Molecular Biology and Evolution

Comparative Analysis of Structural Diversity and Sequence Evolution in Plant Mitochondrial Genes Transferred to the Nucleus

Shao-Lun Liu,[1] *Yan Zhuang,*[1,2] *Peter Zhang,*[1] *and Keith L. Adams*[1]

UBC Botanical Garden and Centre for Plant Research, and Department of Botany, University of British Columbia, Vancouver, BC, Canada

The transfer of functional mitochondrial genes to the nucleus is an ongoing process during plant evolution that has made a major impact on cytonuclear interactions and mitochondrial genome evolution. Analysis of evolutionarily recent transfers in plants provides insights into the evolutionary dynamics of the process and how transferred genes become functional in the nucleus. Here, we report 42 new transferred genes in various angiosperms, including 9 separate transfers of the succinate dehydrogenase gene *sdh3*. We performed comparative analyses of gene structures and sequence evolution of 77 genes transferred to the nucleus in various angiosperms, including multiple transfers of 10 genes in different lineages. Many genes contain mitochondrial targeting presequences, and potentially 5′ *cis*-regulatory elements, that were acquired from pre-existing nuclear genes for mitochondrial proteins to create chimeric gene structures. In eight separate cases, the presequence was acquired from either the *hsp70* chaperonin gene or the *hsp22* chaperonin gene. The most common location of introns is in the presequence, and the least common is in the region transferred from the mitochondrion. Several genes have an intron between the presequence and the core region, or an intron in the 5′UTR (untranslated region) or 3′UTR, suggesting presequence and/or regulatory element acquisition by exon shuffling. Both synonymous and nonsynonymous substitution rates have increased considerably in the transferred genes compared with their mitochondrial counterparts, and the degree of rate acceleration varies by gene, species, and evolutionary timing of transfer. Pairwise and branchwise K_a/K_s analysis identified four genes with evidence for positive selection, but positive selection is generally uncommon in transferred genes. This study provides a detailed portrayal of structural and sequence evolution in mitochondrial genes transferred to the nucleus, revealing the frequency of different mechanisms for how presequences and introns are acquired and showing how the sequences of transferred genes evolve after movement between cellular genomes.

Introduction

Mitochondria are derived from an alpha proteobacterial endosymbiont that lost or transferred to the nucleus most genes following endosymbiosis (reviewed in Gray 1992, 1999). Only a small number of genes remain in mitochondrial genomes, ranging from 67 protein-coding genes in the protist *Reclinomonas americana* to only three genes in some apicomplexans (reviewed in Adams and Palmer 2003). Mitochondrial genomes of animals have a highly conserved set of 13 protein-coding genes that is almost invariant among the over 1,000 sequenced genomes from metazoans (Gissi et al. 2008). In contrast, flowering plant mitochondrial genomes contain up to 40 known protein-coding genes, but the gene number is highly variable among species (Adams et al. 2002). The genes in mitochondrial genomes mostly code for components of the respiratory chain and ribosomal proteins, with a few other genes present in some species.

A major factor accounting for the variable gene content among mitochondrial genomes is transfers of mitochondrial genes to the nucleus during eukaryotic evolution. Functional gene transfers have contributed to the coevolution of the mitochondria and the nucleus. They are an ongoing evolutionary process in land plants and some green algae (reviewed in Adams and Palmer 2003; Bonen and Calixte 2006) but not in animals. Only one potential case of functional transfer has been reported in animals: *atp9* in the sponge *Amphimedon queenslandica* (Erpenbeck et al. 2007). Functional transfers involve activation and expression of the gene in the nucleus after physical transfer from the mitochondrion (reviewed in Adams and Palmer 2003). To become active, a newly transferred gene must gain a promoter and other regulatory elements for proper expression and a sequence for targeting the protein product to the mitochondrion if the protein does not already have the necessary targeting information. Many transferred genes have gained a mitochondrial targeting presequence that is translated and then removed from the protein after this has been imported into the mitochondria. Some genes have gained a mitochondrial presequence from a pre-existing gene for a mitochondrial protein (reviewed in Adams and Palmer 2003). After activation of the transferred nuclear copy, both this and the mitochondrial copy can be expressed for a period of time, however short. Coexpression of nuclear and mitochondrial copies of a gene, at least at the transcript level, has been shown for *cox2* in some legumes, *rpl5* in *Triticum* (wheat), and *sdh4* in *Populus* (Adams et al. 1999; Sandoval et al. 2004; Choi et al. 2006). In contrast to functional gene transfers, nonfunctional gene transfers have occurred in which the transferred gene becomes a pseudogene in the nucleus. Mitochondrial pseudogenes have been found in the nucleus of many eukaryotes, including a diverse array of animals, and the nonfunctional gene transfers are an ongoing evolutionary process in many, if not most, eukaryotes (reviewed in D'Errico et al. 2004; Leister 2005).

Over the past several years, numerous functional mitochondrial gene transfers have been characterized in plants. Several reports were about a single gene in a single species, but a few cases of transfer have been studied in detail in several closely related species to determine their phylogenetic distribution and evolutionary timing (Adams et al. 1999; Ong and Palmer 2006; Hazle and Bonen 2007). Another approach has been to study the same gene in distantly related angiosperm species that probably represent

[1] Equal contributions.
[2] Present address: Plant Biology Graduate Group, UC Davis.

Key words: mitochondria, genome evolution, targeting sequence, intron.

E-mail: keitha@interchange.ubc.ca.

Mol. Biol. Evol. 26(4):875–891. 2009
doi:10.1093/molbev/msp011
Advance Access publication January 23, 2009

separate transfers to the nucleus. That was done first with *rps10* in a variety of angiosperms (Adams et al. 2000) and then extended on a smaller scale to *sdh3*, *sdh4*, and *rps19* (Adams, Rosenblueth, et al. 2001; Adams et al. 2002). The above-mentioned studies have provided insights into how newly transferred genes become activated in the nucleus.

In this study, we identified, sequenced, and analyzed 42 transferred genes in a variety of flowering plants. We then assembled a data set of 77 transferred genes by adding the new sequences to previously published sequences to create the first large-scale analysis of gene structures and sequence evolution among functional gene transfers. We asked several questions about the transferred gene structures including: What is the frequency of genes containing a mitochondrial targeting presequence, compared with genes without a presequence? Is it common for a transferred gene to gain a presequence from another gene for a mitochondrial protein, or from a gene for a nonmitochondrial protein? How common are introns in different locations of the genes (presequence, core, UTRs, etc.)? We performed analyses of synonymous and nonsynonymous substitution rates to determine how much they increase after gene transfer to the nucleus, and if it varies by gene, species, or approximate phylogenetic timing of the transfer. We also tested for evidence of positive selection. The results of this study provide detailed insights into the process of transferred gene activation as well as the subsequent sequence evolution in the nucleus.

Materials and Methods

Sequence Database Searches and Targeting Evaluation

Genes transferred to the nucleus were identified by using TblastN searches of National Center for Biotechnology Information's (NCBI) expressed sequence tag (EST) database, using an angiosperm mitochondrial copy of the gene as a query. GenBank numbers for identified ESTs are in supplementary table S1, Supplementary Material online. Mitochondrial targeting of each gene was predicted using TARGETP v.1.1 (Emanuelsson et al. 2000), MITOPROT (Claros and Vincens 2006), PREDOTAR v.1.03 (Small et al. 2004), and WOLF PSORT (Horton et al. 2007). Both MITOPROT and TARGETP predict N-terminal targeting presequence cleavage sites. Results of the targeting prediction programs are shown in supplementary table S2, Supplementary Material online. We considered a sequence to be mitochondrially targeted if two or more of the prediction programs predicted mitochondrial targeting.

In addition to mitochondrial targeting, ESTs were evaluated for the following features to determine if they are derived from nuclear genes: Some genes contain a region derived from a pre-existing nuclear gene in a chimeric structure with the gene transferred from the mitochondrion. PCR amplification of most newly identified genes using genomic DNA, followed by sequencing, was done to identify spliceosomal introns that are present in nuclear but not mitochondrial genes. Absence of the mitochondrial copy in some species was inferred from a DNA gel blot hybridization survey of 280 angiosperm species (Adams et al. 2002). Genes with considerably diverged sequences compared with the mitochondrial copy in other angiosperms were considered candidate transferred genes because most mitochondrial genes in angiosperms, with only a few exceptional taxa (Mower et al. 2007), have highly conserved sequences. However, sequence divergence results were interpreted with caution. Most newly identified genes contain two or more of the above features. Exceptions are *rpl5* in *Fragaria* and *rps1* in *Ocimum*. Both of them have high K_s values compared with genes in the mitochondrion of other species. In addition, *rpl5* in *Fragaria* has a long 5′ extension of the open reading frame that may be a targeting presequence, and *rps1* has been lost from the mitochondrion of most members the Lamiales order of which *Ocimum* is a member (Adams et al. 2002).

Plant Growth, DNA Extraction, Gene Amplification and Sequencing

DNA was extracted from seedlings, leaves, or fruits using the Qiagen DNAeasy kit (Mississauga, ON, Canada) according to the manufacturer's instructions, from the following species: *Beta vulgaris*, *Citrus reticulata*, *Cucumis sativus*, *Eschscholzia californica*, *Fragaria* × *ananassa*, *Gossypium arboreum*, *Ipomoea batatas*, *Lactuca sativa*, *Malus* × *domestica*, *Ocimum basilicum*, *Prunus persica*, *Rhododendron catawbiens*, *Triticum aestivum*, and *Vaccinium corymbosum*. Young leaves of *Euphorbia cyparissias*, *Aquilegia formosa*, and *Nymphaea tetragona* were obtained from the UBC Botanical Garden for DNA extractions.

Genes transferred to the nucleus newly identified in this study were amplified from genomic DNA by gradient polymerase chain reaction (gradient-PCR) using gene-specific primers (listed in supplementary table S1, Supplementary Material online). The mitochondrial copies of *cox2*, *rpl2*, *rps11*, *rps12*, *rps14*, and *sdh3* were PCR amplified from *Eschscholzia* and *Nymphaea* using primers listed in supplementary table S1, Supplementary Material online. The PCR was performed in a reaction mixture (10 μl) consisting of 4.88 μl of ddH_2O, 1 μl of genomic DNA (20 μg/μl), 1 μl PCR buffer (500 mM KCl, 100 mM Tris base, pH 8.3), 1 μl of 2.5 mM $MgCl_2$ solution, 1 μl of 0.2 mM dNTPs (Fermentas, Burlington, ON, Canada), 0.5 μl of 0.4 μM each primer, and 0.12 units of Taq DNA polymerase. The gradient-PCR conditions were 96 °C for 4 min, and 30 cycles of 96 °C for 40 s, a gradient of 50–60 °C for 40 s (one reaction every 2 °C), and 72 °C for 1 min, with a final extension after cycling of 72 °C for 10 min. The PCR products were run on 1.5% agarose gels and extracted from the gel using a QIAquick Gel Extraction Kit (Qiagen) or sequenced directly after purification with Sephadex columns. The sequencing was performed as in Liu and Adams (2008). The sequences have been assigned GenBank numbers EU924189–EU924199.

Sequence Alignment and Phylogenetic Analysis

Core-region sequences of each gene were aligned using the Perl script TRANSALIGN with the program ClustalW (Bininda-Emonds 2005). Aligned sequences were manually inspected and adjusted using the program BIOEDIT (Hall 1999). Only the core regions of the transferred genes were

used for further analyses (supplementary figs. S1 and S2, Supplementary Material online). Core regions refer to the sequence that was transferred from the mitochondrion and is homologous to the sequence in other eukaryotes, and noncore regions refer to 5′ extensions of the reading frames that were acquired after transfer to the nucleus. Percent sequence identity of the transferred genes to their mitochondrial counterparts in other taxa was determined using EBI pairwise alignment algorithms with EMBOSS::water (local) method (http://www.ebi.ac.uk/Tools/emboss/align/index.html). For phylogenetic analysis, the maximum likelihood (ML) method and bootstrapping (500 replicates) were conducted using programs DNAML and SEQBOOT in the package PHYLIP v.3.68 (Felsenstein 2008). Results from the phylogenetic analysis were viewed and exported using the program TREEVIEW (Page 1996).

