The Complete Chloroplast Genome of *Ginkgo biloba* Reveals the Mechanism of Inverted Repeat Contraction

Ching-Ping Lin†, Chung-Shien Wu, Ya-Yi Huang, and Shu-Miaw Chaw*

Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

*Corresponding author: E-mail: smchaw@sinica.edu.tw.
†These authors contributed equally to this work.

Accepted: 28 February 2012

Data deposition: The accession number of complete chloroplast genome of *Ginkgo biloba*: AB684440.

Abstract

We determined the complete chloroplast genome (cpDNA) of *Ginkgo biloba* (*common name: ginkgo*), the only relict of ginkgophytes from the Triassic Period. The cpDNA molecule of ginkgo is quadripartite and circular, with a length of 156,945 bp, which is 6,458 bp shorter than that of *Cycas taitungensis*. In ginkgo cpDNA, rpl23 becomes pseudo, only one copy of ycf2 is retained, and there are at least five editing sites. We propose that the retained ycf2 is a duplicate of the ancestral ycf2, and the ancestral one has been lost from the inverted repeat A (IRα). This loss event should have occurred and led to the contraction of IRs after ginkgos diverged from other gymnosperms. A novel cluster of three transfer RNA (tRNA) genes, trnY-AUA, trnC-ACA, and trnSeC-UCA, was predicted to be located between trnC-GCA and rpoB of the large single-copy region. Our phylogenetic analysis strongly suggests that the three predicted tRNA genes are duplicates of trnC-GCA. Interestingly, in ginkgo cpDNA, the loss of one ycf2 copy does not significantly elevate the synonymous rate (KS) of the retained copy, which disagrees with the view of Perry and Wolfe (2002) that one of the two-copy genes is subjected to elevated KS when its counterpart has been lost. We hypothesize that the loss of one ycf2 is likely recent, and therefore, the acquired KS of the retained copy is low. Our data reveal that ginkgo possesses several unique features that contribute to our understanding of the cpDNA evolution in seed plants.

Key words: *Ginkgo*, chloroplast genome, trnSeC, inverted repeat contraction, duplication.

Introduction

Ginkgo (*Ginkgo biloba* L.), also known as maidenhair tree, is a well-known living gymnosperm fossil with edible seeds, medicinal efficacy, and ornamental value (Pang et al. 1996). Fossil records suggest that during the late Mesozoic and early Tertiary era (ca. 120–60 Ma), the genus Ginkgo reached its highest species diversity and was widespread in the Northern Hemisphere (Gong et al. 2008). Today, ginkgo is the only living species left within the family Ginkgoaceae, and its natural habitat is restricted to small areas in China (Shen et al. 2005).

The presence of two large inverted repeats (IRs) is one of the most remarkable features in the chloroplast genomes (cpDNAs). In land plants, dynamic expansion/contraction of IRs has been previously reported in some lineages, such as Apioidae (Plunkett and Downie 2000), monocots (Wang et al. 2008), ferns (Wolf et al. 2010), and Pinaceae which have extremely reduced IRs (Lin et al. 2010). The fluctuating lengths of IRs contribute to increase/decrease of cpDNA sizes and can be utilized to address phylogeny but with the need of caution (Wolf et al. 2010).

Using gene mapping and cross-hybridization methods, Palmer and Stein (1986) constructed the first cpDNA map of ginkgo and reported its IR length of approximately 17 kb. Apparently, the IR of ginkgo is significantly shorter than those of most angiosperms (ca. 20–28 kb, Chumley et al. 2006) and *Cycas taitungensis* (ca. 25 kb, Wu et al. 2007), which indicates that ginkgo has experienced an IR contraction. However, the mechanism of IR contraction in ginkgo remains unclear.

Therefore, this study aimed to 1) elucidate the cpDNA organization of ginkgo with reference to other gymnosperms and 2) expand our understanding of cpDNA diversity and evolution as part of our long-term gymnosperm cpDNA
evolutionary study. Here, we report several unique characteristics of ginkgo cpDNA, propose the underlying mechanism of its IR contraction, and discuss the evolution of an unusual transfer RNA (tRNA) gene cluster.

