

Evolutionary Stasis in Cycad Plastomes and the First Case of Plastome GC-Biased Gene Conversion

Chung-Shien Wu and Shu-Miaw Chaw*

Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

*Corresponding author: E-mail: smchaw@sinica.edu.tw.

Data deposition: This project has been deposited at NCBI and DDBJ under the accessions NC_026036, LC049068, LC049070, LC049336, LC049207, LC049069, LC049067, and LC040885.

Accepted: June 22, 2015

Abstract

In angiosperms, gene conversion has been known to reduce the mutational load of plastid genomes (the plastomes). Particularly, more frequent gene conversions in inverted repeat (IR) than in single copy (SC) regions result in contrasting substitution rates between these two regions. However, little has been known about the effect of gene conversion in the evolution of gymnosperm plastomes. Cycads (Cycadophyta) are the second largest gymnosperm group. Evolutionary study of their plastomes is limited to the basal cycad genus, *Cycas*. In this study, we addressed three questions. 1) Do the plastomes of other cycad genera evolve slowly as previously observed in the plastome of *Cycas taitungensis*? 2) Do substitution rates differ between their SC and IR regions? And 3) Does gene conversion occur in the cycad plastomes? If yes, is it AT-biased or GC-biased? Plastomes of eight species from other eight genera of cycads were sequenced. These plastomes are highly conserved in genome organization. Excluding ginkgo, cycad plastomes have significantly lower synonymous and nonsynonymous substitution rates than other gymnosperms, reflecting their evolutionary stasis in nucleotide mutations. In the IRs of cycad plastomes, the reduced substitution rates and GC-biased mutations are associated with a GC-biased gene conversion (gBGC) mechanism. Further investigations suggest that in cycads, gBGC is able to rectify plastome-wide mutations. Therefore, this study is the first to uncover the plastomic gBGC in seed plants. We also propose a gBGC model to interpret the dissimilar evolutionary patterns as well as the compositionally biased mutations in the SC and IR regions of cycad plastomes.

Key words: plastome, biased mutation, GC-biased gene conversion, cycad, gymnosperm.

Introduction

Chloroplasts are photosynthetic plastids (hereafter abbreviated as plastids) and contain their own genomes, called plastomes. During photosynthesis, an excess of solar energy in plastids leads to increased reactive oxygen species (ROS) that can cause DNA damage in plastomes (Kumar et al. 2014). Given that plastomes are constantly exposed to the mutagenic agents of ROS, efficient DNA repair systems are required to prevent plastomes from dysfunction. Moreover, the uniparental, typically maternal, inheritance of plastomes lacks sexual recombination to eliminate deleterious mutations. As a result, plastomes are expected to accumulate mutations over time because of Muller's ratchet (Muller 1964).

In leaf cells, a plastid can contain approximately 1,000 copies of plastomes (Bendich 1987). The nature of multiple plastomes allows for highly efficient gene conversion in asexual genetic systems, correcting mutations and reducing the mutational load of plastomes (Khakhlova and Bock 2006). In plastids, gene conversion takes place in recombination-

dependent replications and is able to repair broken replication forks and maintain plastome stability (Maréchal and Brisson 2010). To date, direct measurements of gene conversion events are only known for the start codons of *ycf1* and *ycf2* genes in transgenic tobacco plastids, in which AT-biased gene conversion was suggested (Khakhlova and Bock 2006). However, this AT-biased gene conversion in the tobacco plastome was considered exceptional, given that biased gene conversion generally favors incorporation of GC bases in most genomes studied (Smith 2012).

Plastomes of seed plants usually contain a pair of inverted repeats (IRs), except for a few lineages, such as legumes and conifers, where one copy of IRs has been completely lost or extremely reduced (e.g., Perry et al. 2002; Cai et al. 2008; Lin et al. 2010; Wu et al. 2011; Guo et al. 2014; Hsu et al. 2014; Wu and Chaw 2014). In angiosperm plastomes, genes located in IRs feature relatively slower substitution rates than those in the single-copy (SC) regions (Wolfe et al. 1987; Maier et al. 1995). These dissimilar substitution rates might result from

greater frequency of gene conversion in IR than in SC regions (Birky and Walsh 1992; Perry and Wolfe 2002). Consequently, IR regions are hotspots of gene conversion and are most suitable for studying the gene conversion machinery in plastomes.

Gene conversion and its effect on the evolution of gymnosperm plastomes are poorly understood. Whether the SC and IR regions of gymnosperm plastomes have similar or contrasting substitution rates remains to be investigated. Phylogenetic-scale analyses have previously revealed that seed plants, including gymnosperms and angiosperms, typically are compositionally biased toward AT-richness in their plastomes (Kusumi and Tachida 2005; Smith 2009, 2012). Therefore, if gene conversion per se is AT-biased in gymnosperm plastomes, we would expect 1) reduced substitution rates in their IR regions and 2) more AT-biased mutations in IR than SC regions. Verifying these possibilities requires comparison of mutational patterns between SC and IR regions in gymnosperm plastomes.

After conifers, cycads (Cycadophyta) are the second largest of the five major groups of living gymnosperms. They include some 300 living species of trees or shrubs in nine or ten genera and three families in the tropics and subtropics of the Old and New Worlds (Chaw et al. 2005). The extant cycads were once considered “living fossils” because their morphologies have changed little since the Upper Triassic, about 235–201 Ma (Renner 2011). However, recent molecular dating has indicated that speciation of living cycads has occurred from a recent global rediversification, not more than 12 Ma (Nagalingum et al. 2011). Cycads have high horticultural value because of their attractive pinnate leaves, elegant giant cones, palm-like trunks, and rarity (Brenner et al. 2003). Nonetheless, many cycad species are facing declining populations because of human activities such as destruction of the natural habitat and removal of plants from the wild for collection (Donaldson 2003). About 40% of cycad species are currently endangered, and four species are listed to become extinct in the wild (IUCN Red List, February 2015) (<http://www.iucnredlist.org/>).