Pairwise and Branchwise Analyses of Nucleotide Substitution Rates

Nonsynonymous (K_a) and synonymous (K_s) nucleotide substitution rates were calculated by using the program CODEML in the package PAML v3.15 (Yang 1997). For pairwise analysis, mitochondrial genes from the following four taxa were used as references: a basal angiosperm (either *Amborella* or *Nymphaea*), a basal eudicot in the Ranunculales order (*Eschscholzia*, *Ranunculus*, or *Dicentra)*, a full-length gene from a eudicot (usually *Nicotiana*), and a monocot. A total of 513 pairwise comparisons were done using a Perl script. For branchwise analysis, constrained tree topologies following well-established phylogenetic relationships (Soltis et al. 2005) were used (supplementary fig. S1, Supplementary Material online). Branch-site model followed by likelihood ratio test (LRT) was used to test if there has been positive selection acting on individual codons following the procedure described in Liu and Adams (2008). Briefly, two LRTs, Model 1 (M1)–Model A test1 (MA_{test1}) and Model A test2 (MA_{test2})–MA_{test1} were conducted to test if the detection of positive selection receives statistical support (Yang 1997; Yang and Nielson 2002). To see if the second model is significantly better than the first model, the value derived from the formula, $2\delta L = 2(\mathrm{Ln1} - \mathrm{Ln2})$, where Ln1 and Ln2 indicate the ML index for the first and second model, was compared against chi-square distribution. The degree of freedom is 2 for M1–MA_{test1} and 1 for MA_{test2}–MA_{test1}. For positively selected codons, Bayes Empirical Bayes analysis was used to test if it is statistically significant. Those codons that show lower Bayesian posterior probability (<0.95) are not considered as strong evidence for positively selected sites due to the increase of false positive rate ($>5\%$; Yang et al. 2005).

Determination of GC Content and Codon Usage

The GC content at silent sites (GC3s) and the effective number of codon usage (ENC) were estimated by using the program CODONW (http://www.molbiol.ox.ac.uk/cu/codonW.html). The *t*-test and Wilcoxon rank test were implemented to test if there is a significant difference for GC3s or ENC between mitochondrial genes and the transferred nuclear genes. Pearson's correlation test was used to determine if there is any significant correlation between K_s and GC3s or K_s and ENC. The statistically significant level was set at 95%. The statistical analyses were conducted using the program R (http://www.r-project.org/).

Results and Discussion

Identification of 42 Ribosomal Protein and Succinate Dehydrogenase Genes Transferred to the Nucleus

We used searches of the EST database at NCBI to identify new cases of mitochondrial genes that have been transferred to the nucleus (see Methods). We obtained genomic DNA sequences by PCR amplification and sequencing for most of the genes to identify intron positions. Forty-two transferred genes were discovered, and they are presented by gene in the following paragraphs. Gene structures, sizes, and intron positions are shown in figure 1. Core regions refer to the sequence that was transferred from the mitochondrion and is homologous to the sequence in other eukaryotes, and noncore regions refer to extensions of the reading frames that were acquired after transfer to the nucleus.

Sdh3: We identified genes for succinate dehydrogenase subunit 3 (*sdh3*) that have been transferred to the nucleus in nine species. Six of the nine transferred *sdh3* genes have regions derived from other nuclear genes for mitochondrial proteins (figs. 1 and 2). In one case, *sdh3* in *Malus* (apple), the gene transferred from the mitochondrion was inserted into a pre-existing gene for the mitochondrial protein translation elongation factor EF-Tu (figs. 1 and 2) and became expressed using the targeting sequence and regulatory elements of the host gene. Surprisingly, *sdh3* in *Ipomoea*, *Arabidopsis*, and *Populus* have presequences and introns derived from the mitochondrial chaperonin *hsp70*. To evaluate the evolutionary timing of the *hsp70* region acquisition in each *sdh3* gene, we performed a phylogenetic analysis of the presequence of *hsp70* from several eudicots and the *hsp70*-derived regions of *sdh3* from *Ipomoea*, *Arabidopsis*, and *Populus*. Each of the *hsp70*-derived regions of *sdh3* branches with the *hsp70* of the same or a closely related species (fig. 3). These results indicate that each of the *hsp70*-derived regions is a separate and recent acquisition. Coincidentally, two of the *sdh3* genes in distantly related lineages, from *Triphysaria* (a parasitic asterid) and *Gossypium* (cotton), contain a presequence derived from mitochondrial chaperonin *hsp22*, also by separate acquisitions. Other transferred *sdh3* genes with regions derived from pre-existing nuclear genes for mitochondrial proteins include: *sdh3* in *Vaccinium* (blueberry) from cytochrome c reductase, *sdh3* in *Lactuca* (lettuce) from an uncharacterized protein that is predicted to be mitochondrial targeted, *sdh3* in *Ocimum* (basil) from NAD-dependent malate dehydrogenase, and *sdh3* in *Euphorbia* has a short region that is similar to 2-oxoacid dehydrogenase (figs. 1 and 2). The nine new cases of *sdh3* in the nucleus add to the five cases previously reported (Adams, Rosenblueth, et al. 2001; Choi et al. 2006). Most of the 14 transferred *sdh3* genes have a different structure (fig. 1), probably as a result of separate gene activations following transfers to the nucleus.

Sdh4: We identified *sdh4* genes that were transferred to the nucleus in four species. Most of the nuclear *sdh4* genes

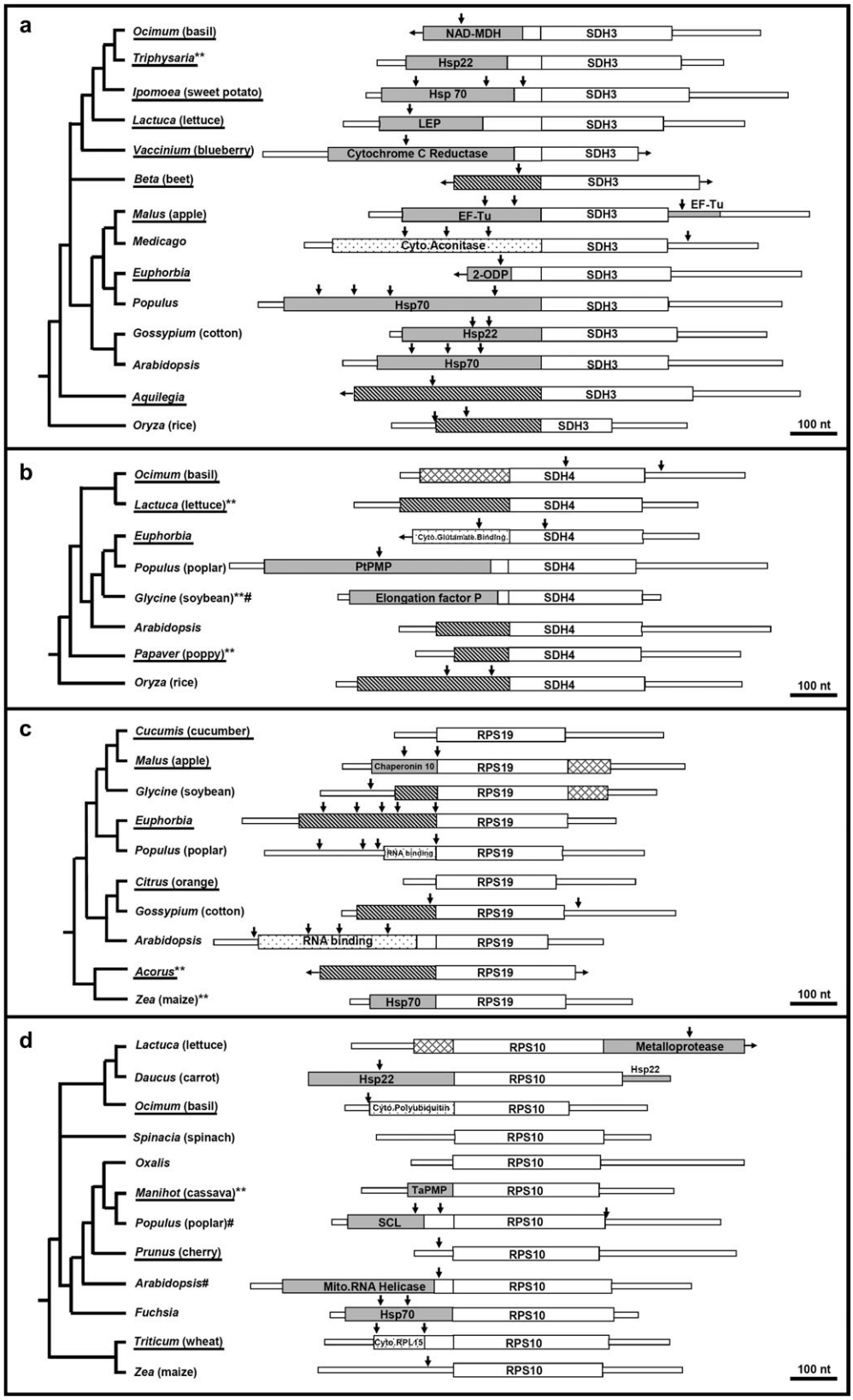


FIG. 1.

