RESEARCH ARTICLE

Michael P. Cummings · Jacqueline M. Nugent
Richard G. Olmstead · Jeffrey D. Palmer

Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms

Abstract We used the chloroplast gene *rbcL* as a model to study the frequency and relative timing of transfer of chloroplast sequences to the mitochondrial genome. Southern blot survey of 20 mitochondrial DNAs confirmed three previously reported groups of plants containing *rbcL* in their mitochondrion, while PCR studies identified a new mitochondrial *rbcL*. Published and newly determined mitochondrial and chloroplast *rbcL* sequences were used to reconstruct *rbcL* phylogeny. The results imply five or six separate interorganellar transfers of *rbcL* among the angiosperms examined, and hundreds of successful transfers across all flowering plants. By taxonomic criteria, the crucifer transfer is the most ancient, two separate transfers within the grass family are of intermediate ancestry, and the morning-glory transfer is most recent. All five mitochondrial copies of *rbcL* examined exhibit insertion and/or deletion events that disrupt the reading frame (three are grossly truncated); and all are elevated in the proportion of nonsynonymous substitutions, providing clear evidence that these sequences are pseudogenes.

Electronic Supplementary Material is available for this article if you access the article at http://dx.doi.org/10.1007/s00294-003-0378-3. A link in the frame on the left on that page takes you directly to the supplementary material.

Keywords Chloroplast · Mitochondrion · Interorganellar transfer · Pseudogene

Electronic Supplementary Material is available for this article if you access the article at http://dx.doi.org/10.1007/s00294-003-0378-3. A link in the frame on the left on that page takes you directly to the supplementary material.

Communicated by F.-A. Wollman

M.P. Cummings (✉)
The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
Marine Biological Laboratory,
7 MBL Street, Woods Hole,
MA 02543-1015, USA
E-mail: mike@mbl.edu
Tel.: +1-508-2897260
Fax: +1-508-4574727

J.M. Nugent · R.G. Olmstead · J.D. Palmer
Department of Biology,
Indiana University,
Bloomington, IN 47405, USA

Present address: J.M. Nugent
Biology Department, The National University of Ireland,
Maynooth, Co. Kildare, Ireland

Present address: R.G. Olmstead
Department of Botany,
University of Washington,
Box 355255, Seattle, WA 98195, USA

Introduction

The movement of DNA between the physically distinct, but functionally interacting genetic systems of the nucleus, plastid, and mitochondrion has been an important force in structuring these genomes. For example, the greatly reduced size of organelle genomes compared with their once free-living bacterial progenitors is in part a consequence of massive gene transfer to the nucleus (Gray 1983; Palmer 1992).

The first report of chloroplast sequences in the mitochondrial genome was that of Stern and Lonsdale (1982) who described a 12-kb sequence in maize mitochondrial DNA (mtDNA) originating from the inverted repeat region of the chloroplast genome. Many subsequent studies have demonstrated the presence of numerous chloroplast sequences in the mitochondrial genome of all angiosperms examined [but not *Marchantia* (Oda et al. 1992) or any green algae (Turmel et al. 2002; http://megasun.bch.umontreal.ca/ogmp/projects/other/mt_list.html)], whereas no sequences of mitochondrial or nuclear origin have ever been discovered in plant chloroplast DNAs (cpDNAs; for a review see Lonsdale 1989; see also Stern and Palmer 1986; Nugent and Palmer

1988; Schuster and Brennicke 1988; Siculella and Palmer 1988; Folkerts and Hanson 1989; Palmer 1992; Nakazono and Hirai 1993; Marienfeld et al. 1999; Kubo et al. 2000).

While many studies have documented the number and extent of chloroplast-derived sequences present in various angiosperm mtDNAs, very little is known about the dynamics and frequency of these sequence transfer events and the subsequent evolution of the transferred sequences in the mitochondrion. Nugent and Palmer (1988) examined the timing of transfer of cpDNA sequences into the mitochondrial genome of six crucifer species. Although the study demonstrated sequential transfer of various sequences from the chloroplast to the mitochondrion during evolution of Brassicaceae, the study could only hint at the dynamics of DNA transfer operable throughout angiosperm evolution—predicting a relatively slow, but ongoing process.