Research of plastomes remains very limited in cycads. Previously, we reported that the plastome of *Cycas taitungensis* has slower substitution rates than those of other seed plants (Wu et al. 2007). In this study, we sequenced the complete plastomes of eight representatives of cycad genera other than the basal-most genus *Cycas*. We addressed three key questions. 1) Do the plastomes of other cycad genera evolve slowly as was previously observed in the plastome of *Cycas taitungensis*? 2) Do substitution rates differ between the SC and IR regions of cycad genera? And 3) Does gene conversion occur in the cycad plastomes? If yes, is it AT-biased or GC-biased?

To address the above-mentioned questions, we compared synonymous and nonsynonymous substitution rates between cycads and other gymnosperms. The substitution rates of non-coding loci between IR and SC regions of sampled cycad

genera were also compared. We reasoned that investigating gene conversion by the use of noncoding loci can avoid the effects of selection. Meanwhile, biased mutations of these two regions were evaluated on the basis of the estimated equilibrium GC content (GC_{eq}). This study is the first to uncover the plastomic GC-biased gene conversion (gBGC) in seed plants. In order to explain our novel finding, we proposed a model to link the gBGC mechanism and the evolution of plastomes in cycads.

Materials and Methods

DNA Extraction and Plastome Sequencing

Leaves were harvested from individuals of the eight sampled cycads (supplementary table S1, Supplementary Material online) growing in the greenhouse at Academia Sinica (Taipei). For each sampled species, DNA was extracted using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Extracted DNA was sequenced at Yourgene Bioscience (New Taipei City, Taiwan) using an Illumina GAI sequencer. Sequencing depth was 2.5 GB of 90-bp paired-end reads for each species.

Plastome Assembly and Annotation

The raw sequencing reads were quality-trimmed and de novo-assembled using CLC Genomics Workbench v5.5.1 software (CLC Bio, Aarhus, Denmark). Contigs with length < 1 kb and sequence coverage < 50 × were discarded. The remaining contigs were analyzed by a BLAST search against the plastome of *Cycas taitungensis* (Wu et al. 2007). Contigs that matched the reference plastomic sequences with E -value < 10^{-10} were designated as plastomic contigs. DNA fragments between plastomic contigs were obtained using the *Taq* 2X Master Mix Red PCR kits (Ampliqon, Copenhagen, Denmark) with our species-specific primers. Sequences of PCR amplicons were obtained using an ABI 3730 DNA Analyzer (Life Technologies, Taipei, Taiwan). Plastome annotation was performed using DOGMA (Wyman et al. 2004) and tRNAscan-SE 1.21 (Schattner et al. 2005). The annotated genes were confirmed by their alignment with their orthologous genes from published gymnosperm plastomes (supplementary table S1, Supplementary Material online).

Sequence Alignment and Phylogenetic Tree Construction

Sequence alignment was conducted using MUSCLE (Edgar 2004) implemented in Mega 5.2 (Tamura et al. 2011). Sequences of tRNAs, rRNAs, and noncoding loci were first aligned with the default parameters, then realigned with the refining alignment option. Sequences of protein-coding genes were aligned with the Align Codons option using default parameters. Phylogenetic trees were constructed using PhyloBayes 3.3f (Lartillot et al. 2009) based on the concatenated amino acid sequences of 72 genes common to cycad plastomes with the parameters “-gtr” and “-cat.” Two

independent runs were conducted, with each run yielding 5,000 trees. The first 25% of trees in each run were discarded as burn-in, and the remaining trees in the two runs were compared using the bpcmp program of PhyloBayes to generate a consensus tree with “maxdiff” = 0.012.

Identification of Syntenic Loci

In cycad plastomes, if orthologous coding genes/exons were found in the same order, they were classified as syntenic coding loci. Intergenic spacers/introns were classified as syntenic noncoding loci if they possessed the same flanking genes/exons among the examined cycad genera. Pseudogenes and intergenic spacers flanked by pseudogenes were not included. According to their locations, the defined syntenic noncoding loci were further divided into SC and IR groups for estimating their substitution rates. Noncoding loci that contained IR boundaries were designated as the IR group because their sequences largely resided in the IR regions.

Estimation of Substitution Rates and Sequence Divergences

Alignments of all orthologous protein-coding genes shared by *Amborella trichopoda* and the five gymnosperm groups were concatenated to generate a data set with 61,444 characters. Synonymous (*ds*) and nonsynonymous (*dn*) substitution rates of the examined gymnosperm species (supplementary table S1, Supplementary Material online) were estimated with the CodeML program of PAML 4.8 (Yang 2007) using the options of seqtype = 1, runmodel = -2, and CodonFreq = 3. *Amborella trichopoda* was the reference for estimating the *ds* and *dn* substitution rates because it has been considered the basal-most angiosperm clade (*Amborella* Genome Project 2013). For the syntenic noncoding loci, pairwise mutation rates between *Cycas* and the eight cycads were estimated with the Baseml program of PAML 4.8 using the REV model. Sequence divergences of syntenic loci were estimated using DnaSP v5 (Librado and Rozas 2009).

Inference of Ancestral Noncoding Sequences and Estimation of Nucleotide Changes

Alignments of syntenic noncoding loci were divided into SC and IR data sets based on their locations, then alignments in the two data sets were separately concatenated for further analyses. For each data set, ancestral sequences of the eight cycads were inferred using maximum likelihood method and a general time reversible (GTR) + G (four gamma categories) model in Mega 5.2. The tree inferred from 72 plastid protein-coding genes was designated as the “user tree.” Nucleotide changes between ancestral and current sequences were counted by using the “nucleotide substitution pattern” option in the “Seq. Analysis” function of DAMBE 5 (Xia 2013). The polarization of mutations was designated “ancestral nucleotide”-to-

“current one.” For example, a C-to-A mutation indicates that the ancestral nucleotide is “C” and the current one is “A.”

