (Kan et al., 2007), play an important role in the observed great ITS length variation in *Cycas*.

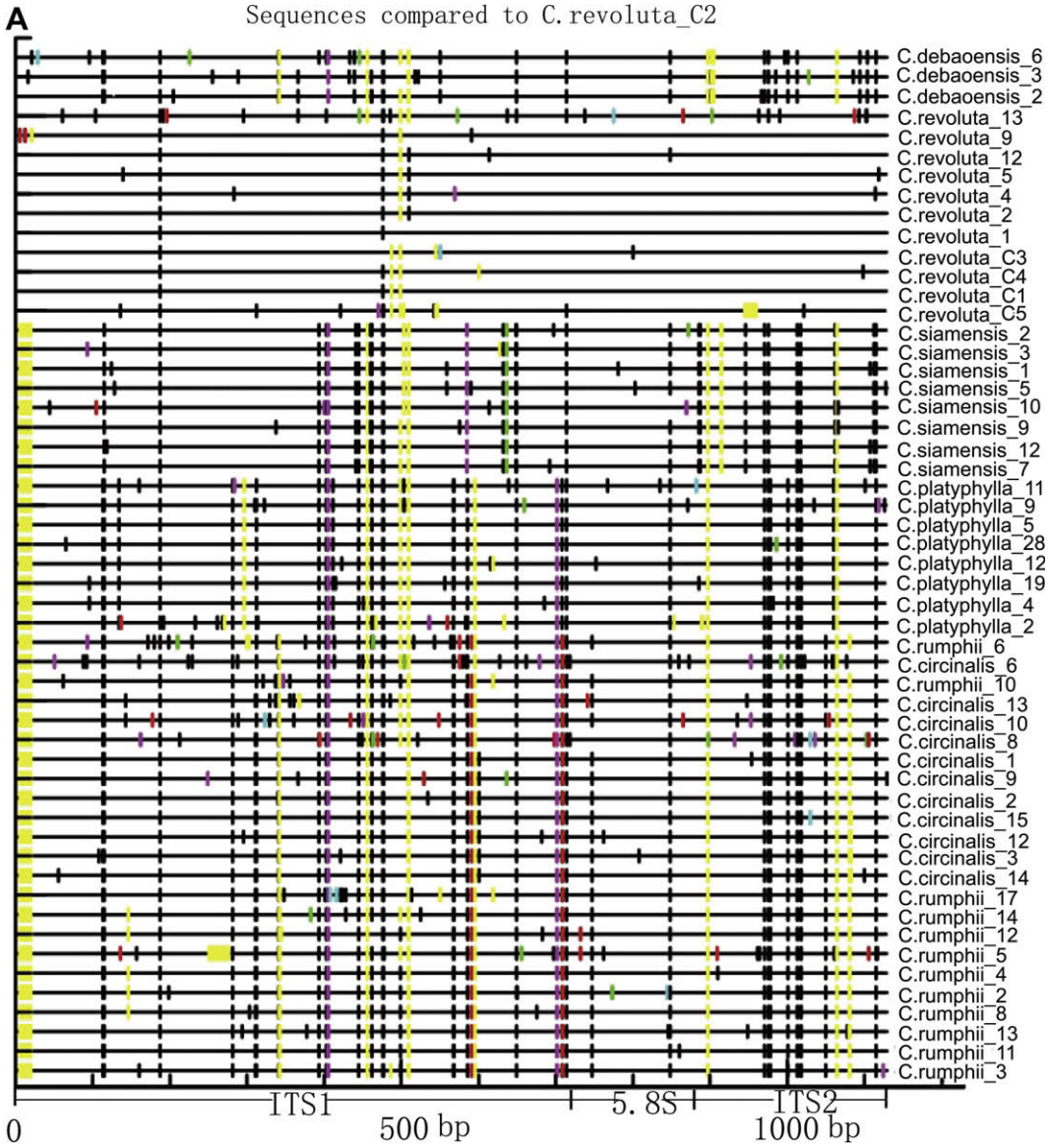

Fig. 2. Schematic illustration of the distribution of substitution sites across the entire ITS region obtained from six species of *Cycas*, using *C. revoluta*_C2 as reference. (A). For 49 presumed functional and four cDNA paralogs. (B). For 41 putative pseudogenes and eight putative recombinants (R). (red = GG > AG, cyan = GA > AA, green = GC > AC, magenta = GT > AT, black = not G > A transition, yellow = gap). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