Our objectives in the present study were to use *rbcL* as a representative chloroplast sequence to assess the distribution of chloroplast sequences in the mitochondrial genome of angiosperms, to determine the frequency and relative timing of chloroplast-to-mitochondrion transfer events with respect to the species divergence, and to examine the dynamics of sequence evolution following intracellular gene transfer. The primary advantages of using *rbcL* (which encodes the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase) as a model are the vast database for chloroplast *rbcL* sequences currently available for flowering plants (e.g., Källersjö et al. 1998) and the fact that mitochondrial *rbcL* sequences have already been characterized from several angiosperms (Lonsdale et al. 1983; Moon et al. 1987; Stern 1987; Nugent and Palmer 1988; Unseld et al. 1997).

Materials and methods

DNA manipulations

mtDNA samples were isolated from either green leaves or etiolated seedlings, according to the DNase I procedure of Kolodner and Tewari (1972). Southern blot hybridization was carried out at 60 °C in Blotto solution [0.5% non-fat dry milk, 1% SDS, 4× saline-sodium citrate (SSC), preheated to 60 °C]. After hybridization, the blot was washed in 2× SSC, 0.5% SDS, with three 5-min washes at room temperature followed by three 30-min washes at 60 °C.

The clones used for sequencing had several sources. The *Brassica rapa* clones containing *rbcL* were derived from the mitochondrial clone bank of Palmer and Shields (1984) and the chloroplast clone bank of Nugent and Palmer (1988). The maize mitochondrial clone containing the *rbcL* gene, pLSH20 of Lonsdale et al. (1983) was kindly provided by D.M. Lonsdale. Chloroplast and mitochondrial copies of *rbcL* from *Ipomoea coccinea* were amplified by PCR and subsequently cloned into the plasmid vector BlueScript SK+ (Stratagene), as described by Olmstead et al. (1992).

All sequencing was by the dideoxynucleotide chain termination method, using a combination of subcloning into M13mp18 and M13mp19, cloning into plasmid vectors, a series of nested deletions created by Exonuclease III digestion (Erase-A-Base, Promega), and universal and specific synthetic oligonucleotide primers, including

some kindly provided by G. Zurawski. The original sequences reported here were deposited in GenBank under accession numbers AY167975–AY167986.

Data analyses

Phylogenetic relationships were determined for the transposed copies of *rbcL* present in the mitochondrial genome, along with sequences taken from GenBank for genuine chloroplast *rbcL* from the same species and other related taxa. Sequence alignments were done using the program CLUSTAL W ver 1.81 (Thompson et al. 1994). Phylogenetic relationships were determined by maximum likelihood analyses using the program PAUP* ver 4.0b10 (Swofford 2002). For each data set, ten separate heuristic searches were conducted, using random sequence addition and tree bisection/reconnection branch-swapping. The likelihood model parameters were as follows: six nucleotide substitution categories, Γ distribution model for substitution rate variation among sites approximated with four discrete rate classes represented by the mean of each class, and a proportion of the sites assumed invariant. Values of the Γ distribution shape parameter, α, nucleotide substitution rates, proportion of invariant sites, and base frequencies were estimated from the data, based on a initial neighbor-joining tree. These initial parameter estimates were applied in heuristic searches. Parameters were re-estimated for the final tree and applied in the analysis of the 2000 bootstrap samples. Alternative topologies were compared with a log likelihood test (Kishino and Hasegawa 1989) using PAUP*.

Ancestral sequences were inferred with the method of Yang et al. (1995), using the program baseml from PAML ver 3.1 (Yang 1997). The mitochondrial and chloroplast copies of *rbcL* were each compared with the inferred sequence for their most recent common ancestor, to determine the nucleotide substitution patterns. The numbers of nonsynonymous substitutions per nonsynonymous site, d_N, synonymous substitutions per synonymous site, d_S, and ratio of the two, $\omega = d_N/d_S$, were determined with the method of Yang and Nielsen (2000), using the program yn00 from PAML.