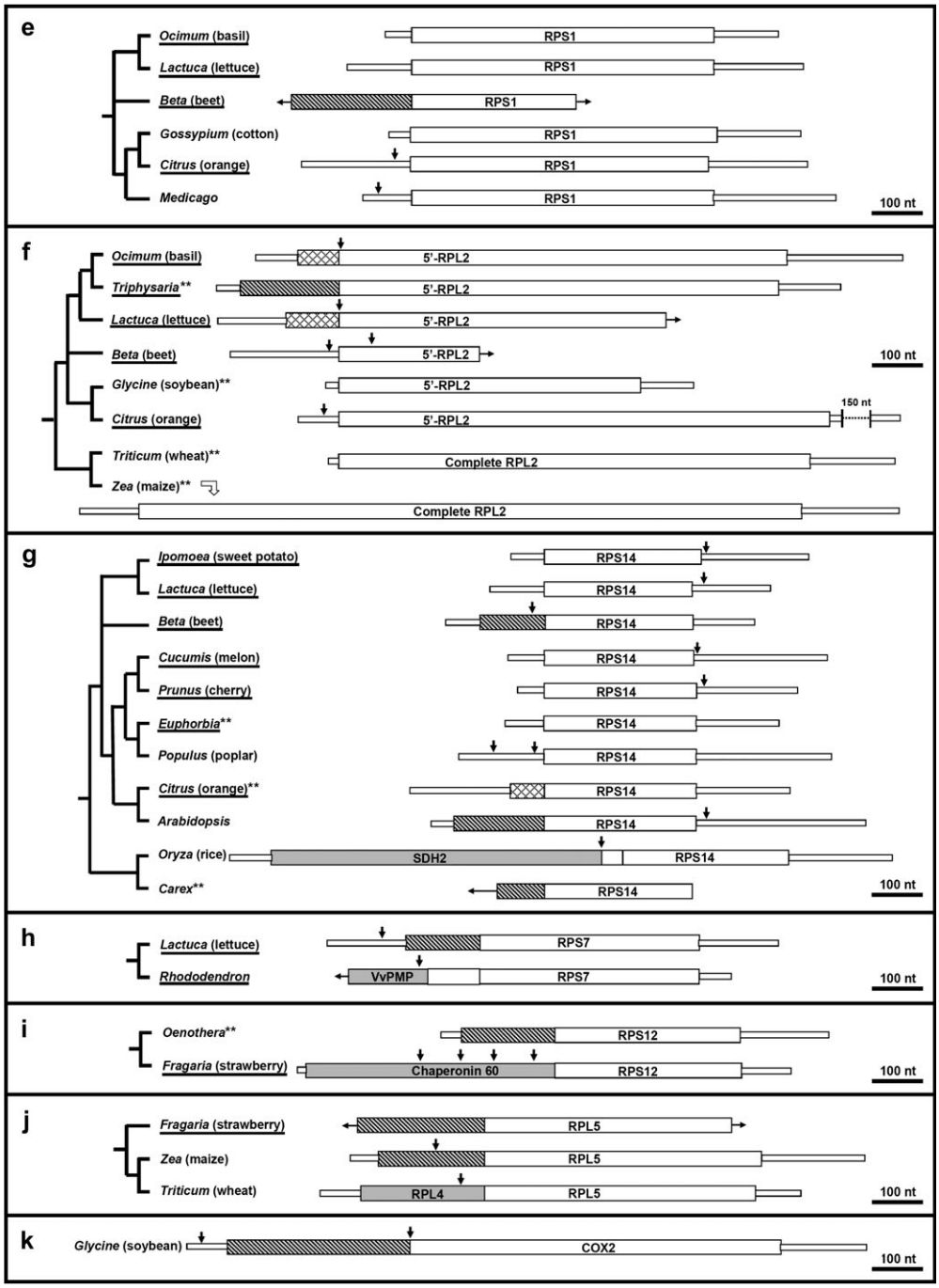


FIG. 1.—Structures of transferred genes. Schematic showing the structure of mitochondrial genes transferred to the nucleus, including genes characterized in this study (underlined) and previously published genes. GenBank numbers are listed in supplementary table S1, Supplementary Material online. (*a*) *rps19*, (*b*) *sdh3*, (*c*) *sdh4*, (*d*) *rps10*, (*e*) *rps1*, (*f*) *5′rpl2*, (*g*) *rps14*, (*h*) *rps7*, (*i*) *rps12*, (*j*) *rpl5*, and (*k*) *cox2*. Vertical arrows indicate intron positions. Horizontal arrows indicate incomplete sequences. Core regions (regions transferred from the mitochondrion) are indicated in boxes with gene names. Different shadings refer to different origins of noncore regions. Gray shading % indicates regions derived from pre-existing nuclear genes for mitochondrial proteins. Dot shading % indicates regions derived from genes whose products do not function in mitochondria. Diagonal line shading % indicates 5′noncore regions that may include presequences derived from unknown sources. Checked shading % indicates 5′noncore regions that probably do not contain presequences or 3′extensions of the reading frame of unknown origin. A box with no shading in the noncore region indicates sequence greater than 25 bp in a chimeric gene that does not appear to be derived from the host gene. Double asterisks indicate those cases for which only cDNA sequences are available. Number signs (#) indicate previously published chimeric genes where the donor gene was identified in this study. The phylogenetic trees are based on well-established phylogenetic relationships (Soltis et al. 2005). Scale bars equal to 100 nt. Small gene fragments of *rps10* (Adams et al. 2000) and *rps14* (Ong and Palmer 2006) are not shown. Abbreviations and gene names: 2-ODP, 2-oxoacid dehydrogenase protein; EF-Tu, elongation factor Tu; Hsp22, heat shock protein 22; Hsp70, heat shock protein 70; LEP, late embryogenesis protein; NAD-MDH, NAD-dependent malate dehydrogenase; PtPMP, *Populus trichocarpa* putative mitochondrial protein; RPL4, ribosomal protein L4; RPL5, ribosomal protein L5; RPL15, ribosomal protein L15; SCL, succinyl-Co ligase; TaPMP, *Triticum aestivum* putative mitochondrial protein; VvPMP, *Vitis viniferia* putative mitochondrial protein.

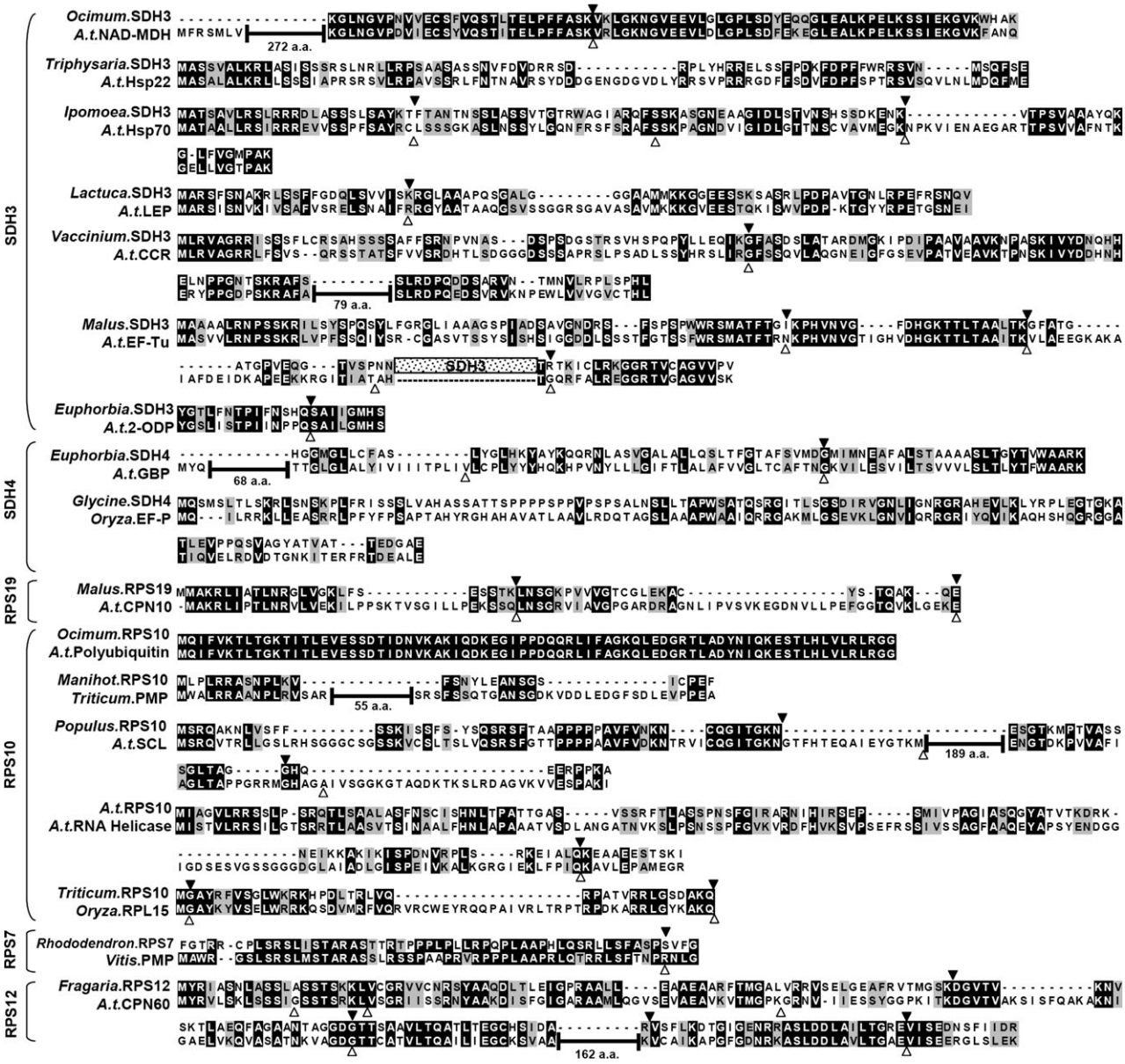


FIG. 2.—Alignments of transferred genes identified in this study that have noncore regions derived from other genes. Black arrowheads refer to intron positions in the corresponding genomic DNA sequence. White arrowheads indicate intron positions in the sequences of the homologous genes. Nonoverlapping regions over 50 amino acids are abbreviated by brackets. Identical amino acids are shown with white letters on a black background; amino acids with similar chemical properties are shown on a gray background. The core region of *sdh3* in *Malus* is shown as a dotted rectangle. Gene names and abbreviations: A.t., *Arabidopsis thaliana*; 2-ODP, 2-oxoacid dehydrogenase protein; CCR, cytochrome c reductase; CPN10, chaperonin 10; CPN60, chaperonin 60; EF-P, elongation factor P; EF-Tu, elongation factor Tu; GBP, glutamate binding protein; Hsp22, heat shock protein 22; Hsp70, heat shock protein 70; LEP, late embryogenesis protein; NAD-MDH, NAD-dependent malate dehydrogenase; PMP, putative mitochondrial protein; PPR, pentatricopeptide repeat-containing protein; RPL15, ribosomal protein L15; SCL, succinyl-CoA ligase. GenBank numbers are given in supplementary table S1, Supplementary Material online.

are very divergent in the 5′ half but highly conserved in the 3′ half (see supplementary fig. S2, Supplementary Material online). *Sdh4* from *Euphorbia* has a 5′ extension of the reading frame with sequence similarity to a putative glutamate binding protein that is not predicted to be mitochondrial targeted, and the two genes share an intron position (fig. 2). However, the *Euphorbia sdh4* is predicted to be mitochondrial targeted by Mitoprot and TargetP. Thus, it appears that a region of a nonmitochondrial protein has been co-opted into a mitochondrial targeting sequence followed by considerable sequence divergence. *Sdh4* in *Lactuca* (lettuce) and *Papaver* (poppy) have predicted presequences that are derived from unknown sources, and *sdh4* in *Ocimum* has a 5′ extension of the reading frame that is not predicted to be a presequence. The four transferred *sdh4* genes add to the four previously reported (Adams, Rosenblueth, et al. 2001; Choi et al. 2006). In addition, we identified a candidate gene for the source of the presequence of *sdh4* in *Glycine* (figs. 1 and 2).

Rps19: We identified five new genes for ribosomal protein S19 (*rps19*) that have been transferred to the nucleus.

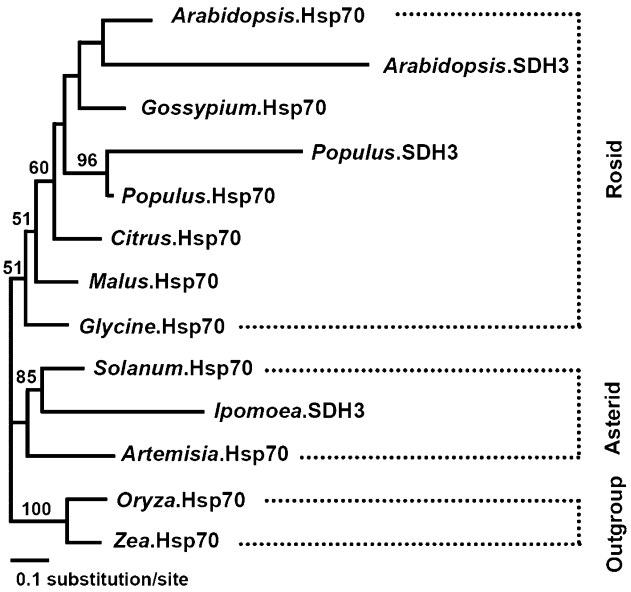


FIG. 3.—Analysis of *sdh3* presequences derived from *hsp70*. ML phylogenetic tree of *hsp70* sequences from various angiosperms and *sdh3* presequences from *Ipomoea*, *Arabidopsis*, and *Populus*. Bootstrap values from 500 replicates of ML analyses are labeled on the internodes.

