# Transfer of Chloroplast Genomic DNA to Mitochondrial Genome Occurred At Least 300 MYA

*Daryi Wang,* Yu-Wei Wu,† Arthur Chun-Chieh Shih,† Chung-Shien Wu,*‡ Ya-Nan Wang,‡ and Shu-Miaw Chaw**

*Research Center for Biodiversity; †Institute of Information Science, Academia Sinica, Taipei 115, Taiwan; and ‡School of Forestry and Resource Conservation, National Taiwan University, Taipei 106, Taiwan

With the completion of the first gymnosperm mitochondrial genome (mtDNA) from *Cycas taitungensis* and the availability of more mtDNA taxa in the past 5 years, we have conducted a systematic analysis of DNA transfer from chloroplast genomes (cpDNAs) to mtDNAs (mtpts) in 11 plants, including 2 algae, 1 liverwort, 1 moss, 1 gymnosperm, 3 monocots, and 3 eudicots. By using shared gene order and boundaries between different mtpts as the criterion, the timing of cpDNA transfer during plant evolution was estimated from the phylogenetic tree reconstructed independently from concatenated protein-coding genes of 11 available mtDNAs. Several interesting findings emerged. First, frequent DNA transfer from cpDNA to mtDNA occurred at least as far back as the common ancestor of extant gymnosperms and angiosperms, about 300 MYA. The oldest mtpt is *trnV*(*uac*)-*trnM*(*cau*)-*atpE-atpB-rbcL*. Three other mtpts—*psaA-psaB*, *rps19-trnH*(*gug*)-*rpl2-rpl23*, and *psbE-psbF*—were dated to the common ancestor of extant angiosperms, at least 150 MYA. However, all protein-coding genes of mtpts have degenerated since their first transfer. Therefore, mtpts contribute nothing to the functioning of mtDNA but junk sequences. We discovered that the cpDNA transfers have occurred randomly at any positions of the cpDNAs. We provide strong evidence that the cp-derived tRNA-*trnM*(*cau*) is the only mtpt (1 out of 3 cp-derived tRNA shared by seed plants) truly transferred from cpDNA to mtDNA since the time of the common ancestor of extant gymnosperms and angiosperms. Our observations support the proposition of Richly and Leister (2004) that "primary insertions of organellar DNAs are large and then diverge and fragment over evolutionary time."

## Introduction

The endosymbiotic theory proposes that chloroplasts (cp) and mitochondria (mt) were once free-living cyanobacteria and α-proteobacteria, respectively, and that both bacteria became organelles of eukaryotes through a cooperative relationship (Gray et al. 2001; Martin et al. 2002). It has long been known that some large DNA segments in the mtDNA have been independently transferred into the nuclear genomes (nrDNA) (Martin and Herrmann 1998; Timmis et al. 2004; Noutsos et al. 2005) through evolution. For example, in *Arabidopsis thaliana*, a 620-kb mtDNA in origin was detected in its nrDNA (Lin et al. 1999), and in rice, 2 large cpDNA segments, 33 and 131 kb in length, were found in its nrDNA (Yuan et al. 2002). It has been hypothesized that transferred DNA sequences were likely obtained through nonhomologous recombination with DNA fragments that leaked out of cp (Timmis et al. 2004). These DNA transfers and/or integration of entire organellar genomes to nuclear ones were suggested to be part of an ongoing evolutionary process that has influenced the evolution of eukaryotic nrDNAs (Richly and Leister 2004; Huang et al. 2005; Yamauchi 2005).

DNA transfers in plants have been documented, not only from organelles to nucleus but also from cp to mt (Kubo et al. 2000; Notsu et al. 2002; Handa 2003; Turmel et al. 2003; Clifton et al. 2004; Ogihara et al. 2005; Sugiyama et al. 2005). These cp-derived mtDNAs (also termed mitochondrial plastid DNAs, here abbreviated as mtpts) have various natures: for example, some protein-coding sequences of mtpts are pseudogenes (Cummings et al. 2003; Clifton et al. 2004) and some tRNA sequences of mtpts are nonfunctional (Notsu et al. 2002; Clifton et al. 2004). Cummings et al. (2003), Clifton et al. (2004), and Woloszynska et al. (2004) have linked the occurrence of mtpts with the phylogeny of flowering plants. These authors suggest that the transfer of mtpts occurred independently among species after speciation. However, because mtpts of flowering plants have higher substitutional and recombinational rates (Watanabe et al. 1994; Kanno et al. 1997), we consider estimating the origin of the mtpt transfer based solely on sequence similarity highly unreliable. Moreover, some characteristics of mtpt sequences were also noted to have caused bias in phylogenetic analysis. For instance, in maize DNA copy correction has yielded a high identity rate (96%) between mtDNA sequences and their cpDNA counterparts (Clifton et al. 2004).

Although Fauron et al. (1995) claimed that sequence information (e.g., DNA identity) used to date the transfer of mtpts has been lost due to the high frequency of sequence restructuring in mtDNA, they reported a single mtpt transfer event in the common ancestor of rice and maize. This event was inferred from several nonoverlapping but closely adjacent mtDNA segments of the 2 species. Studying the structural dynamics of mtDNAs in 3 cereals (wheat, rice, and maize), Ogihara et al. (2005) found that 2-gene clusters had been conserved among the 3 cereals' mtDNAs. This indicates that the relative positions of gene clusters can be maintained during evolution, and one can trace the first common occurrence. For this reason, we believe inferences about the transfer of mtpt based on gene order to be more robust than those based on phylogenetic analysis of mtpt sequences, as exemplified by the study of Clifton et al. (2004).