Results

Presence of *rbcL* sequences in angiosperm mitochondrial genomes

mtDNA preparations from 20 angiosperms were tested for the presence of *rbcL* sequences by Southern blot hybridization. A 1.2-kb fragment internal to the pea (*Pisum sativum*) chloroplast *rbcL* gene hybridized to many of the DNAs (Fig. 1); and all three mtDNAs not shown in the figure [spinach (*Spinacia oleracea*), watermelon (*Citrullus lanatus*), petunia (*Petunia hybrida*)] did not hybridize to the *rbcL* probe. The hybridization signals marked *c* in Fig. 1 are attributed to contaminating cpDNA, based on the known sizes of chloroplast *rbcL*-containing fragments in these species and on previous *rbcL* hybridizations to side-by-side digests of mtDNA and cpDNA for each of these taxa, except *Lactuca sativa* (see Electronic supplementary material). All but one of the remaining signals (marked *m*) are attributed to bona fide *rbcL* hybridization to mtDNA. Moreover, for maize (Fig. 1, lane 2), the five crucifers (lanes 11–15), and cucumber (lane 17), previous studies have already established the presence of *rbcL* sequences in mtDNA
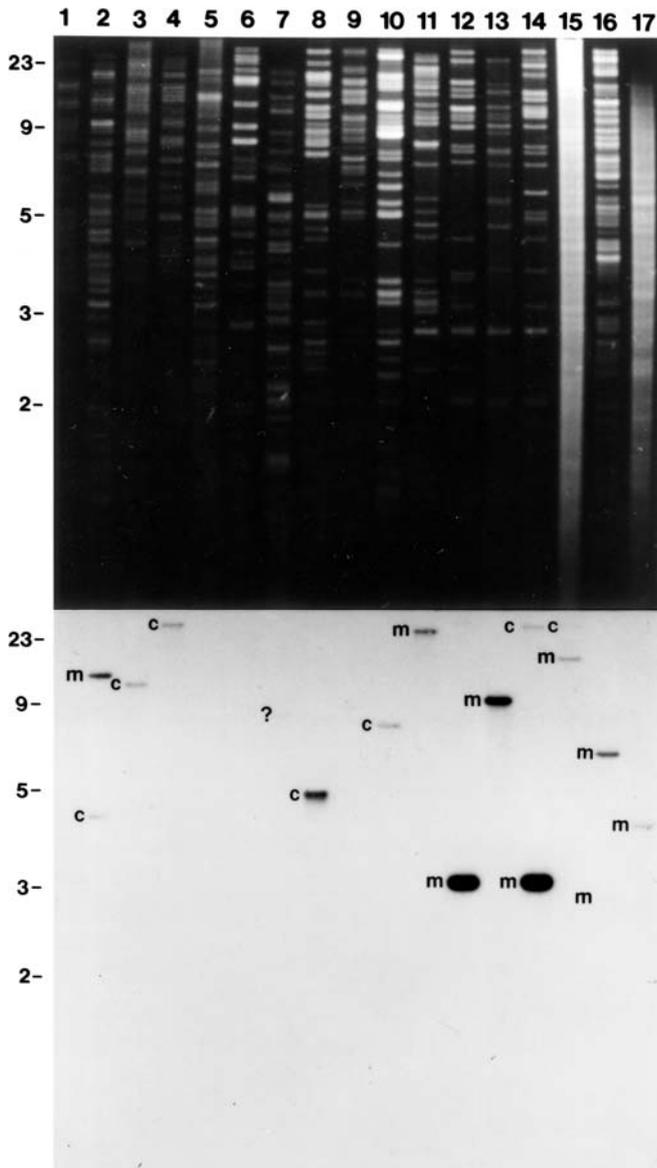
Fig. 1 Detection of *rbcL* sequences in angiosperm mitochondrial (mt)DNAs by Southern blot hybridization. *Top panel* Electrophoresis in a 1.0% agarose gel of mtDNA from *Triticum aestivum* (wheat, *lane 1*, *Pst*I fragments), *Zea mays* (maize, *lane 2*, *Bam*HI), *Lactuca sativa* (lettuce, *lane 3*, *Xho*I), *Helianthus annuus* (sunflower, *lane 4*, *Pst*I), *Solanum lycopersicum* (tomato, *lane 5*, *Bam*HI), *Nicotiana tabacum* (tobacco, *lane 6*, *Pst*I), *Oenothera johansen* (evening primrose, *lane 7*, *Hin*dIII), *Pisum sativum* (pea, *lane 8*, *Pst*I), *Phaseolus vulgaris* (common bean, *lane 9*, *Pst*I), *Vigna radiata* (mung bean, *lane 10*, *Pst*I), *Capsella bursa-pastoris* (shepherds purse, *lane 11*, *Pst*I), *Crambe abyssinica* (*lane 12*, *Pst*I), *Raphanus sativus* (radish, *lane 13*, *Pst*I), *Brassica rapa* (mustard, *lane 14*, *Pst*I), *Arabidopsis thaliana* (*lane 15*, *Pst*I), *Cucurbita maxima* (squash, *lane 16*, *Pvu*II), and *Cucumis sativus* (cucumber, *lane 17*, *Pvu*II). *Bottom panel* Hybridization to a filter blot of the above agarose gel with a ^{32}P-labeled plasmid containing a 1,167-bp, *Pst*I-*Hin*dIII, *rbcL* gene internal fragment from pea chloroplast DNA inserted into plasmid vector pIC20R. *The letters m and c* indicate bands of putative mitochondrial and chloroplast origin, respectively. *Numerals at the left side* provide a scale for approximate size in kilobasepairs