Estimation of GC_{eq}

First, the SC and IR data sets including ancestral and current sequences of noncoding loci were bootstrapped with 100 replications using the “seqboot” program in PHYLIP 3.695 (Felsenstein 2005). We followed the Hershberg and Petrov’s method (2010) for calculating GC_{eq} values. GC_{eq} is the expected GC_{eq} , which was calculated as: rate of AT-to-GC / (rate of AT-to-GC + rate of GC-to-AT), where rate of GC-to-AT = number of GC-to-AT changes / number of GC sites, and rate of AT-to-GC = number of AT-to-GC changes / number of AT sites. Of note, estimation of GC_{eq} values is based on “rate” rather than “count,” which can avoid the potential effect of biased base compositions.

Presentation of Plastome Maps

Plastome maps were drawn by using Circos v0.67 (<http://circos.ca/>, last accessed February 2015) and manually refined with CoreDRAW X5.

Results

“Frozen” Plastomic Organization in Cycads

Maps of the eight sequenced cycad plastomes are circular and showed in figure 1. These plastomes contain a pair of IR regions separating the SC region into two subsets: large SC (LSC) and small SC (SSC) regions (fig. 1). The IR regions range from 24,853 to 26,137 bp. The GC content varies slightly among the plastomes of cycad genera in both coding (40.9–41.2%) and noncoding (36.7–37.9%) regions, as shown in supplementary table S1, Supplementary Material online. Comparative plastomes of cycads, including *Cycas*, reveal two minor organizational differences, including pseudogenization of *rpl23*, *chlB*, *chlL*, and *chlN* in *Stangeria* and absence of *trnT-GGU* from *Cycas*. Overall, these cycad plastomes appear to have a nearly identical organization (fig. 1; supplementary fig. S1, Supplementary Material online), reflecting the frozen plastomic organization in all cycad genera.

Furthermore, all of the eight newly sequenced cycad plastomes contain a residual elongation factor gene, *tufA*, which is flanked by the two protein-coding genes *petL* and *psbE* (fig. 1). Transfer of the plastid *tufA* to the nucleus was predicted to occur in the common ancestor of land plants (Baldauf and Palmer 1990). The retained *tufA* residual was first reported in the plastomes of *Cycas taitungensis*, *Ginkgo biloba*, and the hornwort, *Anthoceros formosae* (Wu et al. 2007). Common retention of the plastomic *tufA* residuals in all cycad genera provides additional evidence that the evolutionary stasis of cycad plastomes has occurred since the common ancestor of cycads, about 270 Ma.

Evolutionary Stasis in Nucleotide Substitution Rates of Cycad Plastomes

Figure 2 shows estimated synonymous (ds) and nonsynonymous (dn) substitution rates to be 0.623–0.649 (substitutions per ds sites) and 0.113–0.120 (substitutions per dn site), respectively, among the plastomes of all cycad genera. They are significantly slower than those of Pinaceae (ds : 0.708–0.731, Mann–Whitney two-sided $P < 0.01$; dn : 0.129–0.136, $P < 0.01$), cupressophytes (ds : 0.748–0.916, $P < 0.01$; dn : 0.14–0.615, $P < 0.01$), and gnetophytes (ds : 1.386–1.537, $P < 0.05$; dn : 0.194–0.2043, $P < 0.05$) but not *Ginkgo* (ds : 0.624, $P = 0.296$; dn : 0.119, $P = 0.29$). This suggests that the plastomic protein-coding sequences evolve slowly in all cycad genera.

In cycad plastomes, 264 syntenic loci were identified, including 136 coding and 128 noncoding loci (supplementary table S3, Supplementary Material online). Estimated sequence divergences of the 264 syntenic loci range from 0% to 11.33%, in which the intergenic spacer between *psbA* and *trnK* is the highest (supplementary table S3, Supplementary Material online), indicating its potential utility in discriminating cycad genera.

Substitution Rates of Noncoding Loci in IR Are Slower than in SC Regions

Figure 3 illustrates the substitution rates estimated from all syntenic noncoding loci in the SC and IR regions, respectively. For each of the eight cycad genera, the substitution rates are significantly slower in the IR than in the SC regions (Mann–Whitney two-sided, all $P < 0.01$). Overall, the rates of SC regions are 2.11- to 2.39- folds higher than those of the IR regions across all cycad genera.

Contrasting Mutational Biases between SC and IR Regions

A phylogenetic tree inferred from the 72 plastomic protein-coding genes common to the eight cycad genera was constructed. Except the node leading to *Zamia* and *Ceratozamia*, all nodes in this tree are supported with 100% posterior probabilities (supplementary fig. S2, Supplementary Material online). This tree was used to infer the ancestral sequences of the noncoding loci for the eight cycads. Then, the inferred ancestral sequences were utilized for counting point mutations (table 1). Supplementary figure S3, Supplementary Material online, illustrates the relative proportions of all six types of nucleotide mutations. In all examined cycads, frequencies of transitions are higher than those of transversions in both the SC and IR regions. However, in the IR regions, the frequency is higher for A-to-G and T-to-C transitions than G-to-A and C-to-T ones (Mann–Whitney two-sided $P < 0.01$). In contrast, the two types of transitions do not differ significantly in the SC regions (Mann–Whitney two-sided $P = 0.874$).