In this study, we estimate the date of an mtpt's appearance. First, the gene boundaries of mtpts among 11 plants

Key words: chloroplast genome, mitochondrial genome, mtpt, gene transfer, seed plants.

E-mail: smchaw@sinica.edu.tw.

*Mol. Biol. Evol.* 24(9):2040–2048. 2007
doi:10.1093/molbev/msm133
Advance Access publication July 3, 2007

**Table 1**
**CpDNAs and mtDNAs of the 11 Plants Examined in This Study**

| Classification | Scientific Name | Abbreviation | Common Name | mtDNA Acc. no. | Reference | cpDNA Acc. no. | Reference |
|---|---|---|---|---|---|---|---|
| Algae | | | | | | | |
| Chlorophyta | | | | | | | |
| | *Chlamydomonas reinhardtii* | Cre | Green algae | NC_001638 | Vahrenholz et al. (1993) | NC_005353 | Maul et al. (2002) |
| Streptophyta | | | | | | | |
| | *Chaetosphaeridium globosum* | Cgl | Charophyte | NC_004118 | Turmel et al. (2002) | NC_004115 | Turmel et al. (2002) |
| Land plants | | | | | | | |
| Bryophyta | | | | | | | |
| | *Physcomitrella patens* | Ppa | Moss | NC_007945 | Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwara M, Sato N, unpublished data | AP005672 | Sugiura et al. (2003) |
| | *Marchantia polymorpha* | Mpo | Liverwort | NC_001660 | Oda et al. (1992) | NC_001319 | Shimada and Sugiura (1991) |
| Seed plants | | | | | | | |
| Gymnosperm | | | | | | | |
| | *Cycas taitungensis* | Cta | Taitung cycad | Acc. no. XX | Chaw SM, et al. unpublished data | AB284317 | Wu et al. (2007) |
| Angiosperm | | | | | | | |
| Monocots | | | | | | | |
| | *Triticum aestivum* | Tae | Wheat | NC_007579 | Ogihara et al. (2005) | NC_002762 | Ogihara et al. (2002) |
| | *Oryza sativa sp. Japonica* | Osa | Rice | BA000029 | Notsu et al. (2002) | NC_001320 | Hiratsuka et al. (1989) |
| | *Zea mays* | Zma | Maize | NC_007982 | Clifton et al. (2004) | NC_001666 | Maier et al. (1995) |
| Dicots | | | | | | | |
| | *Nicotiana tabacum* | Nta | Tobacco | NC_006581 | Sugiyama et al. (2005) | NC_001879 | Kunnimalaiyaan and Nielsen (1997) |
| | *Brassica napus* | Bna | Oilseed rape | AP006444 | Handa (2003) | | |
| | *Arabidopsis thaliana* | Ath | Mouse-ear cress | NC_001284 | Unseld et al. (1997) | NC_000932 | Sato et al. (1999) |

with known cpDNA and mtDNA pairs (table 1) were compared. We then mapped the mtpts onto the independently derived land plant topology. If the mtpt boundaries were shared by more than one clade/branch, we then inferred that the occurrence date of such a shared mtpt went at least as far back as the date of the common node of the clades/branches. Our comparison of gene boundary is founded on the assumption that 2 or more mtpts independently possessing the same gene boundary is extremely rare. The conserved nature (or synteny) of gene order has also been used to detect mtDNA arrangements in metazoans (Fritzsch et al. 2006) and to deduce metazoan phylogeny (Boore 1999; Lavrov and Lang 2005). Although plant mtDNAs are known to have higher recombination rates (Watanabe et al. 1994; Kanno et al. 1997), we believe that gene orders in some mtpts might have maintained over the evolutionary history of plants and that theoretically they are retrievable by the use of a bioinformatics method.

More than 2 decades have passed since the discovery of the first mtpts in maize NB genome (Stern and Lonsdale 1982), and yet no systematic study to analyze their origin and evolution has been conducted, and several questions remain to be answered. In this study, we address 3 questions: (1) When, in terms of diversification of major plant lineages, did the first mtpt transfer occur? (2) How have the mtpts shaped the evolution of mtDNA or adapted to the mtDNA environments? And (3) what are the evolutionary fates of mtpts? To gain a better understanding of mtpt evolution, we have analyzed 11 plant species with known cpDNA and mtDNA pairs, including 2 algae, 1 liverwort, 1 moss, 1 recently sequenced cycad (cpDNA: Wu et al. 2007; mtDNA: Chaw SM, Shih AC, Wu YW, Liu SM, Chow TY, unpublished data), 3 monocots, and 3 eudicots. Our results are striking. First, 7 mtpt transferred scenarios were identified, including one taking place at the common ancestor of extant seed plants, 3 at the common ancestor of extant angiosperms, 2 within the monocots, and one in Brassicaceae. Second, our novel evidence suggested that only 1 of the 3 cp-derived tRNAs shared by the seed plants was actually transferred since the period of the common ancestor of the extant seed plants. Finally, the first scenario of mtpts occurred at least concurrently with the common ancestor of the extant seed plants or before the divergence of gymnosperms and angiosperms.