(Lonsdale et al. 1983; Stern and Palmer 1984; Stern 1987; Nugent and Palmer 1988). The only ambiguous signal with respect to genome of origin is the weak signal for *Oenothera johansen* (lane 7; this signal was clear in the original autoradiogram and on longer exposures, but does not reproduce well), for which a cpDNA restriction site map is not available for *Hin*dIII.

These hybridizations identify one clear-cut new case of *rbcL* sequences in mtDNA, in squash (*Cucurbita maxima*), although mitochondrial *rbcL* has already been reported in another member of the same genus (*C. pepo*, zucchini; Stern 1987), in another genus in the same family (*Cucumis sativus*, cucumber; Stern 1987; Fig. 1), and one ambiguous case in *Oenothera* (see above). Except for the Poaceae, for which maize (lane 2) but not wheat (lane 1; Watanabe et al. 1994) showed mtDNA hybridization, a consistent pattern of entirely positive or negative hybridization was observed within each of the six multiple-sampled families within the present study.

mtDNA from morning glory (*I. coccinea*) was not included in the above Southern blot survey. However, a potential mitochondrial *rbcL* sequence (i.e., a clear pseudogene, see next section) was amplified from morning glory together with an intact *rbcL* gene as part of an independent *rbcL* phylogenetic study. Diagnostic Southern blot experiments were then performed on mtDNA- and cpDNA-enriched fractions from morning glory, which showed that the intact *rbcL* sequence is located in the chloroplast genome and the *rbcL* pseudogene is located in the mitochondrial genome (data not shown).

Relationships of chloroplast and mitochondrial copies of *rbcL*

We sequenced mitochondrial *rbcL* genes from three diverse angiosperms: maize, morning glory, and the crucifer *B. rapa* (for details of the sequences and their alignments, see Electronic supplementary material). Previously determined sequences of mitochondrial *rbcL* were available from rice (Nakazono and Hirai 1993) and *Arabidopsis* (Unseld et al. 1997). Phylogenetic analyses were conducted to determine the relationships of these five transferred sequences of *rbcL* and those *rbcL* genes still resident in the chloroplast genome. This was done to estimate the timing of transfer events from the chloroplast genome to the mitochondrial genome, relative to species divergence. Preliminary analyses strongly suggested that the five mitochondrial *rbcL* sequences are the result of three clearly separate transfer events or pairs of events and therefore separate analyses were subsequently performed that focused on taxa closely related to those containing the mitochondrial *rbcL* sequences analyzed in this study.