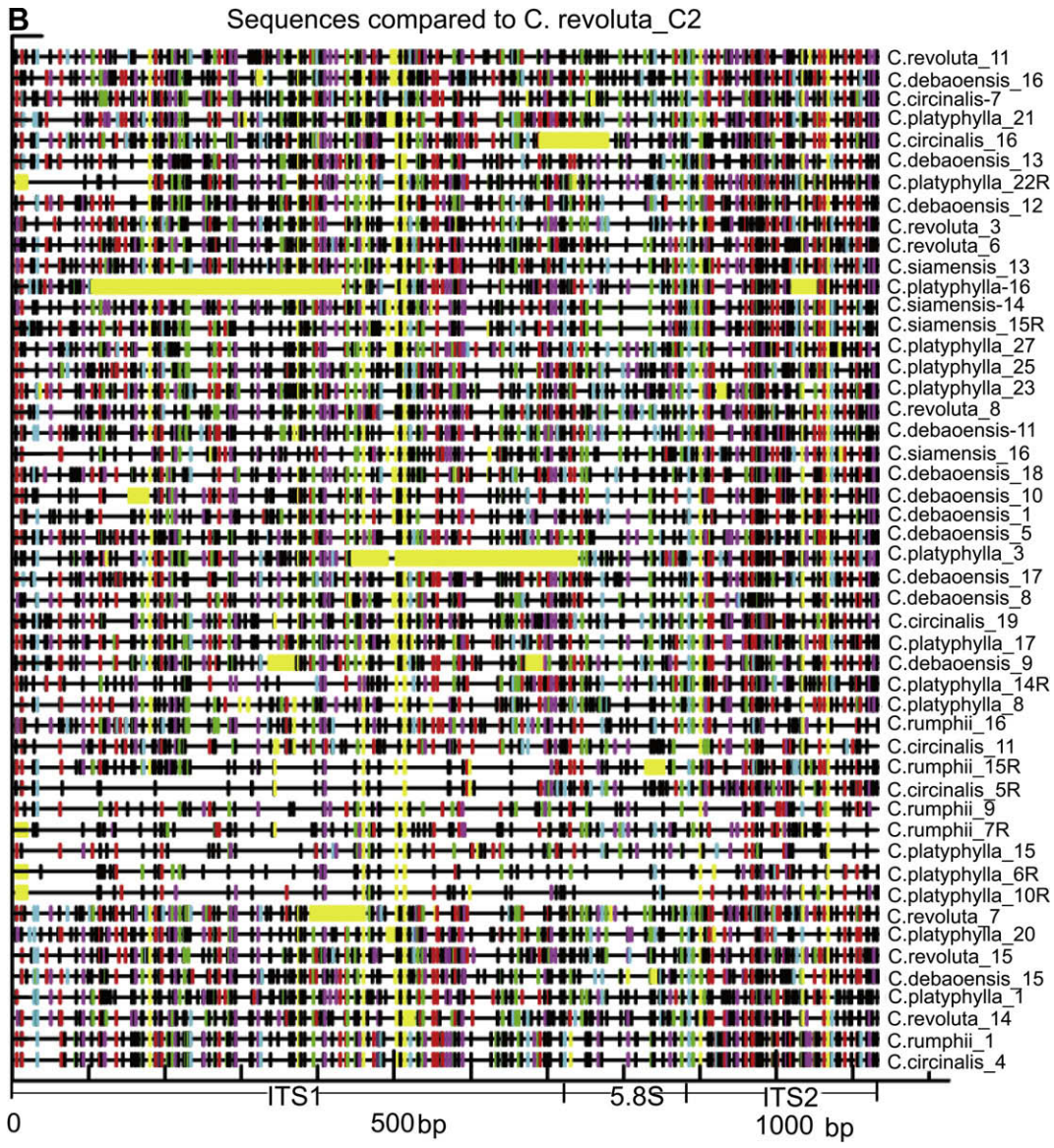

Fig. 2 (*continued*)

Table 4
Nucleotide diversity of individual parts and the entire ITS region of six *Cycas* species analysed.