## Materials and Methods

### Collection of cpDNAs and mtDNAs

Species containing both complete mtDNAs and cpDNAs were selected. They are *Chlamydomonas reinhardtii*, *Chaetosphaeridium globosum*, *Physcomitrella patens*, *Marchantia polymorpha*, *Zea mays*, *Triticum aestivum*, *Oryza sativa*, *Nicotiana tabacum*, *A. thaliana* (table 1), and *Cycas taitungensis* (Chaw SM et al. unpublished data). The corresponding mtDNAs and cpDNAs were retrieved from National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). The

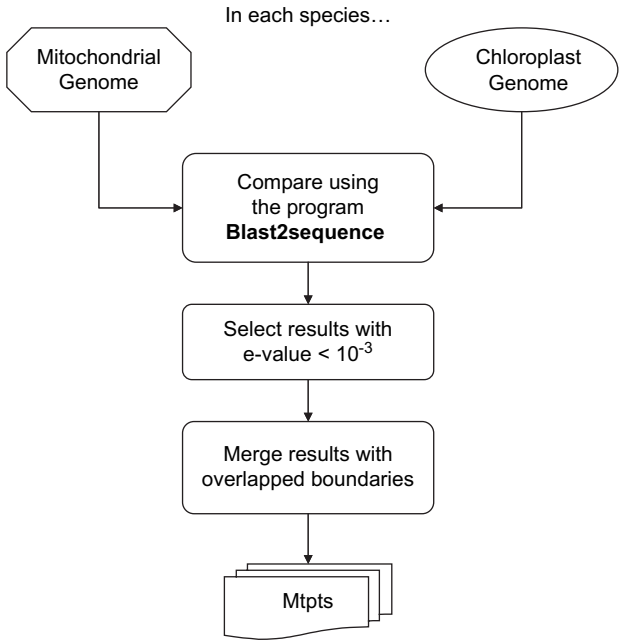


FIG. 1.—A flowchart for the collection of mtpts. At first, the mtDNA and cpDNA of each species were compared using the program Blast2sequence. Then the results with *e*-value smaller than $10^{-3}$ were selected. Finally, overlapped sequences from a specific taxon were combined and considered as the mtpts in its mtDNA.

mtDNA of *Brassica napus* was also retrieved and compared with the cpDNA of *A. thaliana* because they are classified in the same family, Brassicaceae.

## Reconstruction of Phylogenetic Trees

Twenty-one genes common to the selected mtDNAs were downloaded. They were *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad9*, *mttB*, *rps3*, *rps4*, and *rps12*. However, because only 9, i.e., *nad1*, *nad2*, *nad4*, *nad5*, *nad6*, *cob*, *cox1*, *rrnS3* (possibly *rps3* in other species), *rrns4* (possibly *rps4* in other species), of them were annotated in the *Chlamydomonas* mtDNA, this species was not included in the reconstruction of phylogenetic tree. The 21 genes were aligned separately by use of the program MUSCLE (Edgar 2004), and then the aligned sequences were concatenated. Gaps and stop codons were removed manually. Kimura's (1980) 2-parameter model was used to obtain a species-distance matrix based on the concatenated sequences using the program MEGA 3.0 (Kumar et al. 2004). Confidence of nodes was evaluated with bootstrap resampling methods implemented in MEAG 3.0. The robustness of statistical support for the tree branch was determined by 1000 bootstrap replicates.

## The Retrieval of Mtpts

Figure 1 presents a flowchart for the collection of mtpts. Homology search was done by comparing the mtDNA of each species against the cpDNA using the Blast2sequence program. Results with an *e*-value smaller than $10^{-3}$ were selected. Overlapped sequences from the same species were merged. These merged data were considered as the mtpts of a particular sampled taxon, namely the cp-derived regions in each mtDNA. Cp-derived genes, including protein-coding genes, ribosomal RNA genes, and transfer RNA genes that reside in each mtpt, were annotated for future use.

## Collection of Informative Gene Clusters and Inference of the Timing of Mtpt Transfer

In order to identify a possible cluster of transferred genes, mtpts of each species were sorted based on their positions in respective mtDNAs. An mtpt gene cluster was defined as mtpts composed of continuous cp gene sequences (in terms of their corresponding counterparts in cpDNA) without interruption by another mt gene. By definition, an mtpt-gene-cluster should have 2 boundaries at a specific mtDNA. A common transfer event was inferred when more than 2 species possessed at least one common mtpt-gene-cluster boundary. A common transfer event can be revealed by finding the shared boundary among mtpts, which is based on the assumption that a gene boundary would not be deleted in the transfer process. The mtpt-gene-cluster search was done for 2 groups: the monocots (including *Zea*, *Oryza*, and *Triticum*) and the dicots (including *Arabidopsis*, *Brassica*, and *Nicotiana*). If one boundary of an mtpt-gene-cluster is shared by both monocots and dicots, the deduction that the transfer of this mtpt-gene-cluster occurred before the divergence of these 2 plant lineages is reasonable. On the other hand, if the mtpt of an angiosperm lineage has a gene-cluster boundary identical to that of *Cycas*, then the transfer event of such mtpt-gene-cluster should have taken place in the common ancestor of the extant seed plants. Because the probability of 2 random transfers that share a common boundary is infinitesimal, by the use of the "shared boundary of continuous genes" method, it is logical and reasonable to conclude that a common transfer event is present between 2 or more lineages. Note that a single gene transfer cannot form a mtpt-gene-cluster, and such events are therefore excluded from the present analysis.

## Search of Pseudo/Functional Protein-Coding Genes within Mtpts

Each of the cp-derived genes in all mtpts was examined to determine whether the whole gene had been completely transferred. Incomplete genes were designated as pseudogenes. For a complete gene, an open reading frame search was performed using the NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) program to validate its completeness. A gene with one or more stop codons within its corresponding cp gene was also considered a pseudogene.