The phylogenetic tree (Fig. 2) for *rbcL* from Poaceae and an outgroup family is congruent with the parsimony-based tree of Duvall and Morton (1996); and only the
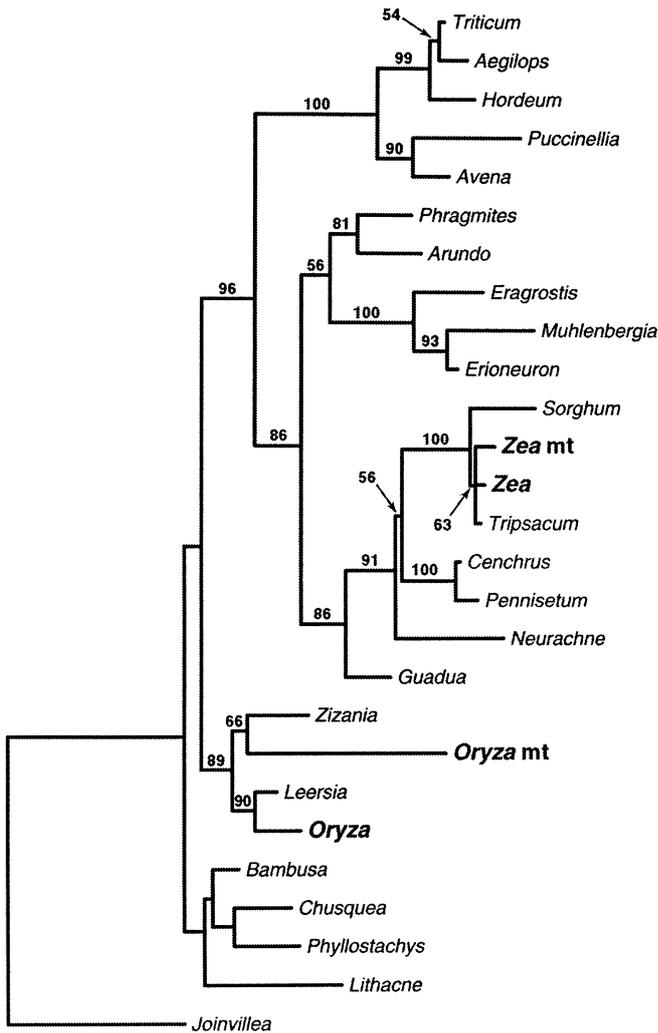
Fig. 2 *rbcL* phylogeny showing two separate interorganellar transfers within Poaceae evolution. Shown is a maximum likelihood tree with branch lengths proportional to the amount of inferred nucleotide differences for Poaceae and the outgroup taxon (Jonivilleaceae: *Joinvillea plicata*). *The letters mt* indicate sequences of putative mitochondrial origin; and all others are chloroplast sequences. *Numerals adjacent to branches* indicate the percentage of bootstrap replicates supporting the clade out of 2,000 bootstrap replicates. The log likelihood value is −4,979.59

Fig. 3 *rbcL* phylogeny showing a single interorganellar transfer early in Brassicaceae evolution. Shown is a maximum likelihood tree with branch lengths proportional to the amount of inferred nucleotide differences for Brassicaceae and outgroup taxa (Capparidaceae: *Polansisia dodecandra, Capparis spinosa, C. hastata, Crateva odora*; Tovariaceae: *Tovaria pendula*; Resedaceae: *Reseda alba*). *The letters mt* indicate sequences of putative mitochondrial origin; and all others are chloroplast sequences. *Numerals adjacent to branches* indicate the percentage of bootstrap replicates supporting the clade out of 2,000 bootstrap replicates. The log likelihood value is −3,418.11

placement of *Guadua* appears contrary to other evidence (E. Kellogg, personal communication). Within the Poaceae, the phylogenetic analysis strongly indicates that the sequences of *rbcL* in the mitochondrial genomes of maize and rice originate from separate transfer events. Each mitochondrial copy of *rbcL* is clearly more closely related to the chloroplast copy from the same species than the two are to each other (Fig. 2). This interpretation is supported by the Southern blot survey, which did not show evidence for *rbcL* in the mitochondrial genome of wheat (Fig. 1). The phylogenetic placement of rice mitochondrial *rbcL* in Fig. 2 [(*Oryza* mt, *Zizania*) (*Oryza, Leersia*)] is obviously incorrect if one assumes strictly vertical inheritance. We therefore evaluated three alternative topologies and estimated the log likelihood