Materials and Methods

Genomic DNA Extraction

Genomic DNA (gDNA) was extracted from fresh young leaves of a ginkgo plant in the greenhouse of Academia Sincica by use of a CTAB-based protocol (Stewart and Via 1993). The purity and integrity of the extracted gDNA were measured and judged by the OD 260/280 ratio and gel electrophoresis, respectively. The gDNA with a 260/280 ratio greater than 1.8 was collected for polymerase chain reaction (PCR) experiments.

Amplification and Sequencing

The cpDNA fragments were amplified using a long-range PCR method with LA Taq (Takara Bio Inc., Shiga, Japan) and specific primers (supplementary table 1, Supplementary Material online). Amplicons were purified (260/280 ratio = 1.8–2.0; 260/230 ratio > 2), then sequenced by use of an Illumina GA IIx sequencer (YOURGENE BIO SCIENCE Co., New Taipei City, Taiwan). We trimmed short reads (73 bp) of paired-end sequencing using CLC Genomic Workbench 4.9 (CLC Bio, Aarhus, Denmark) with an error probability <0.05 and then assembled these trimmed reads in the same software without any reference information. Regions with <200× coverage depth were trimmed off manually, and these trimmed regions were considered as gaps. Finally, the average coverage depth of contigs is approximately 2080×, which is greatly larger than the proposed minimum
coverage depth for cpDNAs (30×) (Straub et al. 2012). All gaps between contigs were filled with sequences of specific PCR products.

Annotation
We used DOGMA (Wyman et al. 2004) and BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to annotate protein coding, ribosomal RNA (rRNA), and tRNA genes. All tRNA genes were further verified by their structures predicted by tRNAscan-SE 1.21 (Schattner et al. 2005).

Examination of RNA-Editing Sites
RNA extraction and reverse transcription polymerase chain reaction experiments involved use of the Plant Total RNA MiniPrep Purification Kit (Gene Mark Co., Taiwan) and the ReverTra Aid First Strand cDNA Synthesis Kit (Fermentas Inc., Glen Burnie, MD), respectively. The obtained cDNAs were used as PCR templates for examining specific RNA-editing sites.

Estimation of Synonymous Rates (Ks)
Ks values of genes were estimated by use of PAL2NAL 1.3 (Suyama et al. 2006). Amborella genes were used as the reference.

Phylogenetic Analyses
Thirty-five tRNA sequences of ginkgo cpDNA were aligned by use of ClustalW (Thompson et al. 1994). The aligned sequences were used to construct a maximum-likelihood (ML) tree with a General time reversible + Gamma + Proportion Invariant (GTR+G+I) model and 1,000 bootstrapping analyses in MEGA 5 (Tamura et al. 2011).

Results and Discussion

Characteristics of Ginkgo cpDNA
The cpDNA of ginkgo (Accession number: AB684440) is a circular molecule of 156,945 bp with a pair of IRs separated by large single-copy (LSC) and small single-copy regions (fig. 1), which agrees well with the restriction mapping of Palmer and Stein (1986), although the total lengths slightly differ. We found that the shortened IR previously noted by Palmer and Stein (1986) is due to the complete loss of the ycf2 from the IRA. We identified 120 unique genes in ginkgo cpDNA: 81 protein-coding genes, 35 tRNA genes, and 4 rRNA genes. A total of 14 genes are duplicated, including three protein-coding genes, six tRNA genes, and four rRNA genes in the IR, as well as one tRNA gene in the LSC region. Thirteen protein-coding genes and eight tRNA genes have introns. The overall AT content is 60.4% (protein-coding genes, 61.1%; tRNA genes, 46.1%; rRNA genes, 44.7%; introns, 60.2%; intergenic spacers, 63.2%). We detected five C-to-U RNA-editing sites and experimentally verified them at the initial codons of petL and rps8 and the terminal codons of petL, rps4, and ndhC.

Table 1
Comparison of cpDNA Features between Ginkgo and Cycas

| Features                     | Ginkgo biloba | Cycas taitungensis |
|------------------------------|---------------|--------------------|
| Size (bp)                    | 156,945       | 163,403            |
| LSC                          | 99,221        | 90,216             |
| SSC                          | 22,258        | 23,039             |
| IR                           | 17,733        | 25,074             |
| % AT content                 | 60.4          | 60.5               |
| % Coding genes               | 54.6          | 57.2               |
| RNA-editing sitesa           | 5             | 37b (Chen et al. 2011) |
| Total number of genes        | 134           | 133                |
| Protein-coding genes         | 84            | 87                 |
| Duplicated genes             | 14            | 15                 |
| tRNA genes                   | 42            | 38                 |
| rRNA genes                   | 8             | 8                  |
| Genes with introns           | 21            | 21                 |