The sizes of the core regions are variable due to 3′ extensions of the reading frame and indels (fig. 1; supplementary fig. S2, Supplementary Material online). Two of the transferred genes contain a presequence: *Malus rps19* is derived from a mitochondrial chaperonin 10 gene and *Euphorbia rps19* is from an unknown source (fig. 1). *Rps19* in *Cucumis* and *Citrus* do not contain a 5′ extension of the reading frame that could be a presequence, as judged by in-frame stop codons present not far upstream of the start codons. The targeting information presumably is contained within the core region of the protein. The five new cases of *rps19* transferred to the nucleus add to the five previously reported (Adams et al. 2002; Fallahi et al. 2005; Choi et al. 2006). Most of the 10 transferred *rps19* genes have a different structure in the noncore region (fig. 1), probably as a result of separate gene activations following transfers to the nucleus.

Rps10: *Rps10* has been transferred to the nucleus in several angiosperm species and multiple transfers to the nucleus are inferred to have occurred during angiosperm evolution (Adams et al. 2000). Here, we have identified four additional transferred *rps10* genes. *Rps10* in *Manihot* (cassava) has a short 5′ extension of the reading frame that is derived from a putative mitochondrial protein with a homolog in grasses (*Triticum* and *Oryza*). *Manihot* is in the same order, Malpighiales, as *Populus*, but *rps10* in *Populus* has presequence derived from succinyl-CoA ligase (figs. 1 and 2). We identified the source of the *rps10* presequence in *Arabidopsis* as a gene for a mitochondrial RNA helicase (fig. 2). *Rps10* from *Prunus* (cherry) does not contain a 5′ extension of the open reading frame that could be a presequence. A few other *rps10* genes have been transferred to the nucleus without gaining a mitochondrial presequence, and these gene products have internal targeting (Adams et al. 2000; Kubo et al. 2000). *Rps10* in *Ocimum* has a 5′ extension of the reading frame that is derived from a nonmitochondrial polyubiquitin gene, but it is not predicted to be a mitochondrial presequence. The open reading frame of *rps10* in *Ocimum* is truncated at the 3′ end, and it is missing a well-conserved region in other angiosperm *rps10* genes (supplementary fig. S2, Supplementary Material online); thus, the gene may be a transcribed pseudogene. *Rps10* in *Triticum* has a 5′ extension of the reading frame that is derived from cytosolic *rpl15*. Surprisingly, the *rpl15*-derived region is predicted to be a presequence. The structure of *rps10* in *Triticum* contrasts with the structures of *rps10* in *Zea* and *Oryza*, both of which do not contain a presequence (Adams et al. 2000; Kubo et al. 2000). The different structures of *rps10* in the three grass species could be the result of separate activation events, from acquisition of a presequence by *rps10* in *Triticum* after a single gene activation event in the common ancestor of these grasses, or loss of the *rpl15* chimeric structure in *Zea* and *Oryza* after a common activation event. To evaluate these three possibilities, we performed phylogenetic analysis of the *rpl15*-derived region of *rps10* from *Triticum* and *rpl15* genes from monocots. The *rpl15*-derived region from *rps10* branches with *rpl15* from *Triticum* indicating that it is a recent acquisition in the PACCAD clade since the divergence from the Bambusoideae, Ehrhartoideae, and Pooideae clade of grasses (supplementary fig. S3, Supplementary Material online).

Rps1: We identified and sequenced transferred *rps1* genes in four species: *Ocimum*, *Lactuca*, *Beta*, and *Citrus*. Only *rps1* in *Beta* contains a 5′ extension of the reading frame that is predicted to be a presequence; the other three genes have no 5′ extension of the reading frame like *rps1* in *Medicago* (Hazle and Bonen 2007). *Rps1* in *Medicago* was transferred to the nucleus after its divergence from a common ancestor with *Pisum*, making it a very recent evolutionary transfer (Hazle and Bonen 2007).

5′rpl2: We identified transferred 5′*rpl2* genes from five eudicots. (The *rpl2* gene is split in most eudicots; Adams, Ong, and Palmer 2001.) The gene structures are variable in different species (fig. 1): One contains a 5′ extension of the reading frame that is predicted to be a presequence, two contain a short 5′ extension of the open reading frame not predicted to be a presequence, and two have no extension of the reading frame and presumably internal targeting information like the previously reported 5′*rpl2* in *Glycine* (Adams, Ong, and Palmer 2001).

Rps14: We identified seven *rps14* genes transferred to the nucleus in eudicots. Only *rps14* in *Beta* has a 5′ extension of the open reading frame that is predicted to be a presequence. The targeting information of the other genes evidently is internal within the region transferred from the mitochondrion. *Rps14* is a pseudogene in the mitochondrion of *Citrus* (#CX071689), *Cucumis* (#AY258274), *Prunus* (#AY464900), and possibly *Ipomoea* as well (it is a pseudogene in tobacco; Sugiyama et al. 2005), and the gene is missing from the mitochondrion of *Beta* and *Euphorbia* (Adams et al. 2002). The transferred *rps14* genes in eudicots add to the previously reported transfers in monocots in which three separate transfers have been inferred (Figueroa et al. 1999; Kubo et al. 1999; Ong and Palmer 2006).

Rps7, *rps12*, and *rpl5*: We identified one or two transfers of three other ribosomal protein genes. *Rps7* in *Rhododendron* has a predicted presequence derived from

a putative mitochondrial protein with a homolog in *Vitis*, whereas *rps7* in *Lactuca* contains a predicted presequence derived from an unknown source. This is the first report of *rps7* genes transferred to the nucleus in plants. Surprisingly, we did not find a transferred nuclear copy in the NCBI databases for four angiosperms that have lost *rps7* from the mitochondrion and have large collections of ESTs: the Solanaceae (*Solanum*, *Lycopersicon*), *Malus*, *Glycine*, and *Citrus*. It is unclear if the gene has been transferred to the nucleus in those species or perhaps the protein has been substituted by another gene product. Finally, we identified and sequenced transferred *rps12* and *rpl5* from *Fragaria* (strawberry), one of which contains a presequence derived from another gene (figs. 1 and 2).

Table 1
Presence/Absence and Origin of Presequences for 70 Genes

Gene	Preseq. Mito.[a]	Preseq. Nonmito.[b]	Unknown Origin[c]	Absence[d]	5′ Ext.[e]
sdh3	8	1	1	—	—
sdh4	2	1	4	—	1
rps19	2	2	3	2	—
rps10	5	1	—	4	2
rps1	—	—	1	5	—
5′rpl2	—	—	1	5	2
rps14	1	—	2	6	1
rps7	1	—	1	—	—
rps12	1	—	1	—	—
rpl5	1	—	1	—	—
cox2	—	—	1	—	—
Total	21	5	16	22	6

5′ Extension refers an extension of the open reading frame at the 5′ end that was acquired after the gene was transferred to the nucleus. Genes with complete 5′ end extensions are included in this analysis.

[a] Presequence from mitochondria targeted gene.
[b] Presequence from non-mitochondria targeted gene.
[c] Unknown origin of presequence.
[d] No presequence or 5′ extension.
[e] 5′ Extension, but not a presequence.

Comparative Analysis of Targeting Sequences: Numerous Chimeric Genes

With a sizable data set of mitochondrial genes transferred to the nucleus, questions can be addressed about the frequency of different ways in which transferred genes become activated and expressed following transfer. Some important questions include: How many transferred genes gained a mitochondrial targeting presequence, and how many became activated without gaining a presequence? For those with presequences, how many were derived from pre-existing nuclear genes for mitochondrial proteins, and how many were derived from pre-existing genes for nonmitochondrial proteins? How many presequences were derived from unknown sources, possibly nongenic regions? For transferred genes with regions derived from pre-existing genes, how much and what region of the pre-existing gene has been incorporated into the transferred gene? To answer these questions, we undertook a comparative analysis of the structures of 70 transferred genes including 36 reported in this study and 34 that were published previously. In all but two cases, the full-length cDNA sequence was available.

Of the 42 genes with a presequence or putative presequence, 21 (50%) have a presequence derived from another nuclear gene for a mitochondrial protein (table 1). Thus, acquisition of a presequence from a pre-existing gene for a mitochondrial protein is a common mechanism by which newly transferred genes gain targeting elements and probably 5′ *cis*-regulatory elements. Remarkably, two nuclear genes for mitochondrial proteins have provided presequences for three or more transferred genes. The mitochondrial chaperonin *hsp70* donated a presequence in five cases of gene transfer and activation, including *sdh3* in *Ipomoea*, *Populus*, and *Arabidopsis*; *rps10* in *Fuchsia*; and *rps19* in *Zea*. The presequence from mitochondrial chaperonin *hsp22* was used for *sdh3* in *Gossypium* and *Triphysaria*, and *rps10* in *Daucus*. Other genes for mitochondrial proteins that have contributed presequences to newly transferred genes include genes for various enzymes and genes whose products are involved in translation.

How do newly transferred genes acquire a presequence from a pre-existing gene? One possibility is that the newly transferred gene could insert into another gene. Indeed there is evidence in three cases for direct insertion of a newly transferred gene into other genes. *Sdh3* in *Malus* and *rps10* in *Daucus* were inserted into mitochondrial EF-Tu and *hsp22*, respectively. *Rps14* in the Poaceae was inserted into an intron of a succinate dehydrogenase gene, and the genes are coexpressed by alternative splicing. Other cases where some of the core region (non–presequence region) of the host gene has been incorporated into the transferred gene structure also may represent cases of insertion of the transferred gene into the host gene.

How much of the host genes are incorporated into the chimeric structures of genes transferred to the nucleus? We examined 20 genes that contain a 5′ noncore region derived from another nuclear gene for a mitochondrial protein for which a complete open reading frame of the chimeric gene is available. We used Mitoprot and TargetP to predict the length of the presequence from the host genes and compared these regions with those of the chimeric transferred genes. Fifteen of the transferred genes contain the predicted presequence of the host gene, and many amino acids in addition (sometimes with deletions); in contrast only five genes (*rps10* in *Fuchsia*, *rps19* in *Malus* and *Zea*, and *rps7* in *Rhododendron*) appear to have acquired just the host gene's presequence or less than 10 amino acids in addition. Thus, transferred genes typically acquire more than just the presequence of the host gene. The patterns are illustrated in the five genes containing regions derived from the chaperonin *hsp70* (supplementary fig. S4, Supplementary Material online). *Rps10* in *Fuchsia* contains the entire presequence and *rps19* in *Zea* contains almost all of the *hsp70* presequence region (lacking eight amino acids). In contrast, *sdh3* from *Arabidopsis* and *Ipomoea* contain the presequence region plus numerous additional amino acids as well as deletions. By far the longest *hsp70*-derived region is from *sdh3* in *Populus*: It has a ~300 amino acid deletion but then contains more of the *hsp70*-derived region to include a considerable amount of the *hsp70* core region overall.

Presequences acquired from genes for nonmitochondrial proteins account for 5 of 42 cases of genes with presequences (12%). These genes include *rps19* in *Arabidopsis* and *Populus*, *sdh3* in *Medicago*, *sdh4* in *Euphorbia*, and

rps10 in *Triticum*. Thus, this mechanism appears to be rarely used during transferred gene activation. That is not surprising considering that most genes for nonmitochondrial proteins will not contain elements for mitochondrial targeting. Coincidentally, two of the transferred *rps19* genes (from *Arabidopsis* and *Populus*) have independently acquired a presequence from a glycine-rich RNA-binding protein (Choi et al. 2006). A homologous RNA-binding protein gene from tobacco was shown to be localized to the nucleoplasm (Moriguchi et al. 1997), suggesting that the *Arabidopsis* RNA-binding protein does not function in the mitochondrion. Thus, there may have been certain mutations that occurred at the 5′ end of the RNA-binding protein gene after association with *rps19* that allowed the gain of mitochondrial targeting. The noncore region of *rps10* in *Triticum* was derived from a cytosolic ribosomal protein, and the noncore region is predicted to serve as a presequence. *Sdh4* in *Euphorbia* has a noncore region that has some similarity to a putative glutamate binding protein that is not predicted to be mitochondrial targeted, and *sdh3* in *Medicago* has a noncore region derived from a cytosolic aconitase gene.