Region	Sequence type[a]	Sequence number	Polymorphic sites	Total no of mutations	K[b]	π[c]	Tajima's D[d]
ITS1	C	5	15	15	6.200	0.009	−1.015
	F	49	169	179	19.315	0.031	−1.770*
	P	47	341	520	82.100	0.182	0.232
ITS2	C	5	2	2	0.800	0.003	−0.973
	F	49	75	79	8.324	0.035	−1.782*
	P	47	179	267	48.061	0.215	0.675
5.8S	C	5	1	1	0.400	0.002	−0.817
	F	49	29	29	2.247	0.014	−2.178**
	P	47	52	78	8.716	0.121	−0.906
ITS	C	5	18	18	7.400	0.007	−1.057
	F	49	273	287	29.886	0.029	−1.860**
	P	47	574	858	139.286	0.187	0.264

[a] C, F, and P as in Table 3,but the pseudogenes do not include the two sequences with over 100-bp deletions here.
[b] Average number of nucleotide differences (K).
[c] Nucleotide diversity (π).
[d] $*p < 0.10$, $**p < 0.05$.

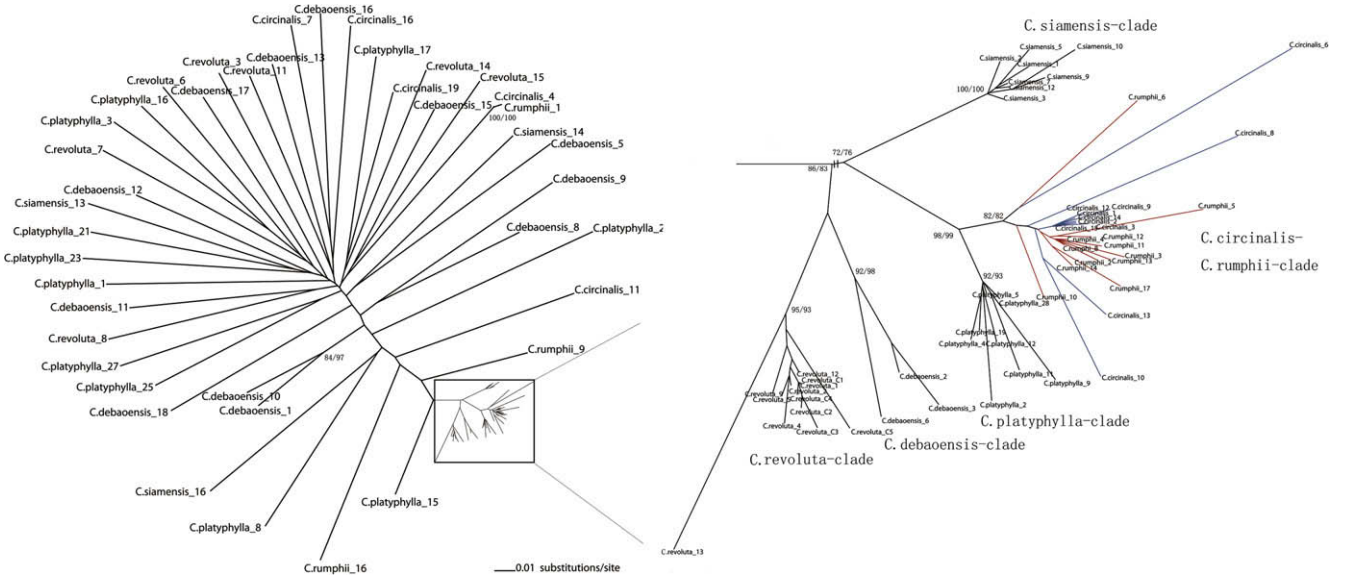

Fig. 3. Unrooted Neighbor-Joining tree constructed from a HKY85 distance matrix using the entire ITS region of 95 paralogs obtained from six species of *Cycas* (excluding eight putative recombinants). Numbers indicate bootstrap values (>50%) for NJ (left) and MP (right) analyses, respectively. || denotes the 14-bp deletion.