## Gain and Loss of tRNA Genes

The cp-derived tRNA genes of each species were also annotated in the process of mtpt retrieval. The function of each tRNA gene was quoted from published mtDNA

studies of *Arabidopsis*, *Zea*, *Oryza*, *Nicotiana*, *Triticum*, and *Cycas* (Unseld et al. 1997; Notsu et al. 2002; Chaw et al. 2004; Clifton et al. 2004; Sugiyama et al. 2005) and reconfirmed by the tRNAscan-SE program (Lowe and Eddy 1997). The loss and/or gain of cp-derived tRNA genes were mapped onto each of the branches of the phylogenetic tree reconstructed by concatenated mt protein-coding genes from the available mtDNA taxa. The principle of parsimony was employed to justify the distribution of cp-derived tRNA genes. Pseudo tRNA genes were not considered.

## Results and Discussion

### The First Mtpt Occurred At Least As Far Back As The Common Ancestor of Seed Plants

Mtpts of 2 different gene categories, namely, protein-coding and tRNA, were investigated separately in 11 plant species (fig. 2) using the Blast2sequence search of NCBI (see supplementary tables 1–3, Supplementary Material online). We did not identify any mtpt in algae, liverwort, or moss. In contrast, only the sampled 7 seed plants had mtpts. Mtpts in each of the 7 seed plants are listed in supplementary table 1 (Supplementary Material online). As depicted in both supplementary table 1 and figure 2 (Supplementary Material online), the distribution of mtpts among sampled taxa suggests that mtpts are only present in the known mtDNAs of gymnosperms and angiosperms, implying that the existence of mtpts may predate the divergence of extant gymnosperms and angiosperms. Furthermore, we identified 7 different conserved mtpts (table 2) in the sampled 7 seed plant lineages. Each of the mtpts is shared by at least 2 lineages/species.

Figure 3 is an NJ tree reconstructed with concatenated 21 mt protein-coding genes using *Chaetosphaeridium* and *Chara* as the outgroup. Topology of this tree strongly indicates that to the exclusion of 2 bryophytes, *Physcomitrella* (a moss) and *Marchantia* (a liverwort), the seed plants (including *Cycas* and the sampled angiosperms) and the angiosperms form 2 separate monophyletic clades. Within the angiosperms, the monocots and the eudicots also constitute 2 distinct subclades. When distributions (table 2) of shared mtpt-gene-clusters are mapped onto their respective branches, as in figure 3, the occurrence time of a particular mtpt can be considered as at least equivalent to the divergence time of its corresponding branch, which can be obtained from reliable fossil dates. Figure 3 indicates that no mtpt-gene-cluster was found in either the liverwort or moss lineage and that at the node leading to the seed plants there is a conserved and common mtpt-gene-cluster, *trnV*(*uac*)-*trnM*(*cau*)-*atpE*-*atpB*-*rbcL* (mtpt 1). Although the mtpt 1 is only found in the *Cycas* and the *Oryza* lineages, fragments of this long mtpt—including the genes *trnV*(*uac*), *trnM*(*cau*), *atpE*, *atpB*, *rbcL*—were found in *Cycas* and most of sampled angiosperms (see supplementary table 1, Supplementary Material online). These results suggest that this mtpt-gene-cluster might be the most ancient one and that transfer of the mtpt likely predates the split of gymnosperms and angiosperms. However, representative mtDNA from seedless vascular plants deserve an investigation to further fix the date of the first mtpt.

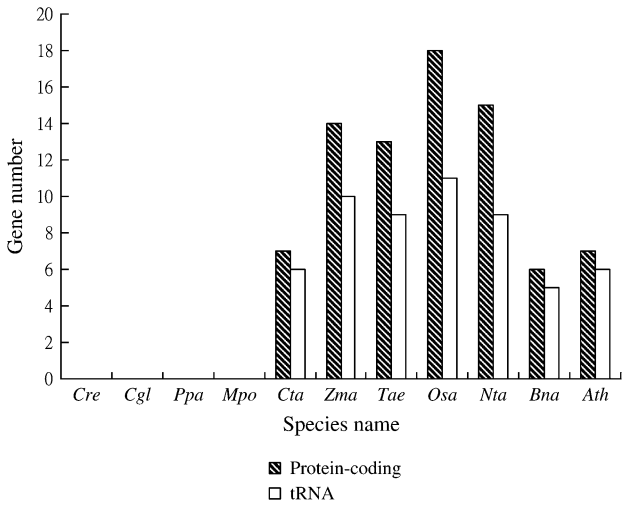


FIG. 2.—Comparison of mtpts in terms of the numbers of tRNA and protein-coding genes in the 11 sampled plant species.

Cummings et al. (2003) did a phylogenetic analysis of *rbcL* sequences from both the cpDNAs and mtDNAs of 5 angiosperms (*Zea mays*, *Oryza sativa*, *Brassica rapa*, *A. thaliana*, and *Ipomoea coccinea*). They discovered that at least 5 independent transorganellar transfers took place during the evolution of angiosperms. Clifton et al. (2004), in addition to using phylogenetic analysis, employed data from gene structure to approach the same issue. Both teams obtained a similar result by the use of sequence similarity comparison, but Clifton et al. (2004), using gene structure comparison, further discovered one transfer scenario that had occurred before the divergence of *Zea mays* and *Oryza sativa*. Using the shared gene boundary as a criterion, we consider that the *rbcL* genes in the mtDNAs of angiosperms came from a single transfer event with the *Cycas* rather than independently in the 2 lineages. Our consideration is further strengthened by the finding that mtDNAs of *Cycas* and most angiosperms contain both cp-derived *rbcL* gene and *trnM*(cau) (supplementary table 1, Supplementary Material online).