Moreover, contrasting biased mutations between the SC and IR regions become evident when the six types of nucleotide mutations are further divided into GC-rich (i.e., A-to-G or C and T-to-G or C) and AT-rich (i.e., G-to-A or T and C-to-A or T) mutations. In the SC regions, counts of GC-rich mutations are fewer than those of AT-rich ones in all cycads except *Dioon* and *Lepidozamia* (table 1). Notably, this bias is opposite in the IR regions. Figure 4 depicts the GC_{eq} estimated from noncoding loci of the SC and IR regions, respectively. In the SC regions, the GC_{eq} values for all cycads are remarkably smaller than equilibrium (i.e., $GC_{eq} = 50\%$), although both *Dioon* and *Lepidozamia* show more GC-rich than AT-rich mutations (table 1). Furthermore, figure 4 shows that in the IR

Table 1
Summary of Nucleotide Mutations in Noncoding Loci

Taxon	Reference ^a	SC		IR ^b	
		Number of AT-to-GC ^c	Number of GC-to-AT ^d	Number of AT-to-GC ^c	Number of GC-to-AT ^d
<i>Bowenia</i>	Node 3	395	431	64	38
<i>Ceratozamia</i>	Node 4	537	699	96	78
<i>Dioon</i>	Node 6	411	394	59	37
<i>Encephalartos</i>	Node 1	151	158	42	14
<i>Lepidozamia</i>	Node 1	226	161	36	16
<i>Macrozamia</i>	Node 2	293	349	47	21
<i>Stangeria</i>	Node 5	1,000	1,365	110	52
<i>Zamia</i>	Node 4	763	991	106	50

^aAncestral sequences corresponding to nodes of the tree in supplementary figure S2, Supplementary Material online, were used for counting nucleotide pair mutations.

^bOnly one IR was taken into consideration.

^cAll GC-rich (i.e., A-to-G, A-to-C, T-to-G, and T-to-C) and ^dall AT-rich (i.e., G-to-A, G-to-T, C-to-A, and C-to-T) mutations were pooled, respectively.

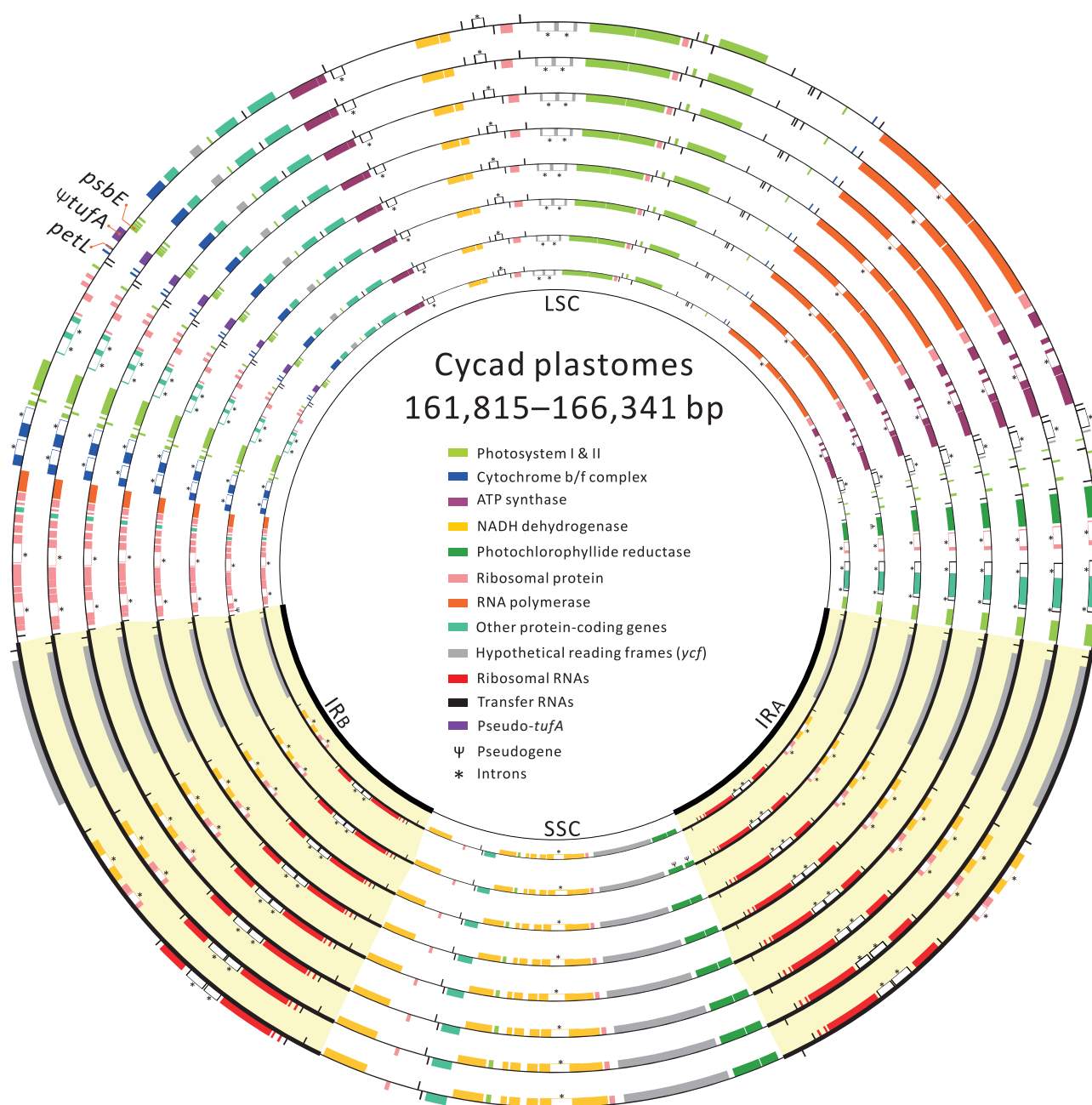


Fig. 1.—Comparisons of eight newly sequenced cycad plastome maps showing their conserved plastomic organizations. The circles from the outermost to the innermost are plastomes of *Bowenia*, *Ceratozamia*, *Dioon*, *Encephalartos*, *Lepidozamia*, *Macrozamia*, *Stangeria*, and *Zamia*. The two IR (IRA and IRB) regions separated by the LSC and SSC regions are highlighted with a yellow background. Genes are color-coded by their functions designated in the center of the maps. Note that a *ΨtufA* gene flanked by *psbE* and *petL* genes is commonly shared in all cycad genera, and that the flanking genes between IRs and LSC, and IRs and SSC are slightly different among genera.

regions, GC_{eq} values of all cycad genera are greater than 50%, except for *Ceratozamia* and *Stangeria* (with GC_{eq} values 47.1% and 46.1%, respectively). This result disagrees with the data shown in table 1, that the IR regions for all cycads have more GC-rich than AT-rich mutations. However, directly counting mutations did not take the effect of polarized base composition into account.