5.3. ITS recombinants

Recombination can explain the existence of chimeric sequences found in several angiosperms (Buckler et al., 1997; Muir et al., 2001) as well as gymnosperm lineages (*Pinus*, Gernandt et al., 2001; *Larix*, Wei et al., 2003). In the present study, the seven paralogs found showed sharp discontinuities in the patterns of sequence similarity and substitution pattern between funcitonal ITS paralogs and pseudogenes (Fig. 2B), and suggests their origin through recombination between the two types of paralogs. However, in *C. siamensis*_15 the major sequence part is similar to *C. circinalis*_19, and a minority similar to *C. siamensis*_14, and could be the result from combination between two ancestral pseudogenes. Thus, the substitution discontinuity here was not as sharp and easy to detect as in the other recombinants. Among these, some unique recombinants, such as *C. platyphylla*_14, *C. platyphylla*_22, *C. platyphylla*_10, and *C. platyphylla*_6, and, *C. rumphii*_15 and *C. rumphii*_7, were obtained from the same PCR reaction per species. But the physical location of nucleotide substitutions showed that they have different recombination break points and have accumulated sequence specific substitutions. Thus, they could be considered true recombinants rather than PCR recombinants (Bradley and Hillis, 1997; Wei et al., 2003). Older recombination will have a greater likelihood of their possible parental sequences being obscured by subsequent mutations. In *Cycas*, the recombinants, whose major parental sequences could not be detected here, might have come from ancient recombination events. It is worthwhile noting, however, that recombination between functional paralogs may be a continuous process, and recent recombinants may be very difficult to identify because the sequence difference between the parental copies is too small (*Cycas* intra-genomic values <4.6%) (Alhiyafi et al., 2007).

5.4. Incomplete concerted evolution of ITS in Cycas

In *Cycas*, the high degree of intra-genomic polymorphism in the ITS region suggests incomplete concerted evolution. Divergent ITS paralogs, including putative pseudogenes, recombinants and multiple functional ITS copies were detected in the genome of each species, i.e. plant, in this study. *Cycas* is assumed to be diploid ($2n = 22$ chromosomes; Johnson and Wilson, 1990; Sax and Beale, 1934), and there is no evidence to indicate that it is a paleopolyploid, having undergone consecutive diploidisation. Thus, a genome duplication, and hence duplication of nrDNA loci, may not be responsible for the intra-genomic polymorphisms here.

The 18S–26S rDNA arrays and their RNA products constitute essential components of the eukaryotic nucleolar organizer region (NOR). NORs interspersed on non-homologous chromosomes, or located at inactive loci can lead to a low rate of concerted evolution (Komarova et al., 2004). Most diploid angiosperms have only one or two NORs per genome (Long and Dawid, 1980; Rogers and Bendich, 1987; Wendel et al., 1995). Thus, their ITS paralogs may be greatly homogenized through concerted evolution, though exceptions exist when hybridization is involved (such as in *Quercus*; Muir et al., 2001). In diploid gymnosperms, the number of NORs is often much higher, e.g. three in *Larix* (Lubaretz et al., 1996), or 10–12 in *Pinus* and *Picea* (Quijada et al., 1998), and the high intra-specific/genomic ITS polymorphisms observed in Pinaceae (Kan et al., 2007; Karvonen and Savolainen, 1993; Wei et al., 2003; Wei and Wang, 2004) and *Gnetum* (Won and Renner, 2005) may be linked to this aspect.