Comparisons of Ginkgo and a Cycad cpDNAs
Because cycads and ginkgo are the two most ancient lineages of gymnosperms, we compared their cpDNA features. The cpDNA organizations of both ginkgo and C. taitungensis are similar (table 1), except that ginkgo has only a single copy (SC) of ycf2 and its rpl23 has become pseudo, and Cycas lost the trnT-GGU originally located between psbD and trnL-UUC in the LSC region (fig. 1 in Wu et al. 2007). These events led to a downsizing of ginkgo cpDNA. In addition, ginkgo cpDNA contains a specific cluster of three novel tRNA genes (trnSeC-UCA, trnC-ACA, and trnY-AUA) that are located between the rpoB and the trnC-GCA of the LSC region (fig. 1).

Pseudogenization of rpl23
In addition to the dysfunctional tufA reported by Wu et al. (2007), rpl23, which is retained in many land plants and is near the junction of IRA and LSC regions, becomes pseudo in ginkgo (viz. Ψrpl23). In gymnosperms, loss of rpl23 was previously reported in gnetophyte cpDNAs (Wu et al. 2009). The Ψrpl23 of ginkgo has a truncated 5′ region as compared with the functional rpl23 of Cycas. These data suggest that rpl23 was independently lost from these two gymnosperm lineages.

trnH-GUG as an Evolutionary Footprint Caused by IR Contraction
As mentioned previously, loss of an ycf2 copy downsized the IRs and the cpDNA of ginkgo. Therefore, two questions are raised: whether the IRA of ginkgo originally had a syntenic ycf2 copy, as is found in most seed plants, and if the ginkgo did lose the ycf2 copy from the ancestral IRA, to what extent the retained ycf2 copy evolved. To answer these questions, we first compared the boundaries of IRs among two
**FIG. 2.**—(A) Comparison of IR boundaries among two representative ferns (*Psilotum nudum* and *Angiopteris evecta*), gymnosperms (*Cycas taitungensis*, *Ginkgo biloba*, and *Gnetum parvifolium*), and angiosperms (*Amborella trichopoda* and *Nicotiana tabacum*). (B) Hypothetical scenario illustrating IR contraction in ginkgo cpDNA. The ancestral IRs of gymnosperms should have expanded to include a \textit{trnH-GUG}, and then \textit{ycf2} was lost from the ancestral IRA during ginkgo evolution. The evolutionary footprints, two \textit{trnH-GUG}, and IR contraction are indicated. The tree topology was modified from Wu et al. (2011). *: genes with introns. \(J_{\text{LA}}\): junction between LSC and IRA; \(J_{\text{LB}}\): junction between LSC and IRB regions.

**Mechanism of Inverted Repeat Contraction**

*Genome Biol. Evol.* 4(3):374–381. doi:10.1093/gbe/evs021 Advance Access publication March 8, 2012 377
representatives of ferns and five representatives of IR-containing seed plants, including Cycas (representative of cycads), ginkgo, Gnetum (representative of gnetophytes), Amborella (representative of basal angiosperms), and Nicotiana (representative of eudicots) (fig. 2A).

Because the IRs of leptosporangiate ferns (e.g., Adiantum and Alsophila) independently expanded to encompass rps7, 3′ rps12, and ycf2 (Wicke et al. 2011), two eusporangiate ferns (Psilotum and Angiopteris) that retain ancestral cpDNA organizations (Gao et al. 2011) were included to simplify evolutionary inferences. As compared with the IR boundaries of ferns, those of all seed plants, except ginkgo, expanded to include ycf2 sequences (fig. 2A). This indicates that duplication of ycf2 is a common trait among the cpDNAs of seed plants. Wu et al. (2007) proposed that the ycf2 of IR was duplicated from that of IRα. Of note, IRs of both cycads and gnetophytes retain a trnH-GUG. In contrast, this tRNA gene is absent from the IRs of both ferns and angiosperms, which suggests that duplication of trnH-GUG is gymnosperm specific. Because the cpDNA of ginkgo has two respective trnH-GUG sequences near its IR boundaries, each of the ancestral IRs of ginkgo should have expanded to include a trnH-GUG sequence, and subsequently the IRs were contracted by loss of at least the ycf2 sequence from the IRα (fig. 2B). As a result, the trnH-GUG that adjoins the current IRα could be considered an evolutionary footprint due to the contraction of the ancestral IRα.