associated with each: (1) the rice mitochondrial sequence as sister to the rice chloroplast sequence {*Zizania* [(*Oryza* mt, *Oryza*) *Leersia*]}, implying that the transposition event took place after the divergence of rice from *Leersia*, (2) the rice mitochondrial sequence as sister to the chloroplast sequences of rice and *Leersia* {*Zizania* [*Oryza* mt (*Oryza, Leersia*)]}, implying that the transposition event took place after the divergence of rice and *Leersia* from *Zizania*, but before the divergence of rice from *Leersia*, (3) the rice mitochondrial sequence as sister to the other three chloroplast sequences {*Oryza* mt [*Zizania* (*Oryza, Leersia*)]}, implying that the transposition event took place in an ancestor common to all three genera. None of the log likelihood values for these alternative topologies was significantly different from that of the tree shown in Fig. 2, as determined by the log likelihood test of Kishino and Hasegawa (1989). Therefore, we conclude that the transferred *rbcL* gene was probably vertically transmitted through the mitochondrial lineage of *Oryza* and its relatives, but that poor resolution of the *rbcL* phylogeny precludes a precise estimate of the timing of transfer.
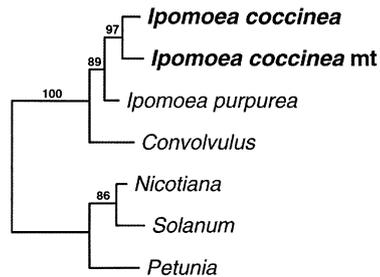
Fig. 4 *rbcL* phylogeny showing a recent interorganellar transfers within the genus *Ipomoea*. Shown is a maximum likelihood tree with branch lengths proportional to the amount of inferred nucleotide differences for Convolvulaceae and outgroup taxa (Solanaceae: *Solanum lycopersicum*, *Nicotiana tabacum*, *Petunia hybrida*). *The letters mt* indicate sequences of putative mitochondrial origin; and all others are chloroplast sequences. *Numerals adjacent to branches* indicate the percentage of bootstrap replicates supporting the clade out of 2,000 bootstrap replicates. The log likelihood value is −2,951.05

The phylogenetic tree (Fig. 3) for Brassicaceae and related families is largely congruent with the parsimony-based tree of Rodman (1993). In contrast to the separate phylogenetic placement and origin of the two grass mt *rbcL* sequences, the *Brassica* and *Arabidopsis* mitochondrial *rbcL* sequences were each members of a trichotomy, with the third member being the entire set of ten Brassicaceae chloroplast *rbcL* sequences analyzed, including the three very closely related *Brassica* sequences and the one *Arabidopsis* sequence. From this it can be inferred that, most likely, a single transfer event occurred at about the time of the origin of the monophyletic Brassicaceae from the paraphyletic Capparaceae. This inference is congruent with that of Nugent and Palmer (1988), who showed that *rbcL* is present in the same mitochondrial genome location in diverse crucifers.

The relationships inferred for taxa within Convolvulaceae and Solanaceae as shown in Fig. 4 are congruent with the parsimony tree of Soltis et al. (2000). The tree shows that the chloroplast and mitochondrial copies of *rbcL* from *I. coccinea* share a more recent common ancestor than either does with a second species in the genus, *I. pupurea*. This implies that this transfer of *rbcL* from the chloroplast genome to the mitochondrial

genome occurred rather recently, subsequent to the divergence of the two *Ipomoea* species.

Comparison of chloroplast and mitochondrial sequence evolution

Examination of chloroplast and mitochondrial sequences from the same species virtually eliminates possible confounding effects associated with differences in species history because, except for a very few unusual plants, both the chloroplast and the mitochondrion are always transmitted through the same, usually maternal, parent. Comparing mitochondrial and chloroplast sequences with their inferred common ancestor, rather than directly with each other, allows one to account for differences subsequent to the transfer event that have occurred in each of the sequences independently. Furthermore, using the inferred most recent common ancestral sequence as a reference allows for polarization of insertion, deletion, and substitution events (e.g., substitution of A-to-T is distinguishable from T-to-A). We inferred ancestral sequences, based on the trees shown in Figs. 2, 3, 4, using the method of Yang et al. (1995).