In *C. revoluta*, FISH rDNA sites were found on 16 out of the 22 chromosomes, and of these only 13 rDNA sites were functional, based on results from silver staining (AG-NORs) (Hizume et al., 1992). This suggests that some of the NOR sites in *C. revoluta* are non-functional, and their paralogs could have escaped from concerted evolution, and have become pseudogenes, "orphan-like" ribosomal-DNA genes (Childs et al., 1981). However, the number and distribution of NOR arrays are evolutionary labile (Álvarez and Wendel, 2003; Datson and Murray, 2006), and we do not know the physical origin of the pseudogenes isolated from the *Cycas* genomes here, whether they cluster together in one locus or occur interspersed among functional copies. Furthermore, the number of sites in the other *Cycas* species, beside *C. revoluta*, is as yet unknown. For a better understanding of the dynamic of NOR number and concerted evolution in *Cycas*, FISH and Ag-NOR data for more species in *Cycas* are needed.

Interestingly, the nrDNA ITS phylogenetic tree (Fig. 3) showed that, although the functional clones from *C. revoluta*, *C. debaoensis*, *C. siamensis* and *C. platyphylla* formed four separating monophyletic groups, those from *C. rumphii* and *C. circinalis* were mixed together. This phylogenetic pattern indicates that incomplete lineage

sorting or hybridization is likely a further process resulting in intra-genomic ITS polymorphism in *Cycas*, at least for *C. rumphii* and *C. circinalis*. Between the two species, the highly morphological similarity and widespread taxonomic confusion (Hill, 1994) suggest their recent divergence. Thus, perhaps, there had been is insufficient time for complete lineage sorting and elimination of ancestral nrDNA ITS polymorphism following diversification from their most recent common ancestor. And, to our knowledge, no morphological intermediates have been reported, indicative of hybridization between *C. rumphii* and *C. circinalis*, and they do not occur in close geographical proximity (Table 1).

5.5. Pseudogene origin and ITS phylogenetic signals

The ITS phylogeny (Fig. 3) shows that in *Cycas*, functional paralogs and pseudogenes are well separated and display completely different phylogenetic patterns. The functional paralogs of each species form separate monophyletic clades, except for those species having greater seed dispersal potential (see above). This suggests that the phylogenetic analysis can provide some insights into the evolution in *Cycas*. However, the currently debated evolutionary relationships and delimitations of intra-generic taxonomic units cannot be addressed here, because a too small sample size (only 6 out of about 90 species) was sampled.

In contrast, pseudogenes have little functional constraint and evolve much more rapidly than the functional paralogs (over 6 times faster than functional copies), even outpacing speciation. Because of the high number of random acquisition of mutations in pseudogenes, they cluster together randomly as seen from the ITS tree. This is likely due to long-branch attraction (Anderson and Swofford, 2004). The short internal and extremely long terminal branches are typical for such a LBA scenario (Fig. 3).

Evidence for the origin of the pseudogenes comes from the 14-bp deletion near the ITS1 5′-end, and support a scenario where they predate the diversification of the genus: in *C. revoluta* and *C. debaoensis*, both the functional paralogs and pseudogenes, lack the 14-bp deletion. But in *C. siamensis*, *C. rumphii*, *C. circinalis*, and *C. platyphylla*, the deletion is present in their functional ITS paralogs, and absent in the pseudogenes, except for four recombinants (*C. platyphylla*_6, 10, and 22, and *C. rumphii*_7). There is a debate about the basal lineage in *Cycas*. Section *Asiorientalis* (i.e. *C. revoluta* and *C. taitungensis*) is generally accepted as the basal lineage in *Cycas* (Hill, 1999) because of the tomentose ovules of its members (vs glabrous in the other sections). This is also supported by the earliest known fossil in *Cycas*, *C. fujiana* Yokoyama (Buckley, 1999). This Eocene fossil species is very similar to *C. revoluta*. Both, the morphological similarity and the occurrence of both taxa in Japan, suggest that *C. fujiana* and *C. revoluta* belong to the same lineage, or even the same species (cf. Pant, 1999). Wang (2000), however, argued that the direction of evolutionary change in leaf morphology in *Cycas* is from dichotomous pinnate to undivided leaves, and proposed the *C. micholitzii* complex (represented here by *C. debaoensis*) as the basal lineage. The unrooted NJ tree shows that, irrespective as to whether section *Asiorientalis* or the *C. micholitzii* complex are basal in *Cycas* (both are part of the same lineage), the 14-bp deletion is a synapomorphy for the functional ITS paralogs in *C. siamensis*, *C. rumphii*, *C. circinalis*, and *C. platyphylla* (indicated by || in Fig. 3).