**IR Contraction Has No Effect on the Substitution Rate of the Retained ycf2**

Perry and Wolfe (2002) discovered that in IR-containing lineages, the synonymous rates (Ks) of IR genes are 2.3-fold lower than those of SC genes, whereas in IR-lacking lineages, the mean Ks of formerly IR-residing genes are 1.3-fold higher than those of the remaining genes. The authors concluded that in IR-lacking cpDNAs, decreased copy number rather than intrinsic properties directly elevates the Ks of genes formerly residing in IRs.

With the conclusion of Perry and Wolfe (2002), one should expect an accelerated Ks in the retained ycf2 of ginkgo cpDNA. Figure 3 shows comparisons of the ycf2 and the rest of the IR genes among seven available IR-containing gymnosperm cpDNAs. The Ks values are largely variable among lineages, with the highest in Ephedra of the gnetophytes. To exclude the lineage effect, the Ks of ycf2 was divided by the mean Ks of the rest of the IR genes in respective lineages (the obtained ratios for Cycas 2.74; Bowenia 3.24; Zamia 2.79; Ginkgo 3.38; Ephedra 3.03; Welwitschia 3.79; Gnetum 3.57). Two-tailed Z-test results revealed no difference between ratios for ginkgo and other gymnosperms (P = 0.29). Therefore, in ginkgo cpDNA, the event of losing an ycf2 copy is likely recent, and the retained copy accumulates few mutations.

**Duplications of trnC-GCA Occurred at Least Twice**

We detected three adjacent tRNA genes (trnY-AUA, trnC-ACA, and trnSeC-UCA) in the same orientation, as well as trnC-GCA (a syntenic tRNA gene of all land plant cpDNAs), in the region between petN and rpoB of the LSC in ginkgo. We exclude the possibility that the three clustered tRNA genes derived from horizontal transfers because of no DNA-importing system in chloroplasts (Smith 2011). Intriguingly, these three clustered tRNA genes have high sequence similarity with trnC-GCA (fig. 4A): the sequence similarities between trnC-GCA and trnY-AUA, trnC-ACA, and trnSeC-UCA are 82.1%, 85.1%, and 80.8%, respectively. The ML tree depicted in figure 4B shows that almost all synonymous tRNA species are clustered with each other or to one another and that trnY-AUA, trnC-ACA, trnSeC-UCA, and trnC-GCA are grouped as a monophyletic clade (bootstrap value = 70%), in which trnC-ACA and trnSeC-UCA form a subclade (bootstrap value = 64%). This result suggests that the three clustered tRNAs are duplicates of trnC-GCA, and they might derive from at least two duplication events. The tandem
Mechanism of Inverted Repeat Contraction

**Fig. 4.**—Phylogenetic relationships of the three clustered tRNA genes uniquely found in ginkgo cpDNA. (A) Alignment of trnC-GCA, trnY-AUA, trnC-ACA, and trnSeC-UCA sequences. (B) Unrooted ML tree based on all 35 tRNA genes encoding in ginkgo cpDNA. The trnC-GCA and the three clustered tRNA genes are in yellow shadow. Tree branches leading to synonymous tRNA species have the same colors. Values along branches denote bootstrapping values estimated from 1,000 replicates (only values ≥50% are shown).
duplicated trnF-GAA copies found in the cpDNAs of Brassicaceae were characterized by several parallel gains and losses (Koch et al. 2005). However, the duplicated tRNA genes that we reported here may not be specific to ginkgo or inherited from the common ancestor of ginkgophytes because cpDNAs of extinct ginkgo lineages, for example, G. adiantoides and G. yimaensis (Zhou and Zheng 2003), are unavailable. Interestingly, trnSeC-UCA was also annotated in the cpDNA of Adiantum (Wolf et al. 2003), a leptosporangiate fern, but it is not syntenic with that of ginkgo cpDNA. Gao et al. (2009) proposed that in the Adiantum cpDNA, trnR-CCG was substituted by trnSeC-UCA because the former is not essential. In contrast, in ginkgo cpDNA, trnC-GCA coexists rather than replaced by its duplicates, possibly because trnC-GCA is vital for plant cell development (Legen et al. 2007). In addition, whether trnSeC-UCA of ginkgo cpDNA is functional and what is its evolutionary significance require further scrutiny.