Although precise comparisons are confounded, because of the inability to separate the effects of selection on functional characteristics (e.g., amino acid sequence, codon usage bias) from genome-specific substitution patterns, some broad patterns can be noted. For three of the five comparisons (all but *Brassica* and *Arabidopsis*), mitochondrial *rbcL* has more inferred nucleotide substitutions, n_d, than chloroplast *rbcL*, although the differences are subtle in some cases (Table 1). A clearer trend is seen in ω, the ratio of nonsynonymous substitutions per nonsynonymous site, d_N, to synonymous substitutions per synonymous site, d_S (Table 1). In all cases, mitochondrial *rbcL* has a higher ω value (in four of five cases, much higher) than chloroplast *rbcL* (Table 1). These observations suggest that mitochondrial *rbcL* is evolving as a pseudogene. Consistent with this hypothesis, three of the mitochondrial genes (from *Ipomoea* and both crucifers) are grossly truncated and all five mitochondrial genes have sustained one or more internal frameshift mutations (see Table 1 for total

Table 1 Divergence of chloroplast and mitochondrial sequences of *rbcL* from their inferred most recent common ancestral sequence. *Brassica* and *Arabidopsis* represent the same transfer event (see text).n_d Number of nucleotide differences, d_N nonsynonymous substitutions per nonsynonymous site, d_S synonymous substitutions per synonymous site, $\omega = d_N/d_S$,*indels* number of internal insertions/deletions (with number of frameshifts in parentheses)

Taxon	Length (bp)	Chloroplast					Mitochondrion				
		n_d	d_N	d_S	ω	indels	n_d	d_N	d_S	ω	indels
Zea mays	1,410	2	0.0000	0.0059	0.0000	0	4	0.0029	0.0000	>1.0000	5 (5)
Oryza sativa	1,411	15	0.0037	0.0347	0.1073	0	39	0.0267	0.0392	0.6804	4 (3)
Brassica rapa	1,118	26	0.0069	0.0904	0.0759	0	10	0.0072	0.0110	0.6541	1 (1)
Arabidopsist haliana	574	7	0.0045	0.0403	0.1122	0	5	0.0118	0.0000	>1.0000	2 (2)
Ipomoea coccinea	1,102	6	0.0034	0.0133	0.2600	0	7	0.0037	0.0123	0.2974	2 (2)

number of indels and associated frameshifts for each of the five genes).

Discussion

The phylogenetic analyses of chloroplast and mitochondrial *rbcL* sequences presented here allow us to infer at least five independent trans-organellar transfers of this gene during angiosperm evolution. Four of these transfers are illustrated in the trees presented in Figs. 2, 3, 4, while a fifth transfer follows from the fact that each of these transfers is family-specific and therefore the unsequenced mitochondrial *rbcL* genes of cucurbits (Cucurbitaceae) must be of separate derivation. Both species of *Cucurbita* examined (*C. maxima*, squash; *C. pepo*, zuchini) contain mitochondrial *rbcL* sequences, whereas only one of two examined *Cucumis* species (*C. sativus*, cucumber; but not *C. melo*, muskmelon) does (Stern 1987; this study). Therefore, sequencing and phylogenetic analyses of cucurbit *rbcL* sequences are needed to determine whether this heterogeneous pattern reflects two independent *rbcL* transfers in the family (as seen in the grass family; Fig. 2), i.e., one each in *Cucurbita* and *Cucumis* lineages, or a more ancient transfer (as in the crucifer family; Fig. 3) with subsequent loss of mitochondrial *rbcL* in *C. melo*.

To our knowledge, this is the first study to document many independent (and relatively recent) transfers of a particular chloroplast sequence to the mitochondrial genome during plant evolution. In some sense, this is surprising, for by 1988 copies of *rbcL* were known to be present in the mitochondrial genome of a diverse array of angiosperms (i.e., maize, rice, crucifers, cucurbits; Lonsdale et al. 1983; Moon et al. 1987; Stern 1987; Nugent and Palmer 1988). However, most of those transfers were characterized only by Southern blot hybridization. Our ability in this study to derive a clear understanding of whether the widespread presence of *rbcL* in plant mtDNA reflects primarily one or a few ancient transfers (and many subsequent losses in those plants shown to lack it in their mtDNAs) or many relatively recent transfers results from our having: (1) sequenced mitochondrial *rbcL* from three diverse plants and (2) carried out the first phylogenetic analyses designed to elucidate the frequency and recency of *rbcL* transfer. Many other chloroplast-to-mitochondrial transfers have been inferred to be recent; and these inferences rest essentially entirely on the basis of a high sequence similarity of trans-organellar duplicates from within a plant (e.g., Marienfeld et al. 1999), as opposed to formal phylogenetic analysis as presented here. However, other transfers are known to be relatively ancient and thus it was not at all clear when going into this study what picture would emerge for *rbcL*, with respect to its frequency and timing of transfer. The independent transfers of *rbcL* shown here in the grass lineages leading to rice and maize stand in contrast to the more ancient transfer of other chloroplast sequences (such as the *rps19/trnH* region) in their common ancestor, early in grass evolution (Kanno et al. 1997), much less the truly ancient transfer of a tRNA gene cluster very early in angiosperm evolution, in the common ancestor of monocots and a broad array of dicots (Kubo et al. 1995).

