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.