The remaining 15 of 42 cases (36%) have a presequence derived from an unknown source; the percentage would be higher, 21 of 47 (45%), if the 6 genes are included whose 5′ ends are incomplete but may well contain a presequence derived from an unknown source. The sources of all the acquired presequences in transferred genes of *Populus* have been identified, but the donor of the presequence of *rps14* and *sdh4* in *Arabidopsis*, as well as *sdh3* and *sdh4* in rice, have not been identified despite the complete or mostly complete genome sequences. We found that the presequence of *sdh4* in *Arabidopsis* has a low level of identity (24% over 67 amino acids with three gaps) to a gene for a PPR protein (#CAB94132), but it might not be the actual donor. It is possible that some of the presequences whose sources cannot be identified might be derived from a sequence that is now part of the open reading frame of the transferred gene, and thus, no similar region is present elsewhere in the genome. It is also possible that the donor or recipient sequence (or both) has diverged too much for identification. In those species whose genome has not been sequenced, the presequences derived from unknown sources may represent presequences derived from a lineage-specific gene without a homolog in a species whose genome has been sequenced. A final possibility is that presequences in transferred genes were acquired from other sequences in the genome that did not possess targeting activity and the targeting function was gained de novo. Mitochondrial presequences vary considerably in primary sequence, but overall tertiary structural features, such as the ability to form an amphiphilic alpha helix, appear to be conserved (reviewed in Glaser et al. 1998). The sequence flexibility of presequences is highlighted by the finding that 2.5% of *Escherichia coli* clones generated in a shotgun screen exhibited mitochondrial targeting activity when added to a truncated yeast gene for cytochrome oxidase subunit 4 (Baker and Schatz 1987). Thus, gain of mitochondrial targeting ability by certain sequences de novo is not difficult to envision, at least for some mitochondrial proteins.

Twenty-two of the 70 transferred genes (31%) have no 5′ extension of the reading frame that could act as a mitochondrial presequence, as judged by the presence of an in-frame stop codon upstream of the first methionine in the core region. The targeting information presumably is located somewhere within the core region. The frequency of genes with no presequence is consistent with a study of all mitochondrial ribosomal proteins transferred to the nucleus in *Arabidopsis* and rice (Bonen and Calixte 2006), most of which were transferred to the nucleus before the evolution of angiosperms that identified 25% of the genes with no presequences. Recently, it was shown that the ribosomal proteins 5′*rpl2* and *rps4* from *Arabidopsis*, and *rpl2* and *rps19* from rice, can be targeted to mitochondria even though the genes that code these proteins are present in the mitochondrial genome (Ueda et al. 2008). Thus, when one of those genes is transferred to the nucleus it would not need to acquire targeting presequence, relieving one of the requirements for gene activation, and potentially allowing more transfers to take place during evolution. Most of the 5′*rpl2* genes studied here do not contain a presequence or predicted presequence (fig. 1), and two of the *rps19* genes studied here do not contain a presequence. It is possible that other genes in this study that do not contain presequences, including *rps10*, *rps14*, and *rps1*, did not need to gain a presequence for gene activation in the nucleus because the gene transferred from the mitochondrion had the necessary targeting information. Mitochondrial copies of *rps1*, *rps14*, and *rps10* were not tested by Ueda et al. (2008) for targeting ability. However, the product of mitochondrial *rps10* in soybean has been shown not to have mitochondrial targeting activity unless certain amino acids at the 5′ end are mutated (Murcha et al. 2005).

Many of the genes described above have different noncore regions in different species, often as chimeras from different nuclear genes, or the gene in some species lacks a noncore region. The unique structural features could have been derived from separate gene activations, likely following separate gene transfers to the nucleus. Alternatively, they could have resulted from extensive recombinational events in each lineage after a single, relatively ancient gene activation following transfer to the nucleus. Another possibility is gene duplication after transfer to the nucleus, before gene activation, and acquisition of different noncore regions by each copy, followed by retention of different copies in different lineages. To distinguish among these possibilities, we analyzed the noncore regions of genes transferred to the nucleus before the evolution of angiosperms. We analyzed 28 ribosomal protein genes, including all of those identified by Bonen and Calixte (2006) that have a 5′ noncore region and two succinate dehydrogenase genes (*sdh1* and *sdh2*). Also, we analyzed two ribosomal protein genes that were transferred to the nucleus near the base of eudicots (3′*rpl2* and *rps11*). We selected seven eudicots from each of the following genera or families from both rosids and asterids: *Arabidopsis*, *Gossypium*, *Populus*, Rosaceae (*Malus*, *Prunus*, or *Fragaria*), Fabaceae (*Glycine* or *Medicago*), *Solanum*, and Asteraceae (*Lactuca*, *Centaurea*, or *Helianthus*). Sequences upstream of the core region for each of the 32 genes and each of the seven taxa (if sequences were available) were aligned, and their noncore regions, including presequences, were compared (supplementary fig. S5, Supplementary Material online).

Almost all of the 30 genes transferred to the nucleus before the evolution of angiosperms, and the two genes transferred at the base of eudicots have noncore regions that are easily alignable, moderately to strongly conserved, and clearly homologous (supplementary fig. S5, Supplementary Material online). The only exceptions are *rpl19*, where the sequences from the seven eudicots show minimal similarity and the noncore region of *rpl19* in *Solanum* is derived from cytochrome c reductase; *rpl15* in *Solanum* whose noncore region is not similar to another gene; and *rps5* in the eudicots, especially in *Lactuca*, although none of the noncore regions show similarity to other genes. There was no sequence similarity among the *rps11* noncore regions in the eudicots, but there is an intron at about the same position downstream of the start codon in each gene suggesting the possibility of extensive sequence divergence following a single origin; *rps11* in *Solanum* has a presequence derived from *rps18* (also observed by others—Kubo N, personal communication) but none of the others have a noncore region derived from another gene.

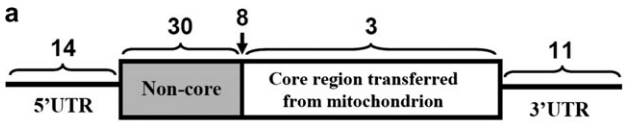
The high conservation of the noncore sequences in anciently transferred ribosomal protein and succinate dehydrogenase genes contrasts with the lack of noncore sequence conservation of genes transferred to the nucleus during angiosperm evolution (fig. 1). The above analysis best supports the hypothesis that the different noncore regions in different lineages for the same gene were derived from separate gene activations in each lineage, probably following separate and relatively recent evolutionary transfers to the nucleus, for the following genes: *sdh3*, *sdh4*, *rps19*, *rps10*, *rps7*, *rpl5*, and *rps12*. Multiple transfers and activations have been inferred previously for five of the preceding genes (Adams et al. 2000; Adams, Rosenblueth, et al. 2001; Adams et al. 2002; Sandoval et al. 2004), and the analysis presented here gives additional support for those inferences.

Changes in the noncore region by shuffling or recombination following gene transfer and activation appear to be rare. From our analysis, we detected only two obvious examples, *rpl19* and *rps11* in *Solanum*, and possibly *rps10* in *Triticum*. Gene duplication followed by the gain of different noncore regions by the duplicates also appears to be rare. *Rps11* in rice (Kadowaki et al. 1996) is the only reported example among evolutionarily recent gene transfers, but two anciently transferred genes in *Arabidopsis* (*rpl7*/*rpl12* and *sdh2*) have two or four copies with different presequences. Surprisingly, the noncore regions of three of the four copies of the *rpl7*/*rpl12* gene in *Arabidopsis* align well to the noncore region in other eudicots suggesting a complicated pattern of gene duplication and noncore region acquisition (supplemental fig. S6, Supplementary Material online). Of the three copies of *sdh2* in *Arabidopsis*, two have similar sequences and are expressed in different patterns, but expression of the third copy with the different noncore sequence was not detected (Figueroa et al. 2001; Elorza et al. 2004), and it might be a pseudogene.

Variable Locations of Introns in Transferred Genes

Where are introns located in genes transferred to the nucleus during recent angiosperm evolution? How common is intron-mediated exon shuffling for the potential acquisition of presequences or regulatory elements? To answer these questions, we analyzed a data set of 62 evolutionarily recent gene transfers for which genomic DNA and cDNA sequences are available, including 37 genomic DNA sequences generated in this study and 25 published sequences. Five possible locations of introns include: in the 5′UTR, in the noncore region of the open reading frame (including the presequence, if present), between the noncore and the core regions, in the core region (the sequence transferred from the mitochondrion), and in the 3′UTR (fig. 4). The most common position for introns is within the noncore region, with 30 of 62 genes (48%) having one or more intron(s) in that region. The second most common position for introns is within the untranslated regions (UTRs). Fourteen of 54 genes for which genomic DNA sequence is available in the 5′UTR have an intron in the sequenced region (26%), and 11 of 57 genes for which the genomic DNA sequence is available in the 3′UTR have an intron in the sequenced region (19%). The number of genes with introns in the UTRs may well be higher because the genomic DNA sequences are not complete in the UTR regions in many cases; this issue might affect the number of introns in the 5′UTRs more than the number in the 3′UTRs because the sequenced regions of the 3′UTRs are frequently longer than the 5′UTRs. Eight of 62 genes (13%) have an intron located between the core and noncore regions. We counted an intron between the core and noncore region if it was at the junction or within 30 nucleotides upstream of the core/noncore junction. Only three genes have an intron in the core region: *5′rpl2* in *Beta* and *sdh4* in *Ocimum* and *Euphorbia*.

What are the implications of the locations of introns in the transferred genes? Introns that are present in the UTR regions, or between the presequence and the core region, support the possibility of an exon shuffling–type recruitment of sequences that confer expression and/or targeting



b

Gene	5'UTR	Non-core region	Between non-core & core	Core region	3'UTR	No introns
sdh3	1	13	0	0	2	0
sdh4	0	3	0	2	1	1
rps19	3	3	4	0	1	2
rps10	3	5	0	0	2	2
rps1	2	0	0	0	0	3
5′rpl2	2	0	2	1	0	0
rps14	1	2	1	0	5	0
rps7	1	1	0	0	0	0
rps12	0	1	0	0	0	0
rpl5	0	2	0	0	0	1
cox2	1	0	1	0	0	0
Total	14	30	8	3	11	9

FIG. 4.—Locations of introns in genes transferred to the nucleus. The number of genes with one or more introns in each region is indicated.

in the process of gene activation (Bonen and Calixte 2006). In our data set, there are only 8 (or 10, if 2 of the *sdh4* genes with introns in the poorly conserved 5′ end are counted) cases of an intron located between the core and noncore regions, suggesting that the acquisition of the noncore region by exon shuffling is an uncommon mechanism. In contrast, a study of all mitochondrial ribosomal protein genes transferred to the nucleus in *Arabidopsis* and rice showed that 28% of the genes had an intron between the core–noncore region (Bonen and Calixte 2006). Introns in the 5′UTR may have been involved in the acquisition of 5′ *cis*-regulatory elements, such as promoters, that are essential for a newly transferred gene to become active in the nucleus. Likewise, introns in the 3′UTR may have allowed acquisition of 3′ *cis*-regulatory elements, such as elements conferring transcript stability. However, there is also the possibility that introns could be gained after the gene has become an active gene in the nucleus and thus not be involved in acquisition of presequences and regulatory elements; conversely, loss of introns after gene activation would result in intron-mediated sequence acquisitions that are undetectable.