Thus, the pseudogenes are likely to have an early origin, predating the diversification of *Cycas*. Combined with their high rate of ITS evolution, many parallel, homoplasious, nucleotide changes will have occurred, causing LBA problems. This resulted in the pseudogenes forming non–monophyletic clades. It is obvious, however, that the usefulness of these fast evolving pseudogenes for phylogenetic reconstructions is rather limited. Though, they could be very useful as outgroups in future phylogenetic studies on *Cycas* as Buckler and Holtsford (1996a, 1996b) pointed out that these types of pseudogenes can be taken as better outgroups than sister species and be more useful for groups without closely related extant taxa. Future studies in this plant group with more samples will show whether this is indeed the case.

Acknowledgments

We thank Dr. Li Nan and Dr. Zhang Shouzhou, Shenzhen Fairylake Botanical Garden, for help in the collection of study material, and Yang Junbo, Kunming Institute of Botany, CAS, for supporting the molecular work. Financial support was provided by the Yunnan Province Natural Science Foundation (0755051K01) and the Knowledge Innovation Program of the Chinese Academy of Sciences (O8LY081K01). R.B.G.E. is supported by the Scottish Government Rural and Environment Research and Analysis Directorate (RERAD).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.11.020.

References

Alhiyafi, J., Sebesan, C., Shiyong, L., Harriott, M., Jurczyszyn, D., Ritchie, R.P., Gonzales, F.S., Ram, J.L., 2007. Computational methods for analysis of cryptic recombination in the performance of genomic recombination detection software. AINA Workshops 1, 707–712.

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

Anderson, F.E., Swofford, D.L., 2004. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. Mol. Phylogenet. Evol. 33, 440–451.

Bailey, J.A., Liu, G., Eichler, E.E., 2003. An *Alu* transposition model for the origin and expansion of human segmental duplications. Am. J. Hum. Genet. 73, 823–834.

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

Bradley, R.D., Hillis, D.M., 1997. Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. 14, 592–593.

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

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

Buckler, E.S., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145, 821–832.

Buckley, R., 1999. A brief review of the fossil cycads. Available from: <http://www.plantapalm.com>.

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.

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., Walters, T.W., Chang, C.C., Hu, S.H., Chen, S.H., 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast *mat*K gene, *trn*K intron, and nuclear rDNA ITS region. Mol. Phylogenet. Evol. 37, 214–234.

Childs, G., Maxson, R., Cohn, R.H., Kedes, L., 1981. Orphans, dispersed genetic elements derived from tandem repetitive genes of eucaryotes. Cell 23, 651–663.

Datson, P.M., Murray, B.G., 2006. Ribosomal DNA locus evolution in *Nemesia*: transposition rather than structural rearrangement as the key mechanism? Chromosome Res. 14, 845–857.

Denduangboripant, J., Cronk, Q.C.B., 2000. High intraindividual variation in internal transcribed spacer sequences in *Aeschynanthus* (Gesneriaceae): implication for phylogenetics. Proc. R. Soc. Lond. 267, 1407–1415.

Doyle, J., 1991. DNA protocols for plants – CTAB total DNA isolation. In: Hewitt, G.M., Johnston, A. (Eds.), Molecular Techniques in Taxonomy. Berlin, Springer, pp. 283–293.

Grimm, G.W., Denk, T., 2007. ITS evolution in *Platanus* (Platanaceae): homoeologues, pseudogenes and ancient hybridization. Ann. Bot. 101, 403–419.