Figure 3 further depicts that 3 other mtpt-gene-clusters—*psaA-psaB*, *rps19-trnH*(*gug*)-*rpl2-rpl23* (mtpt 3), and *psbE-psbF*—are positioned at the node leading to the sampled angiosperms, implying that transfers of these 3 mtpts probably took place before divergence of the extant angiosperms. After the divergence of monocots from eudicots, the monocot lineage acquired the mtpt-gene-cluster *rps12-rps7* while the Brassicaeae (including *Arabidopsis* and *Brassica*) of the eudicots have taken up the mtpt-gene-cluster *ycf1-trnN*(*guu*). Moreover, after the split of *Oryza* and *Triticum* from *Zea*, the mtDNAs of the former 2 lineages have adopted an additional mtpt-gene-cluster, *rpl14-rps8*. Note that our estimates might represent the lower bound of a specific mtpt because our survey was conservatively based on the retrieved mtpt fragments in the 11 available living plants.

Likewise, the mtpt-gene-cluster (mtpt 3; see table 3) that is common to the *Zea*, the *Oryza*, and the *Nicotiana*

**Table 2**
**Distributions of Mtpt-Gene-Clusters in 7 Seed Plant Lineages**

| Mtpt-Gene-Clusters | Cta | Zma | Osa | Tae | Nta | Ath | Bna |
|---|---|---|---|---|---|---|---|
| (1) *trnV*(*uac*)-*trnM*(*cau*)-*atpE-atpB-rbcL* | + | − | + | − | − | − | − |
| (2) *psaA-psaB* | − | − | − | + | + | + | + |
| (3) *rps19-trnH*(*gug*)-*rpl2-rpl23* | − | + | + | − | + | − | − |
| (4) *psbE-psbF* | − | − | − | + | + | − | − |
| (5) *rpl14-rps8* | − | − | + | + | − | − | − |
| (6) *rps12-rps7* | − | + | − | + | − | − | − |
| (7) *ycf1-trnN*(*guu*) | − | − | − | − | − | + | + |

lineages, possibly had been transferred before the diversification of major angiosperm lineages. Clifton et al. (2004) proposed that *Zea* and *Oryza* shared a common recombination event, namely, the cp-derived *rbcL* fragment joining the cp-derived *rpl23-rpl2* fragment. Nonetheless, we consider the joining of mtpt 1 and mtpt 3 (see table 2 and fig. 3) via mt recombination at the common ancestor of the grass family (Poaceae) more likely because *Oryza* still retains both mtpt 1 and mtpt 3 in its mtDNA. Stern and Lonsdale (1982) also discovered that in the maize NB mtDNA a large cpDNA segment (12.6 kb) encompasses the cp *trnA*, *trnI*, *rrn16*, *trnV*, *rps12*, *rps7*, *ndhB*, and *trnL* genes. Our close inspection of the above segments reveals a large insertion/gap (~1.7 kb) located between the 2 gene-clusters, *trnA-trnI-rrn16-trnV* and *rps12-rps7-ndhB-trnL*. Therefore, it seems reasonable to infer that the mtDNAs of the grass family once contained the large 12.6-kb cpDNA segment and that parts of the segment containing the *trnA-trnI-rrn16-trnV* gene-cluster were lost via mt recombination and other mechanisms like deletion and insertions. Our inference is clearly supported by 3 observations: 1) cp-derived *trnA*(*ugc*)-*trnI*(*cau*) and *rrn16* fragments are present in the *Triticum* mtDNA, 2) cp-derived *trnI*(*cau*) and *rrn16* are also found in *Oryza* mtDNA, and 3) *Zea* and *Triticum* share the cp-derived *rps12-rps7* boundary (see supplementary tables 1 and 2 and fig. 3, Supplementary Material online).

## Gain and Loss of tRNA Mtpts in *Cycas* and Sampled Angiosperms

In rice, the tRNA gene sequences between mtDNA and cpDNA have high similarity (97.3−100%), so 11 of 23 tRNA genes in the rice mtDNA were considered to be of plastid origin (Hiratsuka et al. 1989). Sugiyama et al. (2005) asserted that cp-derived tRNA genes—such as *trnH*(*gug*), *trnM*(*cau*), *trnN*(*guu*), *trnP*(*ugg*),