Accelerated Sequence Evolution in Genes Transferred to the Nucleus

The relatively large number of genes transferred to the nucleus in angiosperms allows for comparative analysis of transferred gene sequence evolution, to address questions including: How much has the rate of nucleotide substitutions increased after transfer to the nucleus, and do both synonymous and nonsynonymous substitutions increase? How do substitution rates compare among different transferred genes in the same species and in different species for the same gene? Is there a correlation between K_s and inferred timing of transfer? To answer these questions, we compiled and analyzed a data set of 105 transferred genes from a variety of angiosperm species. We performed pairwise analysis of synonymous (K_s) and nonsynonymous (K_a) substitutions using the ML method. Three comparisons were done for each transferred nuclear gene from eudicots: Nuclear gene versus mitochondrial gene from a representative of the basal eudicot lineage Ranunculales (*Eschscholzia*, *Ranunculus*, or *Dicentra*) as an outgroup to all core eudicots; nuclear gene versus the mitochondrial gene from the basal angiosperms *Nymphaea* or *Amborella* as an outgroup to all other angiosperms; and nuclear gene versus a full-length mitochondrial gene from a eudicot (frequently from *Nicotiana*) because many of the basal eudicot and basal angiosperm sequences were missing short regions at each end (where PCR primers were located). Rate analyses for the monocots were based on comparisons to the basal angiosperms *Nymphaea* or *Amborella*, and either to the mitochondrial copy in another monocot species if a sequence is available or to the basal eudicot representative. Many of the sequences from *Eschscholzia* and *Nymphaea* were generated in this study by PCR amplification and sequencing (see Methods), and the other mitochondrial sequences were obtained from GenBank (supplementary table S1, Supplementary Material online). To assess the effects of C-to-U RNA editing, we did K_s and K_a analyses for *rps19* and *sdh3* either including or excluding sites that are RNA edited in mitochondrial sequences from other eudicots. The differences in K_s and K_a values were minimal; use of different reference taxa for comparison had a larger effect on K_s and K_a values than excluding RNA editing sites (data not shown).

The three genes transferred to the nucleus early during eudicot evolution (at the base of eudicots), 3′*rpl2*, *rps2*, and *rps11* (Adams, Ong, and Palmer 2001; Adams et al. 2002), showed the highest K_s values of all the genes (supplementary table S3, Supplementary Material online). K_s over 2.0 has been inferred to indicate K_s saturation, which is associated with the uncertain estimation of K_s (Blanc and Wolfe 2004). For these three genes, the K_s values in almost every eudicot species analyzed are over 2.0, and K_s is saturated. Another gene where K_s is over 2.0 is *sdh3* in the grasses. *Sdh3* probably was lost functionally from the mitochondrion early during monocot evolution after being transferred to the nucleus, as judged by a DNA gel blot survey (Adams et al. 2002), and we identified transferred nuclear *sdh3* genes in other monocots including two species from the Asparagales (*Allium*: CF440196 and *Phalaenopsis*: CB031893, CB034425) and two species from the Zingiberales (*Zingiber*: DY362466 and *Curcuma*: DY385363). The high K_s for *sdh3* in the grasses is consistent with the relatively ancient transfer to the nucleus. Other genes with high K_s include *rps1* in *Beta* and *Ocimum* (table 2). *Rps1* has been lost from the mitochondrion of most of the Lamiales order and the loss of *rps1* encompasses a major group of the Caryophyllales (Adams et al. 2002), both consistent with an older transfer to the nucleus than many of the other genes. In contrast to the genes mentioned above, two evolutionarily very recent cases of transfer, *rps1* in *Medicago* (Hazle and Bonen 2007) and *cox2* in the Glycininae tribe of legumes (Adams et al. 1999) both have low K_s rates (below 0.5) when compared with other eudicots. Many other cases of transfer have K_s below 0.5 (table 2), consistent with the hypothesis of recent transfers to the nucleus.

There is considerable variability in K_s rates among different species in which the same gene has been transferred to the nucleus (table 2). For example, *rps10* in four species has K_s less than 0.5; in five species, the K_s is between 0.5 and 1.0; and in three species, it is over 1.0. *Sdh4* shows extreme variability, with *Populus* at 0.14 and five species over 2.0; the high K_s rates in several species likely are caused by a divergent 5′ end whereas the 3′ end of *sdh4* maintains higher sequence conservation (supplementary fig. S2, Supplementary Material online). We also compared K_s rates among genes in the same species or family. Four of the five transferred genes in *Arabidopsis* had similar K_s, ranging from about 0.9 to 1.3. Three of these genes were lost from the mitochondrion or became pseudogenes in a common ancestor of *Arabidopsis* and Brassicaceae after divergence from the Caricaceae, suggesting that transfers might have occurred within a similar window of time. All five transferred genes from the Chenopodiaceae (*Beta* and *Spinacia*) have K_s values over 1.0. In contrast, all six transferred genes from *Populus* and three from *Citrus* have K_s values below 0.8. These findings are consistent with the lower reported rate of sequence evolution in *Populus* than in *Arabidopsis*

Table 2
K_s and K_a of Mitochondrial Genes Transferred to the Nucleus

Gene	K_s				K_a		
	<0.5	0.5–1.0	1.0–2.0	Saturated (>2.0)	<0.25	0.25–0.5	>0.5
Sdh3	*Gossypium* (0.30) *Malus* (0.24) *Populus* (0.44) *Triphysaria* (0.36) *Vaccinium* (0.32)	*Aquilegia* (0.68) *Arabidopsis* (0.87) *Euphorbia* (0.92) *Ipomoea* (0.54) *Medicago* (0.60) *Ocimum* (0.80)	*Beta* (1.69) *Lactuca* (1.12)	*Oryza* (16.33)*	*Gossypium* (0.23) *Malus* (0.21)	*Aquilegia* (0.48) *Beta* (0.48) *Euphorbia* (0.32) *Ipomoea* (0.32) *Medicago* (0.30) *Ocimum* (0.43) *Populus* (0.31) *Triphysaria* (0.32) *Vaccinium* (0.32)	*Arabidopsis* (0.54) *Lactuca* (0.67) *Oryza* (0.59)*
Sdh4	*Populus* (0.14)		*Arabidopsis* (1.29) *Glycine* (1.05)	*Euphorbia* (2.12) *Lactuca* (3.49) *Ocimum* (2.58) *Oryza* (9.82)* *Papaver* (6.23)	*Populus* (0.10)		*Arabidopsis* (0.68) *Euphorbia* (1.42) *Glycine* (0.70) *Lactuca* (0.70) *Ocimum* (0.78) *Oryza* (0.98)* *Papaver* (0.74)
rps19	*Citrus* (0.46) *Cucumis* (0.14)	*Malus* (0.89) *Populus* (0.64) *Zea* (0.79)*	*Glycine* (1.73) *Gossypium* (1.04)	*Acorus* (5.29) *Arabidopsis* (3.59) *Euphorbia* (2.94)	*Arabidopsis* (0.19) *Citrus* (0.21) *Cucumis* (0.19) *Zea* (0.19)*	*Euphorbia* (0.31) *Glycine* (0.30) *Gossypium* (0.27) *Malus* (0.28) *Populus* (0.34)	*Acorus* (0.62)
rps10	*Manihot* (0.44) *Oxalis* (0.45) *Populus* (0.30) *Prunus* (0.27)	*Daucus* (0.59) *Fuchsia* (0.71) *Lactuca* (0.76) *Ocimum* (0.55) *Triticum* (0.58)*	*Arabidopsis* (1.18) *Spinacia* (1.29) *Zea* (1.41)*		*Fuchsia* (0.23) *Lactuca* (0.25) *Oxalis* (0.19) *Populus* (0.18) *Prunus* (0.15)	*Arabidopsis* (0.27) *Daucus* (0.46) *Manihot* (0.29) *Spinacia* (0.31) *Triticum* (0.31)* *Zea* (0.32)*	*Ocimum* (0.45)
rps1	*Citrus* (0.30) *Medicago* (0.19)	*Gossypium* (0.58) *Lactuca* (0.74)		*Beta* (2.61) *Ocimum* (4.73)	*Beta* (0.22) *Citrus* (0.14) *Lactuca* (0.20) *Medicago* (0.14)	*Gossypium* (0.31) *Ocimum* (0.45)	
5′rpl2	*Citrus* (0.41)	*Ocimum* (0.97) *Triphysaria* (0.78) *Triticum* (0.90)*	*Beta* (1.77)	*Glycine* (2.14) *Lactuca* (2.05) *Zea* (2.15)*	*Citrus* (0.17)	*Beta* (0.26) *Glycine* (0.36) *Lactuca* (0.48) *Triphysaria* (0.38) *Triticum* (0.30)*	*Ocimum* (0.71) *Zea* (0.46)*
rps14	*Joinvillea*(0.14)*	*Carex* (0.71)* *Citrus* (0.77) *Lactuca* (0.65) *Populus* (0.77)	*Arabidopsis* (1.19) *Beta* (1.74) *Cucumis* (1.83) *Ipomoea* (1.81) *Oryza* (1.30)*	*Prunus* (2.62) *Euphorbia* (2.08)	All taxa		
rps7		*Lactuca* (0.59)	*Rhododendron* (1.08)		*Lactuca* (0.17) *Rhododendron* (0.13)		
rps12		*Fragaria* (0.52) *Oenothera* (0.77)			*Fragaria* (0.11) *Oenothera* (0.09)		
rpl5		*Zea* (0.96)*	*Fragaria* (1.56) *Triticum* (2.00)*			*Fragaria* (0.27) *Triticum* (0.29)* *Zea* (0.26)*	
cox2	*Glycine* (0.43)				*Glycine* (0.08)		

Numbers are average values from two or three comparisons; see supplementary table 2, Supplementary Material online, for more details and for the genes *rps2*, *rps11*, and 3′*rpl2*. Asterisks indicate monocots.

(Tuskan et al. 2006). Other species show a considerable range of K_s rates among different transferred genes, such as the seven genes in *Lactuca* and five genes in *Ocimum*.

Nonsynonymous substitution rates (K_a) generally showed a narrower range of variability than K_s. Excluding *sdh4* in several species, most were below 0.5 and none were above 0.71 (table 2). *Rps14* genes had especially low K_a values with all being under 0.3; that is in stark contrast to the high K_s values in most *rps14* genes. All *rps12* and *rps7* genes also had low K_a values. In contrast, K_a in most *sdh4* genes was high due to the very divergent 5′ end of these genes compared with the mitochondrial *sdh4* and to each other (supplementary fig. S2, Supplementary Material online). K_a rates were expected to be variable among genes because different proteins will experience different selective constraints. Two genes transferred to the nucleus during recent legume evolution, *rps1* and *cox2*, showed very low K_a rates (less than 0.15). K_a for the three genes transferred to the nucleus at the base of eudicots (*rps2*, *rps11*, and *3′rpl2*) showed K_a rates that were comparable with many of the more recently transferred genes (supplementary table S3, Supplementary Material online), suggesting that nonsynonymous rates reach a threshold after a gene has been in the nucleus for a period of time.

To compare the K_a and K_s rates of the transferred nuclear copies with the mitochondrial copies in other species, we calculated K_a and K_s in pairwise comparisons (as was done for the nuclear copies) using several angiosperm sequences for most genes. When comparing K_a and K_s for the mitochondrial copies, the rates were considerably lower (supplementary table S4, Supplementary Material online), in contrast to the genes transferred to the nucleus (supplementary table S3, Supplementary Material online). For each gene, K_s of all the nuclear copies was higher than K_s of all the mitochondrial copies. For almost all genes, the K_a of the nuclear copies was higher than K_a of any of the mitochondrial copies, with exceptions being *rps1* in *Medicago* and *Citrus*. These results indicate that both K_a and K_s increase considerably after a gene is transferred to the nucleus.