Particularly, the current GC content of noncoding loci are AT-rich in both the SC (GC content = 35.5–36.9%) and IR (41.7–42.0%) regions (fig. 4). As a result, the data shown in table 1 might not precisely reflect the mutational biases of noncoding loci. From the estimated GC_{eq} values (fig. 4), two different mutational trends are apparent: 1) toward AT-richness in SC regions and 2) toward GC-richness in IR regions.

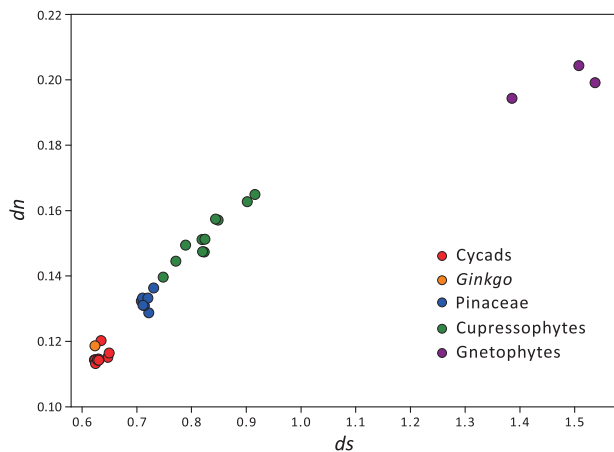


FIG. 2.—Comparison of synonymous (ds) and nonsynonymous (dn) substitution rates among the five major gymnosperm groups. The five groups are cycads, cupressophytes, ginkgo, gnetophytes, and Pinaceae. Each group is color-coded. Of note, all cycads feature slowly evolving substitution rates in their plastomic protein-coding sequences.

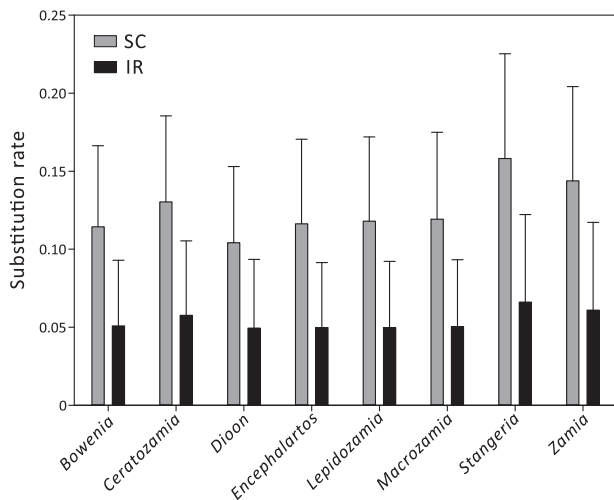


FIG. 3.—Comparison of substitution rates between the SC and IR regions. Substitution rates were estimated from noncoding loci of the eight cycad genera, with *Cycas* used as the reference. For each of the eight cycad genera, reduced substitution rates of the IR region are apparent. Data are mean \pm SD.

These findings strongly suggest that mutational biases are associated with the plastomic structure.

Discussion

Implications of Conserved Cycad Plastomes: Barcoding Utilities

Plastomic genes of cycads were previously reported as highly conserved (Rai et al. 2003; Wu et al. 2007). This study further reveals the conserved plastomic organization and slow

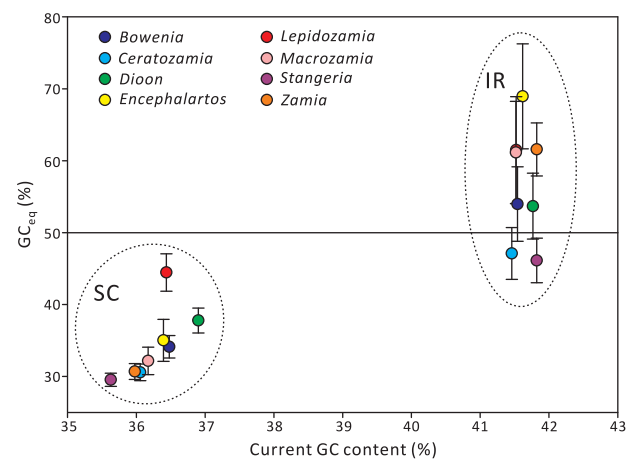


FIG. 4.—Comparison of GC_{eq} values between the SC and IR regions in the eight cycad genera. The GC_{eq} values were estimated from the non-coding loci. The solid horizontal line denotes the evolution of GC content under equilibrium ($GC_{eq}=50\%$). Note that in the SC regions, all GC_{eq} values are less than 50%, whereas in the IR regions, most GC_{eq} values are greater than 50%. Data are mean \pm SD from 100 bootstrapping analyses.

substitution rates across plastomes of all cycad genera (figs. 1 and 2; [supplementary fig. S1, Supplementary Material online](#)), which indicates that the morphology and plastomes of cycad genera changed little after cycad genera diverged in the early Jurassic (Nagalingum et al. 2011).

The slowly evolving cycad plastomes prompted us to question whether the commonly used DNA barcodes of land plants are appropriate for cycads. The two universal barcodes for land plants, *rbcL* and *matK* (CBOL Plant Working Group 2009), show only moderate sequence divergences among all syntenic loci of cycad plastomes ([supplementary table S3, Supplementary Material online](#)). The noncoding loci, such as *atpF-atpH*, *psbK-psbI*, and *psbA-trnH*, were previously examined for their ability to discriminate cycad genera/species (Sass et al. 2007; Nicolalde-Morejón et al. 2011). However, these three noncoding loci are lower in sequence divergence than our newly determined locus of *psbA-trnK*, which has the greatest sequence divergence ([supplementary table S3, Supplementary Material online](#)). Therefore, the noncoding locus *psbA-trnK* could be a powerful marker for identification of cycad genera and species.