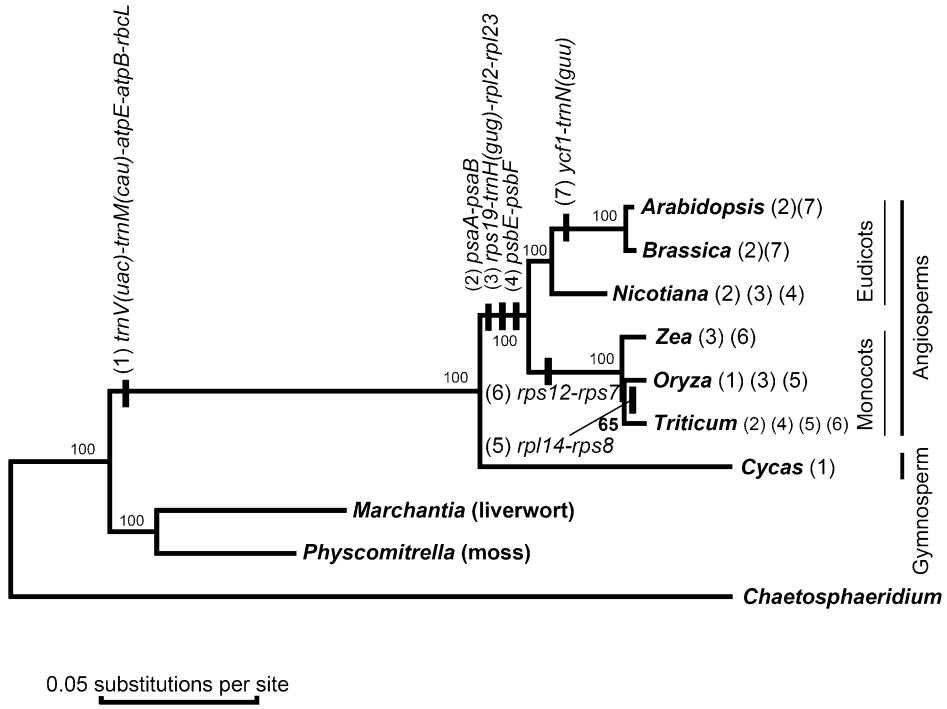


FIG. 3.—Mtpt-gene-clusters were plotted onto the branches of the rooted mtDNA phylogenetic tree for the 10 sampled plant taxa. *Chaetosphaeridium* was used as the outgroup. The known reliable date of a particular node was used to infer the approximate transfer time of a specific mtpt-gene-cluster. The number behind each species indicates a specific mtpt-gene-cluster listed in table 2. The numbers at each node denote bootstrap values per 100%.

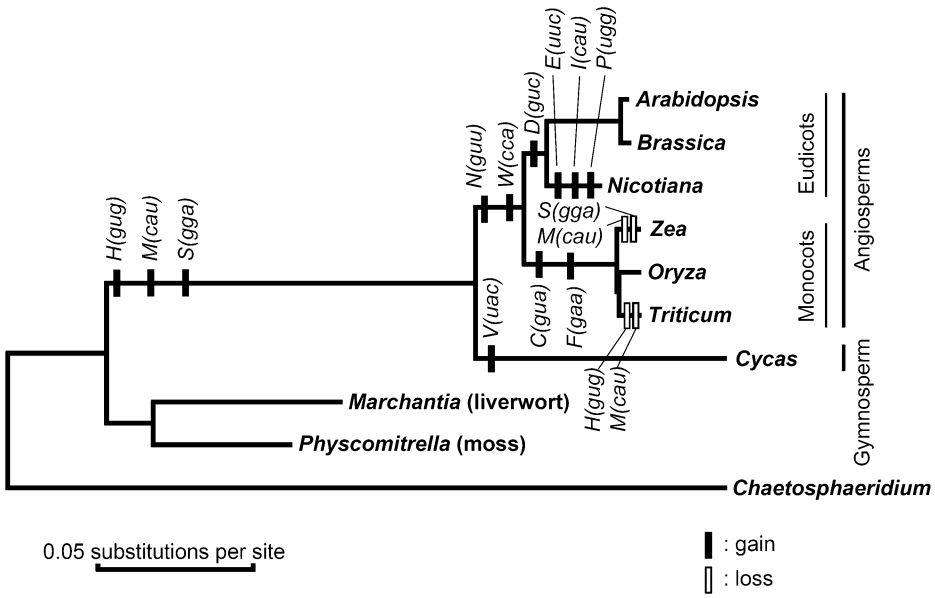


FIG. 4.—Chloroplast-derived tRNA gain (filled bar) and loss (open bar) based on parsimony algorithm are plotted on the branches leading to each lineage. The phylogenetic tree is the same as in figure 3.

*trnS*(*gga*), and *trnW*(*cca*)—are common to all angiosperms and that *trnD*(*guc*) is common to dicots only. However, the transfer date of other tRNA genes from cpDNA to mtDNA has been difficult to estimate because of limited sequence information in the genes (Clifton et al. 2004; Sugiyama et al. 2005).

Based on parsimony algorithm, the distributions of tRNA gain and loss (fig. 4, supplementary table 1, Supplementary Material online) were depicted on the mtDNA phylogeny of 11 plants. Figure 4 suggests that among the 11 cp-derived tRNA genes detected in the sequenced mtDNAs, *trnH*(*gug*), *trnM*(*cau*), and *trnS*(*gga*) appear to be commonly gained by all of the sampled seed plants. Coincidently, these 3 tRNA genes were considered part of the large cpDNA fragment uptaken by the *Oryza* mtDNA (Notsu et al. 2002). Therefore, it is reasonable to infer that after the divergence of gymnosperms from angiosperms, the *Cycas* lineage has gained the *trnV*(*uac*) gene while the common ancestor of extant angiosperms has taken up both the *trnN*(*guu*) and *trnW*(*cca*) genes.

Moreover, the data in figure 4 also suggest that after the split of monocots/Poaceae from eudicots, the monocots/Poaceae acquired *trnC*(*gua*) and *trnF*(*gaa*) while the eudicots gained *trnD*(*guc*). Notably, the *Nicotiana* lineage has independently acquired *trnE*(*uuc*), *trnI*(*cau*), and *trnP*(*ugg*). In contrast, the *Zea* lineage secondarily lost the *trnM*(*cau*) and *trnS*(*gga*), and the *Triticum* lost the *trnH*(*gug*) and *trnM*(*cau*). Sugiyama et al. (2005) proposed that "in contrast to the protein-coding genes, tRNA genes were transferred from cpDNA to mtDNA during the evolution of angiosperms," estimated to be approximately 150 MYA (Chaw et al. 2004). Nevertheless, our data strongly suggest that the transfer of tRNA genes commenced before the divergence of gymnosperms and angiosperms. Collectively, our results indicate that in plants transfers of cp protein-coding and tRNA genes to mtDNA have been co-occurring at least since 300 MYA (Carboniferous period), when the seed plants evolved.