Previous analysis of nucleotide substitution rates in genes transferred to the nucleus have been limited to *cox2* in legumes and *rps12* in *Oenothera* (Laroche et al. 1997), *rps14* in grasses (Ong and Palmer 2006), and *rps10* in several angiosperms (Adams et al. 2000). Those studies showed an increased rate of substitutions especially at silent sites, consistent with the higher rate of substitutions in the nuclear genome (Wolfe et al. 1987). Our analysis of a large data set of transferred genes in a variety of angiosperms has confirmed those findings and revealed additional patterns of rate evolution among genes and taxa.

Analysis of Positive Selection in Transferred Genes

We calculated pairwise K_a/K_s ratios to determine if any genes have experienced positive selection. Evidence for positive selection is provided by a K_a/K_s ratio over 1.0. Only one gene showed a K_a/K_s ratio above 1.0: *rps19* in *Cucumis*. We used branchwise K_a/K_s analysis to evaluate the possibility of positive selection on individual codons based on the phylogenetic context. The branch-site model in PAML was used to detect if there has been adaptive molecular evolution in a specific lineage (Yang and Nielson, 2002). We selected several transferred nuclear genes with reasonably well-conserved sequences across the entire core region and genes/species for which mitochondrial copies were available from relatively closely related taxa. We identified three genes that show evidence for codon-specific positive selection (see Methods): *rps10* in *Daucus*, *Manihot*, and *Ocimum* (table 3). There are 4, 2, and 5 positively selected sites inferred in *Daucus*, *Manihot*, and *Ocimum*, respectively. None of these are codons modified by RNA editing in the mitochondrial copies *rps10* in other species. We used strict criteria to identify codons showing positive selection to avoid false positives, and thus, we could be underestimating the number of positively selected sites.

In addition to the detection of positively selected sites in the above three genes, there are three major types of rate evolution found in the rest of the genes that were analyzed (table 3): 1) Most sites of the sequences are under purifying selection ($K_a/K_s < 1$), such as *cox2* in legumes, *rps1* in *Ocimum*, *Lactuca*, and *Medicago*, etc. 2) There is nearly a one-to-one ratio between purifying selection and neutral evolution, such as *rps1* in *Gossypium* and *Citrus*, *rps10* in *Spinacia* and *Prunus*, etc. 3) Many sites in the sequence are under neutral evolution, including *sdh4* and *rps10* in *Populus*. For the first and second cases, nearly 50% or more of the sites are under purifying selection, indicating that the genes are still under functional constraint. For the last case, the high portion of sites undergoing neutral evolution could indicate either relaxed purifying selection or pseudogenization. For *sdh4* in *Populus*, there is another expressed copy in the mitochondria (Choi et al. 2006), and thus, the nuclear copy might be an expressed pseudogene. However, for *rps10* in *Populus*, there is no evidence for the presence of a copy in the mitochondrion (Choi et al. 2006).

Due to asexual reproduction, genes in mitochondria have been hypothesized to accumulate many deleterious mutations by Muller's Ratchet (e.g., Lynch 1996). Once a mitochondrial gene has been transferred to the nucleus, it would experience a higher recombination rate that facilitates the loss of deleterious mutations and fixation of beneficial mutations, thus releasing mitochondrial genes from Muller's ratchet effect (Blanchard and Lynch 2000). Based on that hypothesis, positive selection might be acting on mitochondrial genes transferred to the nucleus. However, considering that 1% (1 of 106) of genes in pairwise comparisons and 10% (3 of 31) of genes in branchwise comparisons show evidence of positive selection, positive selection for beneficial mutations or to eliminate deleterious mutations is not common in genes transferred to the nucleus.

GC Content and Codon Usage in Genes Transferred to the Nucleus

We analyzed genes transferred to the nucleus for changes in GC content and codon usage compared with their mitochondrial counterparts. The GC content in the silent position (GC3s) and the ENC were calculated. Our results show a trend toward increased GC3s for genes transferred to the nucleus ($P < 0.05$; fig. 5). Also, there

Table 3
Selection and LRT Statistics of the Branch Site Model for Specific Lineages

Gene	Taxon	Purifying Selection (%, $K_a/K_s < 1$)	Neutral Evolution (%, $K_a/K_s = 1$)	Positive Selection (%, $K_a/K_s > 1$)	$M1–MA_{test1}$		$MA_{test1}–MA_{test2}$		Positively Selected Sites
					$2\delta L$	*P* Value	$2\delta L$	*P* value	
cox2	*Glycine*	71.76	28.24	0.00	0.0000	1.0000	0.0000	1.0000	None
rpl5	*Fragaria*	66.73	29.97	3.30	0.2952	0.8628	0.2726	0.6016	None
rpl5	*Triticum*	76.46	23.54	0.00	0.0000	1.0000	0.0000	1.0000	None
rpl5	*Zea*	98.67	0.38	0.95	0.0359	0.9822	0.0359	0.8497	None
rps1	*Citrus*	42.01	54.27	3.72	5.7698	0.0559	5.7348	0.0166*	None
rps1	*Gossypium*	50.01	42.38	7.61	2.9321	0.2308	2.9023	0.0885	None
rps1	*Lactuca*	73.00	27.00	0.00	0.0000	1.0000	0.0000	1.0000	None
rps1	*Medicago*	100.00	0.00	0.00	0.0000	1.0000	0.0000	1.0000	None
rps1	*Ocimum*	66.59	33.41	0.00	0.0000	1.0000	0.0000	1.0000	None
rps7	*Lactuca*	47.60	52.40	0.00	4.6273	0.0989	0.0000	1.0000	None
rps7	*Rhododendron*	92.03	7.97	0.00	0.7565	0.6851	0.0000	1.0000	None
rps10	*Arabidopsis*	80.25	19.75	0.00	0.8294	0.6605	0.0000	1.0000	None
rps10	*Daucus*	40.33	27.04	32.63	11.0323	0.0004*	9.6220	0.0019*	19A, 22Q, 23T, 76A
rps10	*Fuchsia*	94.38	0.00	5.62	3.8686	0.1445	1.6878	0.1939	None
rps10	*Lactuca*	61.53	38.47	0.00	0.0000	1.0000	0.0000	1.0000	None
rps10	*Manihot*	59.74	28.22	12.04	9.3907	0.0091*	9.1687	0.0025*	20A, 23S
rps10	*Ocimum*	56.69	23.62	19.79	19.4388	0.0001*	16.7500	0.0000*	32N, 38S, 72A, 85S, 87G
rps10	*Oxalis*	39.91	60.09	0.00	1.2557	0.5337	0.0001	0.9920	None
rps10	*Populus*	25.01	74.99	0.00	3.6558	0.1607	0.0925	0.7610	None
rps10	*Prunus*	42.76	57.24	0.00	1.3761	0.5026	0.0000	1.0000	None
rps10	*Spinacia*	56.97	33.26	9.77	3.6098	0.1645	3.6098	0.0574	None
rps10	*Triticum*	50.05	30.60	19.35	5.0419	0.0804	4.3771	0.0364*	None
rps10	*Zea*	50.92	49.08	0.00	0.3006	0.8604	0.0000	1.0000	None
rps12	*Fragaria*	70.46	19.07	10.48	3.9202	0.1409	0.5343	0.4648	None
rps12	*Oenothera*	83.16	16.84	0.00	0.0129	0.9936	0.0000	1.0000	None
rps14	*Carex*	65.56	34.44	0.00	0.0000	1.0000	0.0000	1.0000	None
rps14	*Oryza*	73.63	26.37	0.00	0.0000	1.0000	0.0000	1.0000	None
rps19	*Zea*	56.24	43.76	0.00	7.6778	0.0215*	0.0000	1.0000	None
sdh4	*Populus*	20.44	79.56	0.00	0.0002	0.9999	0.0002	0.9887	None

* Asterisk indicates a statistically significant difference ($P < 0.05$). Numbering of positively selected sites corresponds to alignment positions in supplementary figure S6, Supplementary Material online.

is a significant positive correlation between GC3s content and K_s ($P < 0.05$; fig. 5), suggesting that the more time genes are present in the nucleus, the higher their GC3s content. Especially for more anciently transferred genes such as *3′ rpl2*, *rps2*, and *rps11*, most nuclear genes have higher GC3s than their mitochondrial counterparts (supplementary table S3, Supplementary Material online). In contrast, for several recently transferred genes such as *cox2* in *Glycine*, *rps1* in *Medicago*, and *rps12* in *Fragaria*, there is not much difference in GC3s between transferred nuclear genes and mitochondrial genes (supplementary table S3, Supplementary Material online). In few cases, such as *sdh3* in *Medicago* and *Vaccinium*, slightly lower GC3s were found in the nuclear copy. In contrast, GC3s is generally higher in transferred genes of grasses. Previous studies of the transferred *rps19* and *rpl2* genes in wheat (*Triticum*) indicated that the nuclear gene shows a preference for G + C in the silent position, whereas mitochondrial genes show a bias for A + T in the silent position (Fallahi et al. 2005; Subramanian and Bonen 2006). The elevation of GC3s suggests that genes gradually adapt to the nuclear GC content environment after transfer to the nucleus.

ENC values range from 20 to 61, the former indicating an extreme bias and the latter showing equal usage of synonymous codons (Wright 1990). We analyzed ENC in the data set of genes transferred to the nucleus. In contrast to GC3s values, no significant differences in ENC values were observed between transferred nuclear genes and mitochondrial genes ($P > 0.05$; supplementary fig. S7, Supplementary Material online). In addition, there is no correlation between synonymous substitutions and ENC ($P > 0.05$; supplementary fig. S7, Supplementary Material online). These results indicate that there is no change in codon usage for genes transferred to the nucleus in angiosperms.

Transfers to the Nucleus of Angiosperm Chloroplast Genes and Mitochondrial Genes in Nonangiosperm Land Plants

In this study, we analyzed features of gene structures and sequence evolution in 77 mitochondrial genes transferred to the nucleus in a variety of angiosperm species that reflect many transfers to the nucleus. How common are functional transfers in nonangiosperm land plants? Compared with the mitochondrion of angiosperms, in which many species have lost several genes (Adams et al. 2002), nonangiosperm mitochondria have lost few genes as far as examined. The completely sequenced mitochondrial genome of the cycad *Cycas taitungensis* contains 39 protein-coding genes and only appears to

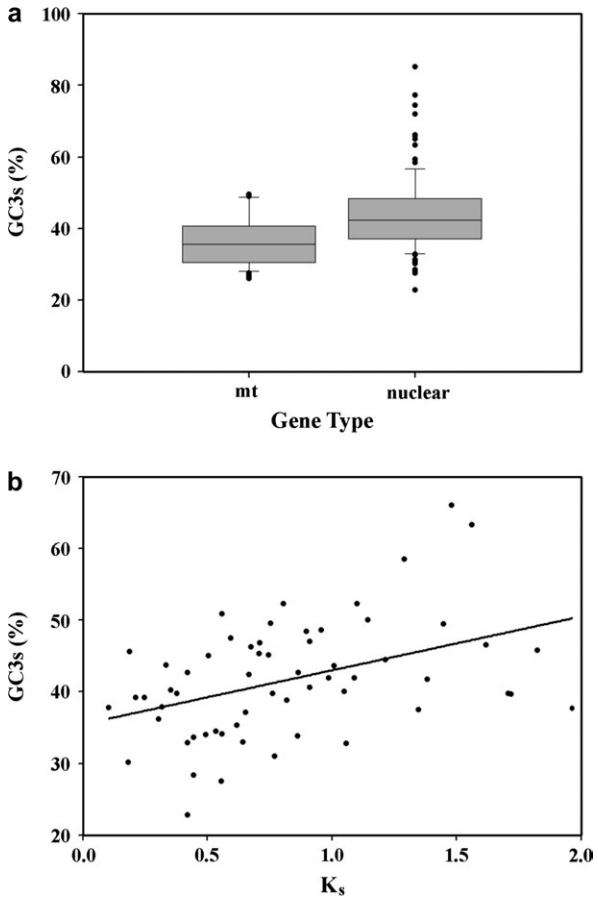


FIG. 5.—GC content analysis. (*a*) Boxplot of GC content at third positions (GC3s) between mitochondrial genes (mt) and genes transferred to the nucleus (nuclear). The outliers are indicated by dots. *t*-test: $t = -5.3245$; $P = 5.357e-07$. Wilcoxon: $W = 1039$; $P = 9.465e-06$. (*b*) Correlation between K_s and GC3s for transferred genes. Pearson's correlation test: $r = 0.42$; $P = 8.275e-03$.