A gBGC Model Interprets the Evolution of Cycad Plastomes

Wolfe et al. (1987) showed that substitution rates were relatively slower in IR than SC regions in plastomes of some angiosperms. Birky and Walsh (1992) proposed that copy-number-dependent variation of substitution rates was resulted from correction of mutations by gene conversion, which occurred more frequently in IR than SC regions. Gene

conversion in IR regions was supported by the study of Perry and Wolfe (2002), who found that substitution rates were accelerated in genes originally located in the IR regions of IR-lacking legume species. Therefore, in the plastomes of cycads, the reduced substitution rates of the IR regions—approximately 2-fold slower than those of the SC regions (fig. 3)—most likely result from gene conversion. Gene conversion is processed neutrally (Galtier et al. 2001). In this study, investigations of gene conversion using noncoding loci can avoid the effects of selection. For examples, the unusual codon-usage of *psbA* gene was proposed to be associated with selection for enhancement of translation efficiency (Morton 1993).

This study provides novel evidence that substitution rates of cycad plastomes are reduced and also GC-biased in their IR regions (figs. 3 and 4). Notably, we also discovered striking AT-biased mutations in the SC regions. This suggests that compositionally biased mutations are associated with copy numbers. As a result, in the IR regions of cycads, GC-biased mutations together with reduced substitution rates can best be explained by gBGC. gBGC, which prefers repairing DNA mismatches with GC bases in hetero-duplexed recombination intermediates (Marais 2003; Duret and Galtier 2009), is ubiquitous in eukaryotes (Pessia et al. 2012; Glémin et al. 2014) and prokaryotes (Lassalle et al. 2015). In addition, elevated GC content in the mitochondrial genome of *Polytomella capuana*,

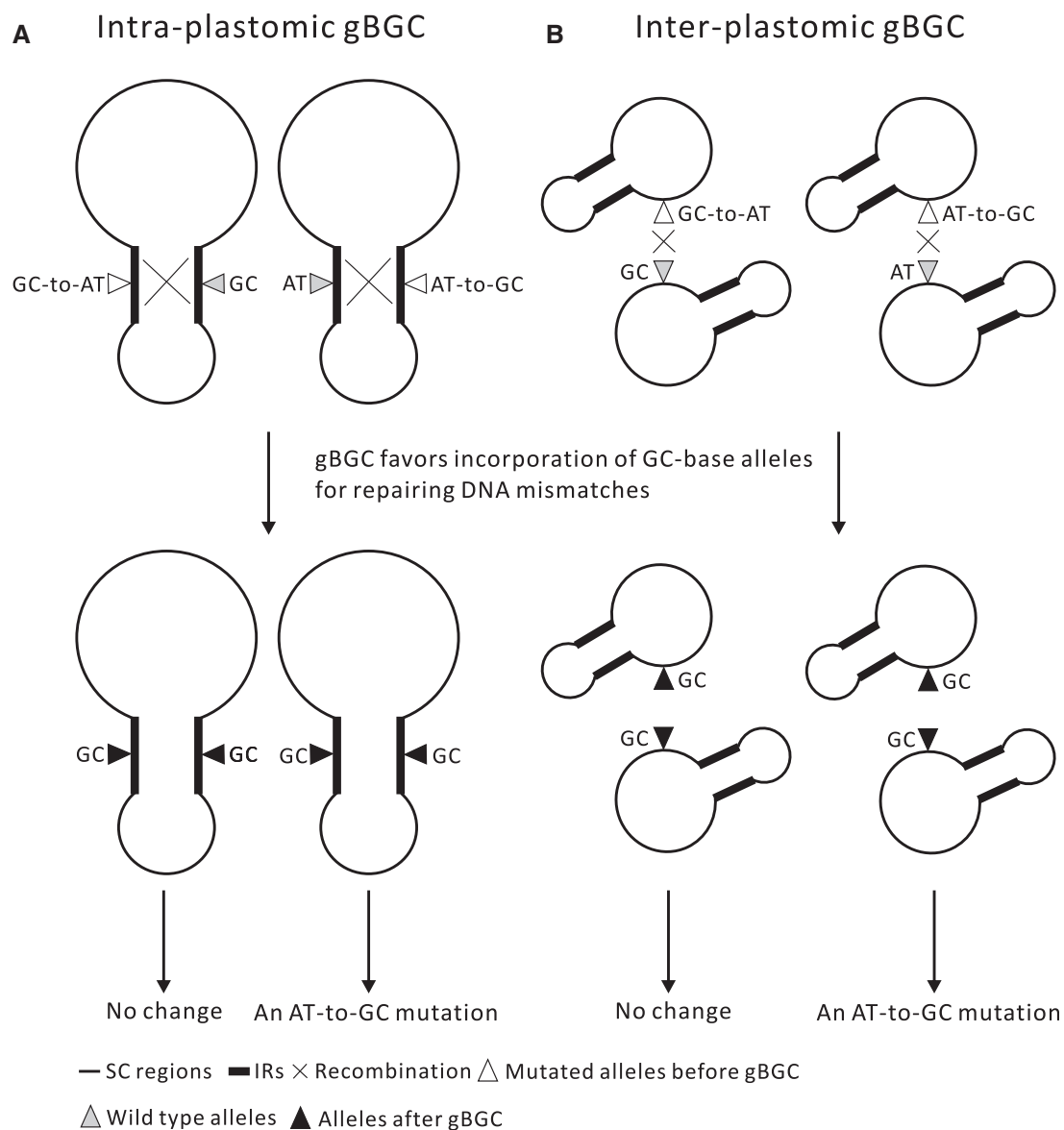


Fig. 5.—A gBGC model for the evolution of cycad plastomes. The fate of point mutations that are eliminated or retained after gBGC presumably depends on the type of mutations. This model interprets both the reduced substitution rates and GC-biased mutations observed in the IR regions of cycads.

a green alga, was suggested to be associated with gBGC (Smith and Lee 2008). Nevertheless, gBGC in plastomes has never been reported before. Here, we propose a gBGC model to interpret the reduced substitution rates and GC-biased mutations observed in the IR regions of cycads (fig. 5). Plastids contain multiple copies of plastomes, where two IR copies have identical sequences. Therefore, intraplasmidic recombination via IR regions (Palmer 1983) and an interplasmidic one via homologous sequences (Khakhlova and Bock 2006) may simultaneously occur in plastids (fig. 5).