As exemplified by the above parsimony method, mapping of commonly shared cp-derived tRNA genes onto the clades/branches of the mtDNA phylogenetic tree provides insight into the date of the gain or loss of a particular tRNA gene. However, an observed common tRNA mtpt of 2 (or more) different lineages might actually result from different transfer events. For example, the 3 cp-tRNA genes, *trnH*(*gug*), *trnM*(*cau*), and *trnS*(*gga*), are shared by all sampled seed plants, and *trnM*(*cau*) is the only one being included in a shared mtpt-gene-cluster (fig. 3). In addition, according to our result using Blast2sequence program, the *trnH*(*gug*) genes from *Cycas*, *Zea*, and *Oryza* were actually transferred from inverted repeat (IR) regions of cpDNAs, and the *trnH*(*gug*) genes of *Nicotiana*, *Arabidopsis*, and *Brassica* were transferred from the large single copy (LSC) regions of cpDNA. These data suggest different origins for *trnH*(*gug*) transfers (supplementary table 1, Supplementary Material online). Moreover, a *trnH*(*gug*) is also embedded in a commonly preserved mtpt-gene-cluster (fig. 3, mtpt 3) at the node leading to the extant angiosperms. Therefore, *trnH*(*gug*) is more likely to have been transferred since the time of the common ancestor of angiosperms rather than that of the seed plants. As for the *trnS*(*gga*), because its counterpart is absent from *Marchantia* mtDNA, Sugiyama et al. (2005) concluded that "the mechanism of integration the *trnS*(*gga*) into mtDNA appears to be different from that of other cp-derived tRNA." We observed that *trnS*(*gga*) is not included in any mtpt-gene-cluster, so it is difficult for us to judge the exact timing of its transfer.

### Protein-Coding Mtpts of Seed Plants Have Lost Functions Since Their First Transfer 300 MYA

A comparative analysis on the number of cp-derived protein-coding genes among 7 seed plants is shown in supplementary table 1 (Supplementary Material online). Note that the protein-coding mtpts vary from 7 in *Brassica* to 22 in *Nicotiana* and that the numbers in *Brassica* (7), *Arabidopsis* (8), and *Cycas* (8) are fewer than in *Zea* (17), *Triticum* (19), *Oryza* (21), and *Nicotiana* (22). We have observed that the trafficking of mtpts seems to be positively correlated with the variations (coefficient value $r^2 = 0.47$) of mtDNA size. Note that in contrast to the larger mtDNA size of *Zea* (570 K), *Oryza* (491 K), *Triticum* (453 K) and *Nicotiana* (431 K), those of *Brassica* (222 K), *Arabidopsis* (367 K), and *Cycas* (415 K) are relatively small. To date, the mechanism causing the diversity of transfer number remains unclear.

Previously, Notsu et al. (2002), Clifton et al. (2004), Ogihara et al. (2005), and Sugiyama et al. (2005) found that mtpts are nonfunctional as a result of sequence alterations (indels) and the absence of RNA editing. We also discovered that mainly due to frameshifts and indels, all the predicted protein-coding mtpts in the mtDNA of *Cycas taitungensis* are nonfunctional (data not shown). In summary, since the first transfer of mtpts in the common ancestor of extant seed plants 300 MYA (fig. 4), all of them have gradually degenerated or have even been thrown out of the mtDNAs. Therefore, mtpts do not add any new function to mtDNA and can be regarded as junk sequences.

### Mtpts Were Transferred Randomly from Any Region of CpDNA

Because cp-transfer occurs frequently (Notsu et al. 2002), it is also interesting to investigate the mtpt sequences' original locations in cpDNA to see if any particular regions in the cpDNA are hot spots of DNA transfer. Among the 79 transferred protein-coding sequences in the 4 sampled angiosperms (*Zea*, *Triticum*, *Oryza*, and *Nicotiana*) with larger mtDNAs, 7, 16, and 56 were from the SSC regions, the IR regions, and the LSC regions of cpDNAs, respectively. These numbers highly correspond to the lengths of SSC (12−20 kb), IR (20−28 kb), and LSC (80−90 kb) in the cpDNAs of angiosperms, suggesting that in angiosperms the cpDNA transfer has occurred randomly at any position in the cpDNAs and in proportion to the length of a cpDNA region. Similarly, transfer of genes from the cpDNA into the nrDNA of rice has been shown to occur randomly with respect to their positions in the plastid genome (Matsuo et al. 2005).

## Conclusions

We propose the use of gene boundaries to seek common transfers of mtpts in a particular node of a phylogenetic tree and to estimate the date of mtpt transfer. Our method is more reliable than earlier ones (Cummings et al. 2003; Clifton et al. 2004; Ogihara et al. 2005), based mainly on mtpt sequence comparisons. As in angiosperms (Notsu et al. 2002; Sugiyama et al. 2005), mtpts should be common in gymnosperms too, as exemplified by *Cycas*. Note that this is a minimal set of mtpts detected with a particularly high confidence approach, single mtpt transfer and those mtpt-cluster lost boundary cannot be detected.