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.

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.

Harpke, D., Peterson, A., 2006. Non-concerted ITS evolution in *Mammillaria* (Cactaceae). Mol. Phylogenet. Evol. 41, 579–593.

Hershkovitz, M.A., Zimmer, E.A., Hahn, W.J., 1999. Ribosomal DNA sequences and angiosperm systematics. In: Hollingsworth, P.M., Bateman, R.M., Gornall, R.J. (Eds.), Molecular Systematics and Plant Evolution. Taylor and Francis, London, pp. 268–326.

Hill, K.D., 1994. The *Cycas rumphii* complex (Cycadaceae) in New Guinea and the Western Pacific. Aust. Syst. Bot. 7, 543–567.

Hill, K.D., 1999. Cycas – an evolutionary perspective. In: Chen, C.J. (Ed.), Biology and Conservation of Cycads – Proceedings of the Fourth International Conference on Cycad Biology. International Academic Publishers, Beijing.

Hill, K.D., 2004. Character evolution, species recognition and classification concepts in the Cycadaceae. In: Walters, T., Osborne, R. (Eds.), Cycad Classification Concepts and Recommendations. CABI Publishing, Wallingford, pp. 23–44.

Hizume, M., Ishida, F., Kondo, K., 1992. Differential staining and *in situ* hybridization of nucleolar organizers and centromeres in *Cycas revoluta* chromosomes. Jpn. J. Genet. 67, 381–387.

Jobes, D.V., Thien, L.B., 1997. A conserved motif in the 5.8S ribosomal RNA (rRNA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) sequences. Plant Mol. Biol. Rep. 15, 326–334.

Johnson, L.A.S., Wilson, K.L., 1990. General traits of the Cycadales. In: Kramer, K.U., Green, P.S. (Eds.), The Families and Genera of Vascular Plants. Springer-Verlag, Berlin, pp. 363–368.

Kan, X.Z., Wang, S.S., Ding, X., Wang, X.Q., 2007. Structural evolution of nrDNA ITS in Pinaceae and its phylogenetic implications. Mol. Phylogenet. Evol. 44, 765–777.

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

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

Kita, Y., Ito, M., 2000. Nuclear ribosomal ITS sequences and phylogeny in East Asian *Aconitum* subgenus *Aconitum* (Ranunculaceae), with special reference to extensive polymorphism in individual plants. Plant Syst. Evol. 225, 1–13.

Komarova, N.Y., Grabe, T., Huigen, D.J., Hemleben, V., Volkov, R.A., 2004. Organization, differential expression and methylation of rDNA in artificial *Solanum* allopolyploids. Plant Mol. Biol. 56, 439–463.

Kumar, S., Tamura, K., Nei, M., 1994. MEGA, Molecular Evolutionary Genetics Analysis software for microcomputers. Bioinformatics 10, 189–191.

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.

Long, E.O., Dawid, I.B., 1980. Repeated genes in eukaryotes. Annu. Rev. Biochem. 49, 727–764.

Lubaretz, O., Fuchs, J., Ahne, R., Meister, A., 1996. Karyotyping of three Pinaceae species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. Theor. Appl. Genet. 92, 411–416.

Mai, J.C., Coleman, A.W., 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. J. Mol. Evol. 44, 258–271.

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

McDade, L.A., 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution 46, 1329–1346.

Möller, M., Kokubugata, G., Jong, K., 2008. Aspects of genome evolution in Gesneriaceae: patterns of 45S-nrDNA site evolution based on evidence from fluorescent *in situ* hybridization (FISH). Int. J. Plant Sci. 169 (5), 667–676.

Muir, G., Fleming, C.C., Schlötterer, C., 2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L.. Mol. Biol. Evol. 18, 112–119.

Nagylaki, T., 1984. The evolution of multigene families under intrachromosomal gene conversion. Genetics 106, 529–548.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

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.