Our model hypothesizes that point mutations, which altered base compositions (i.e., GC-to-AT or AT-to-GC), took place before the occurrence of gBGC. However, gBGC favors incorporation of GC bases for repairing DNA mismatches, regardless of GC bases being derived from wild-type or mutated alleles. Ultimately, GC-to-AT mutations are eliminated, whereas AT-to-GC ones are fixed in plastomes after gBGC (fig. 5), thus resulting in accumulation of GC-biased mutations. Of note, in our model, gBGC can rectify mutations or retain GC-biased mutations. This does not conflict with the correction of mutations by gene conversion in plastomes (Birky and Walsh 1992; Khakhlova and Bock 2006, and data herein). Although AT-biased gene conversion was previously proposed (Khakhlova and Bock 2006) basing on measuring the start codons of *ycf* genes, selective constraints of those protein-coding genes might be overwhelming. Consequently, neutral genetic processes, such as gBGC, might not be observed.

According to our model (fig. 5), gBGC might also occur in the SC regions via interplasmidic recombination. Because gBGC can correct some mutations, we expect a negative correlation between substitution rates and GC_{eq} values among the eight cycads. Indeed, a strong negative correlation between the substitution rates and GC_{eq} values was suggested for SC (Spearman's $r = -0.905$, $P < 0.01$) and IR ($r = -0.699$, $P < 0.05$) regions, respectively. These data clearly support that in cycads, gBGC has shaped plastome-wide mutations and GC content.

We are also interested in the reason behind the AT-rich noncoding loci in the IR regions of cycads (current GC content $< 50\%$; see fig. 4), despite having GC-biased mutations. The GC content of genomes results from a combination of complex mechanisms, including selection, biased gene conversion, and biased mutations (Hershberg and Petrov 2010; Hildebrand et al. 2010; Van Leuven and McCutcheon 2012). Although methylation of DNA is likely absent in plastomes (Fojtová et al. 2001; Ahlert et al. 2009), plastomic context-dependent mutations (i.e., base mutations toward transitions or transversions highly depend on their neighboring bases) was previously demonstrated (e.g., Morton 1995, 1997). Therefore, combination of gBGC and the plastomic context-dependent mutations might have affected the evolution of substitution rates and GC content in cycad plastomes.

Evolution of Substitution Rates in Cycads

Recent molecular dating suggested that the divergence of cycad genera has occurred since about 158.1 Ma, which is earlier than that of gnetophyte genera dated about 146.1 Ma (Lu et al. 2014). However, our data indicate that the *ds* substitution rates are significantly lower in cycads than in gnetophytes (fig. 2), which implies that different substitution rates among gymnosperm lineages are irrelevant with the divergence times of lineages. Determining that whether in cycads, the mitochondrial and nuclear genomes are also slowly evolving is of interest. In estimating substitution rates of some mitochondrial genes, Mower et al. (2007) concluded that cycads had the slowest substitution rates among all seed plants studied. Specific rate acceleration was documented in mitochondria of some seed plant lineages (Cho et al. 2004; Zhu et al. 2014). A correlation of substitution rates among mitochondrial, nuclear, and plastid genomes (Eyre-Walker and Gaut 1997; Drouin et al. 2008) or between organelle genomes (Sloan et al. 2012) was also suggested. However, substitution rates of nuclear genomes are poorly known in cycads. A nuclear genome-wide comparison between cycads and other gymnosperms is definitely needed to clarify whether substitution rates are homogeneously conserved across all three genomes of the cycads. If they are, a lineage effect, such as generation time, may lead to the best possible answer for the evolutionary stasis of cycad plastomes.

In conclusion, this study is the first to uncover the plastomic gBGC in seed plants. Cycad plastomes were observed to possess highly conserved organization. Their noncoding sequences constitute excellent systems for studying the evolution of neutral processes in plastomes. We propose a gBGC model to account for the dissimilar evolutionary patterns as well as the compositionally biased mutations in the SC and IR regions of cycad plastomes. Plastomes of conifers lack one IR (e.g., Lin et al. 2010; Wu et al. 2011; Guo et al. 2014; Hsu et al. 2014; Wu and Chaw 2014). Whether the retained IR copy of conifers has accelerated the substitution rates and relaxed GC-biased mutations is worthy of further investigations.

Supplementary Material

Supplementary tables S1–S3 and figures S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We thank Mr. Yi-Ming Chen for the gifts of two living potted plants, *Bowenia serrulata* and *Stangeria eriopus*, and Dr. Terrance W. Waters for the seeds of *Dioon spinulosum*. These three species grow beautifully in our green house. The authors appreciate the suggestive comments on an earlier version of this manuscript by Naruya Saitou, Carol Eunmi Lee,

Isheng Tsai, and Edi Sudiarto. The authors thank the two anonymous reviewers for their critical reading and comments. This work was supported by research grants from the Ministry of Science and Technology, Taiwan (MOST 103-2621-B-001-007-MY3), and the Investigator's Award of Academia Sinica to S.-M.C.