In good agreement with the conclusion of Notsu et al. (2002) and Clifton et al. (2004) that in angiosperms all the protein-coding mtpts turn out to be degenerated while some of the cp-derive tRNA are still functional in mtDNA, we also discovered similar fates in both protein-coding and tRNA mtpts in the *Cycas* lineage. Moreover, we also found that the protein-coding mtpts of angiosperms occurred randomly at any position of cpDNA while the protein-coding mtpts of *Cycas* occurred only in the LSC region. Collectively, our results appear to support the proposition of Richly and Leister (2004) that "primary insertions of organellar DNAs are large and then diverge and fragment over evolutionary time." As our estimates were based on limited mtDNA representatives, additional mtDNA data from other gymnosperms, ferns, and nonvascular land plants are required to extend our knowledge on the evolution of mtpts in diversified land plant lineages.

## Supplementary Material

Supplementary materials are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

We thank the 2 anonymous reviewers for valuable comments. This work was supported by research grants from Research Center for Biodiversity, Academia Sinica (AS) to S.M.C. and D.W. and in part by a grant from Institute of Information Science, AS to A.C.S.

## Literature Cited

Boore JL. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27:1767–1780.

Chaw SM, Chang CC, Chen HL, Li WH. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol. 58: 424–441.

Clifton SW, Minx P, Fauron CM, et al. (13 co-authors). 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol. 136:3486–3503.

Cummings MP, Nugent JM, Olmstead RG, Palmer JD, et al. 2003. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms. Curr Genet. 43:131–138.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5:113.

Fauron C, Casper M, Gao Y, Moore B. 1995. The maize mitochondrial genome: dynamic, yet functional. Trends Genet. 11:228–235.

Fritzsch G, Schlegel M, Stadler PF. 2006. Alignments of mitochondrial genome arrangements: applications to metazoan phylogeny. J Theor Biol. 240:511–520.

Gray MW, Burger G, Lang BF. 2001. The origin and early evolution of mitochondria. Genome Biol. 2:REVIEWS1018

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

Hiratsuka J, Shimada H, Whittier R, et al. (10 co-authors). 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet. 217:185–194.

Huang CY, Grunheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiol. 138:1723–1733.

Kanno A, Nakazono M, Hirai A, Kameya T. 1997. Maintenance of chloroplast-derived sequences in the mitochondrial DNA of Gramineae. Curr Genet. 32:413–419.

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

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

Kumar S, Tamura K, Nei M. 2004. MEGA 3.0: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.

Kunnimalaiyaan M, Nielsen BL. 1997. Fine mapping of replication origins (ori A and ori B) in *Nicotiana tabacum* chloroplast DNA. Nucleic Acids Res. 25:3681–3686.

Lavrov DV, Lang BF. 2005. Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. Syst Biol. 54:651–659.

Lin X, Kaul S, Rounsley S, et al. (37 co-authors). 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature. 402:761–768.

Lowe TM, Eddy SR. 1997. tRNAscan−SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

Maier RM, Neckermann K, Igloi GL, Kossel 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.

Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol. 118:9–17.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA. 99:12246–12251.

Matsuo M, Ito Y, Yamauchi R, Obokata J. 2005. The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. Plant Cell. 17:665–675.

Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB. 2002. The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell. 14:2659–2679.

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. 2002. The complete sequence of the rice (*Oryza sativa L.*) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics. 268:434–445.

Noutsos C, Richly E, Leister D. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res. 15:616–628.

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Ohyama K. 1992. Transfer RNA genes in the mitochondrial genome from a liverwort, *Marchantia polymorpha*: the absence of chloroplast−like tRNAs. Nucleic Acids Res. 20:3773–3777.

Ogihara Y, Isono K, Kojima T, et al. (19 co-authors). 2002. Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Mol Genet Genomics. 266:740–746.

Ogihara Y, Yamazaki Y, Murai K, et al. (14 co-authors). 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 33:6235–6250.

Richly E, Leister D. 2004. NUMTs in sequenced eukaryotic genomes. Mol Biol Evol. 21:1081–1084.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. 6:283–290.

Shimada H, Sugiura M. 1991. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res. 19:983–995.

Stern DB, Lonsdale DM. 1982. Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. Nature. 299:698–702.

Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. Nucleic Acids Res. 31:5324–5331.

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics. 272:603–615.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.

Turmel M, Otis C, Lemieux C. 2002. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc Natl Acad Sci USA. 99:11275–11280.

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

Unseld M, Marienfeld JR, Brandt P, Brennicke A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet. 15:57–61.

Vahrenholz C, Riemen G, Pratje E, Dujon B, Michaelis G. 1993. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. Curr Genet. 24: 241–247.

Watanabe N, Nakazono M, Kanno A, Tsutsumi N, Hirai A. 1994. Evolutionary variations in DNA sequences transferred from chloroplast genomes to mitochondrial genomes in the Gramineae. Curr Genet. 26:512–518.

Woloszynska M, Bocer T, Mackiewicz P, Janska H. 2004. A fragment of chloroplast DNA was transferred horizontally,

probably from non-eudicots, to mitochondrial genome of *Phaseolus*. Plant Mol Biol. 56:811–820.

Wu CS, Wang YN, Liu SM, Chaw SM. Forthcoming 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 plant. Mol Biol Evol. 24:1366–1379.

Yamauchi A. 2005. Rate of gene transfer from mitochondria to nucleus: effects of cytoplasmic inheritance system and intensity of intracellular competition. Genetics. 171:1387–1396.

Yuan Q, Hill J, Hsiao J, Moffat K, Ouyang S, Cheng Z, Jiang J, Buell CR. 2002. Genome sequencing of a 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. Mol Genet Genomics. 267:713–720.

Dan Graur, Associate Editor

Accepted June 25, 2007