have lost *sdh4* among those genes present in the common ancestor of land plants (Chaw et al. 2008). Thus, there may only be one gene that has been transferred to the nucleus in *Cycas*. The mitochondrial genomes of two nonvascular land plants have been sequenced. *Marchantia polymorpha* (a liverwort) lacks a functional copy only of *nad7* (Oda et al. 1992), and a functional copy is present in the nucleus (Kobayashi et al. 1997). *Physcomitrella patens* (a moss) reportedly lacks *rps10 and rps8* (Terasawa et al. 2007). To determine if *rps10* and *rps8* have been transferred to the nucleus, we searched the NCBI databases and found several ESTs of *rps10* (#FC441329, etc.) that match a sequence (#XM_001751279) from the nuclear genome sequence (Rensing et al. 2008) indicating that *rps10* has been transferred to the nucleus in *Physcomitrella patens*. Overall, it appears that transfer of mitochondrial genes to the nucleus and activation to become functional genes is not common in nonangiosperm land plants, although the number of investigated species is very small.

How common are functional transfers of chloroplast genes to the nucleus in angiosperms? The chloroplast genomes of angiosperms have a relatively constant set of protein-coding genes, with a few exceptions (Jansen et al. 2007), compared with angiosperm mitochondrial genomes (Adams et al. 2002). Few chloroplast genes transferred to the nucleus have been reported; the only examples are *infA* in several angiosperms (Millen et al. 2001), *rpl22* in legumes (Gantt et al. 1991), and *rpl32* in *Populus* (Cusack and Wolfe 2007; Ueda et al. 2007). The chloroplast and mitochondrial genomes of angiosperms provide a stark contrast in stable versus variable gene content and few versus many transfers of genes to the nucleus.

Supplementary Material

Supplementary tables S1–S4 and supplementary figures S1–S7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Mike Barker and Nolan Kane for advice on rate analyses, James Robertson and Moira Scasitelli for comments on the manuscript, and the UBC Botanical Garden for plant material. This study was funded by a grant from the Natural Science and Engineering Research Council (NSERC) of Canada, by start-up funds from the Faculty of Land and Food Systems at the University of British Columbia, and by infrastructure funds from the Canadian Foundation for Innovation. P.Z. was supported in part by an Undergraduate Student Research Award from NSERC.

Literature Cited

Adams KL, Daley DO, Qiu YL, Whelan J, Palmer JD. 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature. 408:354–357.

Adams KL, Ong HC, Palmer JD. 2001. Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene rpl2 and partial or complete gene transfer to the nucleus. Mol Biol Evol. 18:2289–2297.

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 29:380–395.

Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA. 99:9905–9912.

Adams KL, Rosenblueth M, Qiu YL, Palmer JD. 2001. Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. Genetics. 158:1289–1300.

Adams KL, Song K, Roessler PG, Nugent JM, Doyle JL, Doyle JJ, Palmer JD. 1999. Intracellular gene transfer in action: dual transcription and multiple silencing of nuclear and mitochondrial cox2 genes in legume. Proc Natl Acad Sci USA. 96:13863–13868.

Baker A, Schatz G. 1987. Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. Proc Natl Acad Sci USA. 84:3117–3121.

Bininda-Emonds OR. 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. BMC Bioinformatics. 6:156.

Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distribution of duplicate gene. Plant Cell. 16:1667–1678.

Blanchard JL, Lynch M. 2000. Organellar genes: why do they end up in the nucleus? J Mol Evol. 41:397–406.

Bonen L, Calixte S. 2006. Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. Mol Biol Evol. 23:701–712.

Chaw SM, Shih AC, Wang D, Wu YW, Liu SM, Chou TY. 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol. 25:603–615.

Choi C, Liu Z, Adams KL. 2006. Evolutionary transfers of mitochondrial genes to the nucleus in the Populus lineage and coexpression of nuclear and mitochondrial *Sdh4* genes. New Phytol. 172:429–439.

Claros MG, Vincens P. 2006. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem. 241:779–786.

Cusack BP, Wolfe KH. 2007. When gene marriages don't work out: divorce by subfunctionalization. Trends Genet. 23: 270–272.

D'Errico I, Gadaleta G, Saccone C. 2004. Pseudogenes in metazoa: origin and features. Brief Funct Genomic Proteomic. 3:157–167.

Elorza A, León G, Gómez I, Mouras A, Holuigue L, Araya A, Jordana X. 2004. Nuclear SDH2-1 and SDH2-2 genes, encoding the iron–sulfur subunit of mitochondrial complex II in *Arabidopsis*, have distinct cell-specific expression patterns and promoter activities. Plant Physiol. 136:4072–4087.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 300:1005–1016.

Erpenbeck D, Voigt O, Adamski M, Adamska M, Hooper JN, Wörheide G, Degnan BM. 2007. Mitochondrial diversity of early-branching metazoa is revealed by the complete mt genome of a haplosclerid demosponge. Mol Biol Evol. 24:19–22.

Fallahi M, Crosthwait J, Calixte S, Bonen L. 2005. Fate of mitochondrially located S19 ribosomal protein genes after transfer of a functional copy to the nucleus in cereals. Mol Genet Genomics. 273:76–83.

Felsenstein J. 2008. PHYLIP (phylogeny inference package), version 3.6.8. Seattle: University of Washington.

Figueroa P, Gómez I, Holuigue L, Araya A, Jordana X. 1999. Transfer of rps14 from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron–sulphur subunit of succinate dehydrogenase and expression by alternative splicing. Plant J. 18:601–609.

Figueroa P, León G, Elorza A, Holuigue L, Jordana X. 2001. Three different genes encode the iron–sulfur subunit of succinate dehydrogenase in *Arabidopsis thaliana*. Plant Mol Biol. 46:241–250.

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. 1991. Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 10:3073–3078.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity. 101:301–320.

Glaser E, Sjöling S, Tanudji M, Whelan J. 1998. Mitochondrial protein import in plants. Signals, sorting, targeting, processing and regulation. Plant Mol Biol. 38:311–338.

Gray MW. 1992. The endosymbiont hypothesis revisited. Int Rev Cytol. 141:233–357.

Gray MW. 1999. Evolution of organellar genomes. Curr Opin Genet Dev. 9:678–687.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 41:95–98.

Hazle T, Bonen L. 2007. Status of genes encoding the mitochondrial S1 ribosomal protein in closely-related legumes. Gene. 405:108–116.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35:W585–W587.

Jansen RK, Cai Z, Raubeson LA, et al. (11 co-authors). 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA. 104:19369–19374.

Kadowaki K, Kubo N, Ozawa K, Hirai A. 1996. Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. EMBO J. 15:6652–6661.

Kobayashi Y, Knoop V, Fukuzawa H, Brennicke A, Ohyama K. 1997. Interorganellar gene transfer in bryophytes: the functional nad7 gene is nuclear encoded in *Marchantia polymorpha*. Mol Gen Genet. 256:589–592.

Kubo N, Harada K, Hirai A, Kadowaki K. 1999. A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. Proc Natl Acad Sci USA. 96:9207–9211.

Kubo N, Jordana X, Ozawa K, Zanlungo S, Harada K, Sasaki T, Kadowaki K. 2000. Transfer of the mitochondrial rps10 gene to the nucleus in rice: acquisition of the 5′ untranslated region followed by gene duplication. Mol Gen Genet. 263: 733–739.

Laroche J, Li P, Maggia L, Bousquet J. 1997. Molecular evolution of angiosperm mitochondrial introns and exons. Proc Natl Acad Sci USA. 94:5722–5727.

Leister D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet. 21:655–663.

Liu SL, Adams KL. 2008. Molecular adaptation and expression evolution following duplication of genes for organellar ribosomal protein S13 in rosids. BMC Evol Biol. 8:25.

Lynch M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol Biol Evol. 13:209–220.

Millen RS, Olmstead RG, Adams KL, et al. (11 co-authors). 2001. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell. 13:645–658.

Moriguchi K, Sugita M, Sugiura M. 1997. Structure and subcellular localization of a small RNA-binding protein from tobacco. Plant J. 12:215–221.

Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol. 7:135.

Murcha MW, Rudhe C, Elhafez D, Adams KL, Daley DO, Whelan J. 2005. Adaptations required for mitochondrial import following mitochondrial to nucleus gene transfer of ribosomal protein S10. Plant Physiol. 138:2134–2144.

Oda K, Yamato K, Ohta E, et al. (11 co-authors). 1992. Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol. 223:1–7.

Ong HC, Palmer JD. 2006. Pervasive survival of expressed mitochondrial rps14 pseudogenes in grasses and their relatives for 80 million years following three functional transfers to the nucleus. BMC Evol Biol. 6:55.

Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci. 12: 357–358.

Rensing SA, Lang D, Zimmer AD, et al. (70 co-authors). 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science. 319:64–69.
Sandoval P, León G, Gómez I, Carmona R, Figueroa P, Holuigue L, Araya A, Jordana X. 2004. Transfer of RPS14 and RPL5 from the mitochondrion to the nucleus in grasses. Gene. 324:139–147.
Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics. 4:1581–1590.
Soltis DE, Soltis PS, Endress PK, Chase MW. 2005. Phylogeny and evolution of angiosperm. Sunderland (MA): Sinauer Associates. p. 370
Subramanian S, Bonen L. 2006. Rapid evolution in sequence and length of the nuclear-located gene for mitochondrial L2 ribosomal protein in cereals. Genome. 49:275–281.
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.
Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwara M, Sato N. 2007. The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. Mol Biol Evol. 24:699–709.
Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 313:1596–1604.
Ueda M, Fujimoto M, Arimura S, Murata J, Tsutsumi N, Kadowaki K. 2007. Loss of the rpl32 gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. Gene. 402: 51–56.
Ueda M, Fujimoto M, Arimura S, Tsutsumi N, Kadowaki K. 2008. Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. Mol Biol Evol. 25: 1791–1793.
Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA. 84: 9054–9058.
Wright F. 1990. The 'effective number of codons' used in a gene. Gene. 87:23–29.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13: 555–556.
Yang Z, Nielson R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 19:908–917.
Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.

Geoffrey McFadden, Associate Editor

Accepted January 14, 2009

doi:10.1093/molbev/msp061

Comparative Analysis of Structural Diversity and Sequence Evolution in Plant Mitochondrial Genes Transferred to the Nucleus

Shao-Lun Liu, Yan Zhuang, Peter Zhang, and Keith L. Adams

Mol. Biol. Evol. 26: 875–891. 2009. doi:10.1093/molbev/msp011

A footnote in the paper incorrectly cited all four authors as having equal contribution to the article. Only the first two authors equally contributed to this article. The affiliation and footnote should have appeared as follows:

Shao-Lun Liu,[1] Yan Zhuang,[1,2] Peter Zhang, and Keith L. Adams
UBC Botanical Garden and Centre for Plant Research, and Department of Botany, University of British Columbia, Vancouver, BC, Canada

[1] Equal contributions.
[2] Present address: Plant Biology Graduate Group, UC Davis.