Conclusions

We elucidated that the shortened IR of ginkgo cpDNA is a consequence of IR contraction, and the contraction mainly resulted from loss of one ycf2 copy from the IRα. The presence of two trnH-GUG, one near the junction of LSC-IRα and the other upstream of ycf2, are considered as footprints of IR contraction. Unexpectedly, the Ks of the retained ycf2 copy is nonacelerated, which suggests that the loss might be recent in ginkgo evolution. Moreover, we found a unique cluster of three tRNA genes upstream of trnC-GCA in ginkgo cpDNA. The duplicated relationships between the three clustered tRNA genes and trnC-GCA are evident on the basis of their high sequence similarity and phylogenetic evaluation. However, the evolutionary impact of this tRNA gene cluster needs further investigation.

Supplementary Material

Supplementary table 1 is available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We thank the two anonymous reviewers for providing critical comments and valuable suggestions. This work was supported by a research grant from the Biodiversity Research Center, Academia Sinica, to S.M.C.

Literature Cited

Chen H, Deng L, Jiang Y, Lu P, Yu J. 2011. RNA editing sites exist in protein-coding genes in the chloroplast genome of Cycas taitungensis. J Integr Plant Biol. 53:961–970.

Chumley TW, et al. 2006. The complete chloroplast genome sequence of Pelargonium × hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 23:2175–2190.

Drescher A, Ruf S, Cals T Jr, Carrer H, Bock R. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. Plant J. 22:97–104.

Gao L, Yi X, Yang YX, Su YJ, Wang T. 2009. Complete chloroplast genome sequence of a tree fern Alsophila spinulosa: insights into evolutionary changes in fern chloroplast genomes. BMC Evol Biol. 9:130.

Gao L, Zhou Y, Wang ZW, Su YJ, Wang T. 2011. Evolution of the rpoB-psbZ region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. BMC Plant Biol. 11:64.

Gong W, Chen C, Dobes C, Fu CX, Koch MA. 2008. Phylogeography of a living fossil: pleistocene glaciations forced Ginkgo biloba L. (Ginkgoaceae) into two refuge areas in China with limited subsequent postglacial expansion. Mol Phylogenet Evol. 48:1094–1105.

Koch MA, et al. 2005. Evolution of the trnR(GAA) gene in Arabidopsis relatives and the brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. Mol Biol Evol. 22:1032–1043.

Legen J, Wanner G, Herrmann RG, Small I, Schmitz-Linneweber C. 2007. Plastid tRNA genes trnC-GCA and trnV-GUU are essential for plant cell development. Plant J. 51:751–762.

Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. 2010. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biol Evol. 2:504–517.

Palmer JD, Stein DB. 1986. Conservation of chloroplast genome structure among vascular plants. Curr Genet. 10:823–833.

Pang Z, Pan F, He S. 1996. Ginkgo biloba L: history, current status, and future prospects. J Altern Complem Med. 2:359–363.

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.

Plunkett GM, Downie SR. 2000. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Syst Bot. 25:648–667.

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–W689.

Shen L, et al. 2005. Genetic variation of Ginkgo biloba L. (Ginkgoaceae) based on cpDNA PCR-RFLPs: inference of glacial refugia. Heredity 94:396–401.

Smith DR. 2011. Extending the limited transfer window hypothesis to inter-organellar DNA migration. Genome Biol Evol. 3:743–748.

Stewart CN Jr, Via LE. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14:748–751.

Straub SC, et al. 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Ann Bot 99:1–16.

Suyama M, Torres D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.

Tamura K, et al. 2011. MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Wang RJ, et al. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol. 8:36.

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76:273–297.

Wolf PG, Roper JM, Duffy AM. 2010. The evolution of chloroplast genome structure in ferns. Genome 53:731–738.

Wolf PG, Rowe CA, Sinclair RB, Hasebe M. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, Adiantum capillus-veneris L. DNA Res. 10:59–65.

Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. 2009. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. Mol Phylogenet Evol. 52:115–124.

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 organelar genomes with DOGMA. Bioinformatics 20:3252–3255.

Zhou Z, Zheng S. 2003. The missing link in Ginkgo evolution. Nature 423:821–822.

Associate editor: Bill Martin