Ochieng, J.W., Henry, R.J., Baverstock, P.R., Steane, D.A., Shepherd, M., 2007. Nuclear ribosomal pseudogenes resolve a corroborated monophyly of the eucalypt genus *Corymbia* despite misleading hypotheses at functional ITS paralogs. Mol. Phylogenet. Evol. 44, 752–764.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Pant, D.D., 1999. A study in contrasts, the present and past distribution and form of certain cycads. In: Chen, C.J. (Ed.), Biology and conservation of cycads – Proceedings of the Fourth International Conference on Cycad Biology. International Academic Publishers, Beijing.

Quijada, A., Liston, A., Delgado, P., Vázquez-Lobo, A., Alvarez-Buylla, E.R., 1998. Variation in the nuclear ribosomal DNA internal transcribed spacer (ITS) region of *Pinus rzedowskii* revealed by PCR-RFLP. Theor. Appl. Genet. 96, 539–544.

Rogers, S.O., Bendich, A.J., 1987. Ribosomal RNA genes in plants, variability in copy number and in the intergenic spacer. Plant Mol. Biol. 9, 509–520.

Rose, P.P., Korber, B.T., 2000. Detecting hypermutations in viral sequences with an emphasis on G A hypermutation. Bioinformatics 16, 400–401.

Rozas, J., Rozas, R., 1999. DnaSP version 3, an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15, 174–175.

Sanderson, M.J., Doyle, J.J., 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst. Biol. 41, 4–17.

Sax, K., Beale, J.M., 1934. Chromosomes of the Cycadales. J. Arnold Arbor. 15, 255–258.

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

Sone, T., Fujisawa, M., Takenaka, M., Nakagawa, S., Yamaoka, S., Sakaida, M., Nishiyama, R., Yamato, K.T., Ohmido, N., Fukui, K., Fukuzawa, H., Ohyama, K., 1999. Bryophyte 5S rDNA was inserted into 45S rDNA repeat units after the divergence from higher land plants. Plant Mol. Biol. 41, 679–685.

Stevenson, D.W., 1990. Morphology and systematics of the Cycadales. Mem. NY Bot. Gard. 57, 8–55.

Suh, Y., Thien, L.B., Zimmer, E.A., 1992. Nucleotide sequences of the internal transcribed spacers and 5.8S rRNA gene in *Canella winterana* (Magnoliales; Canellaceae). Nucleic Acids Res. 20, 6101–6102.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), Version 4.0. Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface, flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Treutlein, J., Wink, M., 2004. Molecular phylogeny of cycads inferred from *rbc*L sequences. Naturwissenschaften 89, 221–225.

Wang, D.Y., 2000. Studies on Morphology, Anatomy, Taxonomy and Evolution of Cycadaceae. Ph.D. Dissertation. Nanjing, Nanjing Forestry University.

Wei, X.X., Wang, X.Q., 2004. Recolonization and radiation in *Larix* (Pinaceae), evidence from nuclear ribosomal DNA paralogs. 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.

Wolf, M., Achtziger, M., Schultz, J., Dandekar, T., Müller, T., 2005. Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. RNA 11, 1616–1623.

Won, H., Renner, S.S., 2005. The internal transcribed spacer of nuclear ribosomal DNA in the gymnosperm *Gnetum*. Mol. Phylogenet. Evol. 36, 581–597.

Xiang, Q.P., Xiang, Q.Y., Liston, A., Fu, L.K., 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.

Yang, S.L., Meerow, A.W., 1996. The *Cycas pectinata* (Cycadaceae) complex structure and gene flow. Int. J. Plant Sci. 157, 468–483.

Yokota, Y., Kawata, T., Iida, Y., Kato, A., Tanifuji, S., 1989. Nucleotide sequences of the 58S rRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. J. Mol. Evol. 29, 294–301.

Zheng, X.Y., Cai, D.Y., Yao, L.H., Teng, Y.W., 2008. Non-concerted ITS evolution, early origin and phylogenetic utility of ITS pseudogenes in *Pyrus*. Mol. Phylogenet. Evol. 48, 892–903.

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