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-psbB, rps19-trnH(gug)-rpl2-rpl23, and psbE-psbF—were dated to the common ancestor of extant angiosperms, at least 150 MYA. However, all cpDNA transfers and/or integration of entire organellar genomes to nuclear genomes (nrDNAs) have independently occurred 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; Noutsu et al. 2002; Handa 2003; Turnel et al. 2003; Clifton et al. 2004; Oghara 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 (Noutsu 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), Oghara 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...
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) 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.

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

<table>
<thead>
<tr>
<th>Classification</th>
<th>Scientific Name</th>
<th>Abbreviation</th>
<th>Common Name</th>
<th>mtDNA Acc. no.</th>
<th>Reference</th>
<th>cpDNA Acc. no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Algae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyta</td>
<td>Chlamydomonas reinhardtii</td>
<td>Cre</td>
<td>Green algae</td>
<td>NC_001638</td>
<td>Vahrenholz et al. (1993)</td>
<td>NC_005353</td>
<td>Maul et al. (2002)</td>
</tr>
<tr>
<td><strong>Streptophyta</strong></td>
<td>Chaetosphaeridium globosum</td>
<td>Cgl</td>
<td>Charophyte</td>
<td>NC_004118</td>
<td>Turnel et al. (2002)</td>
<td>NC_004115</td>
<td>Turnel et al. (2002)</td>
</tr>
<tr>
<td><strong>Land plants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marchantia polymorpha</td>
<td>Mpo</td>
<td>Liverwort</td>
<td>NC_001660</td>
<td>Oda et al. (1992)</td>
<td>NC_001319</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Seed plants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gymnosperm</td>
<td>Cycas taitungensis</td>
<td>Cta</td>
<td>Taitung cycad</td>
<td>NC_007579</td>
<td>Ogihara et al. (2005)</td>
<td>NC_002762</td>
<td>Ogihara et al. (2002)</td>
</tr>
<tr>
<td>Monocots</td>
<td>Triticum aestivum</td>
<td>Tae</td>
<td>Wheat</td>
<td>NC_000029</td>
<td>Notsu et al. (2002)</td>
<td>NC_001320</td>
<td>Hiratsuka et al. (1989)</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>Zma</td>
<td>Maize</td>
<td>BA000009</td>
<td>Not available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dicots</strong></td>
<td>Nicotiana tabacum</td>
<td>Nta</td>
<td>Tobacco</td>
<td>NC_006581</td>
<td>Sugiyama et al. (2005)</td>
<td>NC_001879</td>
<td>Nielsen (1997)</td>
</tr>
<tr>
<td>Brassica napus</td>
<td>Bna</td>
<td>Oilseed rape</td>
<td>AP006444</td>
<td>Handa (2003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>Ath</td>
<td>Mouse-ear cress</td>
<td>NC_001284</td>
<td>Unsed et al. (1997)</td>
<td>NC_000932</td>
<td>Sato et al. (1997)</td>
<td></td>
</tr>
</tbody>
</table>

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*, *Arabidopsis 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
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 Cyca, 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 extinct 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.

Cummings et al. (2003) did a phylogenetic analysis of rbl 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 Brassicaceae (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
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(ugg)-trnL(cau)* and *rrn16* fragments are present in the *Triticum* mtDNA, 2) cp-derived *trnL(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).

### Table 2

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

<table>
<thead>
<tr>
<th>Mtpt-Gene-Clusters</th>
<th>Cta</th>
<th>Zma</th>
<th>Osa</th>
<th>Tae</th>
<th>Nta</th>
<th>Ath</th>
<th>Bna</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) <em>trnV(uac)-trnM(cau)-atpE- AtpB-rbcL</em></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(2) <em>psbA- psbB</em></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>(3) <em>rps19-trnH(gug)-rpl2-rpl23</em></td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(4) <em>psbE-psbF</em></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(5) <em>rpl14-rps8</em></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(6) <em>rps12-rps7</em></td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(7) <em>ycf1-trnN(guu)</em></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

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)*,
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(uac), trnM(cau), and trnP(uac). 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 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.

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.

![Diagram illustrating tRNA gain and loss in plant species]
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


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.
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.


Dan Graur, Associate Editor

Accepted June 25, 2007