Literature Cited

- Ahlert D, Stegemann S, Kahlau S, Ruf S, Bock R. 2009. Insensitivity of chloroplast gene expression to DNA methylation. *Mol Genet Genomics*. 282:17–24.
- Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. *Science* 342:1241089.
- Baldauf SL, Palmer JD. 1990. Evolutionary transfer of the chloroplast *tufA* gene to the nucleus. *Nature* 344:262–265.
- Bendich AJ. 1987. Why do chloroplasts and mitochondria contain so many copies of their genome? *Bioessays* 6:279–282.
- Birky CW Jr, Walsh JB. 1992. Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. *Genetics* 130:677–683.
- Brenner ED, Stevenson DW, Twigg RW. 2003. Cycads: evolutionary innovations and the role of plant-derived neurotoxins. *Trends Plant Sci*. 8:446–452.
- Cai Z, et al. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J Mol Evol*. 67:696–704.
- CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proc Natl Acad Sci U S A*. 106:12794–12797.
- Chaw SM, Walters TW, Chang CC, Hu SH, Chen SH. 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast *matK* gene, *trnK* intron, and nuclear rDNA ITS region. *Mol Phylogenet Evol*. 37:214–234.
- Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc Natl Acad Sci U S A*. 101:17741–17746.
- Donaldson JS. 2003. Chapter 1: Introduction. In: Donaldson JS, editor. *Cycads status survey and conservation action plan*. Gland (Switzerland): IUCN. p. 1–2.
- 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.
- Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet*. 10:285–311.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Eyre-Walker A, Gaut BS. 1997. Correlated rates of synonymous site evolution across plant genomes. *Mol Biol Evol*. 14:455–460.
- Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Seattle: Department of Genome Sciences, University of Washington.
- Fojtová M, Kovarik A, Matyásek R. 2001. Cytosine methylation of plastid genome in higher plants Fact or artefact? *Plant Sci*. 160:585–593.
- Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159:907–911.
- Glémin S, Clément Y, David J, Ressayre A. 2014. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. *Trends Genet*. 30:263–270.
- Guo W, et al. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. *Genome Biol Evol*. 6:580–590.
- Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet*. 6:e1001115.
- Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet*. 6:e1001107.
- Hsu CY, Wu CS, Chaw SM. 2014. Ancient nuclear plastid DNA in the yew family (taxaceae). *Genome Biol Evol*. 6:2111–2121.
- Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J*. 46:85–94.
- Kumar RA, Oldenburg DJ, Bendich AJ. 2014. Changes in DNA damage, molecular integrity, and copy number for plastid DNA and mitochondrial DNA during maize development. *J Exp Bot*. 65:6425–6439.
- Kusumi J, Tachida H. 2005. Compositional properties of green-plant plastid genomes. *J Mol Evol*. 60:417–425.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
- Lassalle F, et al. 2015. GC-Content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genet*. 11:e1004941.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. 2010. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and sub-families. *Genome Biol Evol*. 2:504–517.
- Lu Y, Ran JH, Guo DM, Yang ZY, Wang XQ. 2014. Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. *PLoS One* 9:e107679.
- Maier RM, Neckermann K, Igloi GL, Kössel H. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol*. 251:614–628.
- Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. *Trends Genet*. 19:330–338.
- Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytol*. 186:299–317.
- Morton BR. 1993. Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on tRNA availability. *J Mol Evol*. 3:273–280.
- Morton BR. 1995. Neighboring base composition and transversion/transition bias in a comparison of rice and maize chloroplast noncoding regions. *Proc Natl Acad Sci U S A*. 92:9717–9721.
- Morton BR. 1997. The influence of neighboring base composition on substitutions in plant chloroplast coding sequences. *Mol Biol Evol*. 14:189–194.
- Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol Biol*. 7:135.
- Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat Res*. 1:2–9.
- Nagalingum NS, et al. 2011. Recent synchronous radiation of a living fossil. *Science* 334:796–799.
- Nicolalde-Morejón F, et al. 2011. A character-based approach in the Mexican cycads supports diverse multigene combinations for DNA barcoding. *Cladistics* 27:150–164.
- Palmer JD. 1983. Chloroplast DNA exists in two orientations. *Nature* 301:92–93.
- Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH. 2002. Evolutionary re-organisation of a large operon in adzuki bean chloroplast DNA caused by inverted repeat movement. *DNA Res*. 9:157–162.
- Perry AS, Wolfe KH. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol*. 55:501–508.
- Pessia E, et al. 2012. Evidence for widespread GC-biased gene conversion in eukaryotes. *Genome Biol Evol*. 4:675–682.

- Rai HS, O'Brien HE, Reeves PA, Olmstead RG, Graham SW. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Mol Phylogenet Evol.* 29:350–359.
- Renner SS. 2011. Living fossil younger than thought. *Science* 334:766–767.
- Sass C, Little DP, Stevenson DW, Specht CD. 2007. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS One* 2:e1154.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33:W686–689.
- Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol Evol.* 4:294–306.
- Smith DR. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Mol Biol.* 71:627–639.
- Smith DR. 2012. Updating our view of organelle genome nucleotide landscape. *Front Genet.* 3:175.
- Smith DR, Lee RW. 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. *Mol Biol Evol.* 25:487–496.
- Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Van Leuven JT, McCutcheon JP. 2012. An AT mutational bias in the tiny GC-rich endosymbiont genome of *Hodgkinia*. *Genome Biol Evol.* 4:24–27.
- Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A.* 84:9054–9058.
- Wu CS, Chaw SM. 2014. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnol J.* 12:344–353.
- Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM. 2011. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol Evol.* 3:1284–1295.
- Wu CS, Wang YN, Liu SM, Chaw SM. 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. *Mol Biol Evol.* 24:1366–1379.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255.
- Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol.* 30:1720–1728.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24: 1586–1591.
- Zhu A, Guo W, Jain K, Mower JP. 2014. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Mol Biol Evol.* 31:1228–1236..

Associate editor: Bill Martin