We find it remarkable that at least five, possibly six, independent interorganellar transfers of *rbcL* are now documented, considering how poorly angiosperms have been sampled in this regard (only 8 of over 300 families and 22 of over 300,000 species have been examined). By extrapolation, it seems reasonable to infer that *rbcL* transfer to mtDNA has occurred many more times during angiosperm evolution, perhaps even hundreds of times.

The frequent presence of *rbcL* and other chloroplast sequences in mtDNA, combined with the exceptionally low substitution rate in plant mtDNA (Gaut 1998), highlights the potential for PCR amplification of non-chloroplast sequences, even when using primers based on chloroplast sequences. At the same time, however, the common occurrence of frameshift mutations in these transferred sequences (Table 1), combined with the rarity of pseudogenes in cpDNA, provides excellent clues that a particular PCR-generated sequence might be the product of interorganellar transfer [e.g., see the unambiguous pseudogene of *rbcL* identified by PCR in *Beaumontia* (Sennblad et al. 1998) and the likely *rbcL* pseudogene identified in *Canella* (the *Canella* B sequence) by Qiu et al. (1993)].

From the standpoint of taxonomic rank, the crucifer transfer is the most ancient (in the common ancestor of all Brassicaceae; Fig. 3) of the four *rbcL* transfer cases for which sequence and phylogenetic information are available, the two grass transfers are of intermediate ancestry [occurring within the rice and maize subfamilies (Oryzoideae and Panicoideae, respectively); Fig. 2], and the morning-glory transfer is the most recent (within genus *Ipomoea*; Fig. 4). From the standpoint of molecular divergence, the timing of transfer could in principle be estimated from the amount of either chloroplast or mitochondrial sequence divergence (Table 1). However, because substitution rates can vary widely in both mitochondrial and chloroplast genomes of plants (Gaut et al. 1992; Eyre-Walker and Gaut 1997; Gaut 1998) and because relatively few substitutions have occurred (Table 1), it is probably unwise to invoke any molecular clock assumptions to infer the relative (or absolute) timing of these transfers.

Two lines of evidence, frequent truncations and frameshift mutations and higher value for ω (Table 1), make it clear that the five sequenced mitochondrial *rbcL* loci are all pseudogenes. This is not surprising, given that the only known function in land plants of the large subunit of ribulose-1,5-bisphosphate carboxylase-oxygenase (the *rbcL* gene product) is in photosynthetic carbon fixation. However, it is not out of the question to imagine that *rbcL* might occasionally be co-opted into

some pathway of mitochondrial carbon metabolism, given that functional *rbcL* genes are present in a number of nonphotosynthetic archaebacteria and eubacteria (Maeda et al. 1999) and in the permanently nonphotosynthetic plastid of the euglenoid alga *Astasia longa* (Siemeister and Hachtel 1990).

References

Duvall MR, Morton BR (1996) Molecular phylogenetics of Poaceae: an expanded analysis of *rbcL* sequence data. Mol Phylogenet Evol 5:352–358

Eyre-Walker A, Gaut BS (1997) Correlated substitution rates among plant genomes. Mol Biol Evol 14:455–460

Folkerts O, Hanson MR (1989) Three copies of a single recombination repeat occur on the 443 kb mastercircle of the *Petunia hybrida* 3704 mitochondrial genome. Nucleic Acids Res 17:7345–7357

Gaut BS (1998) Molecular clocks and nucleotide substitution rates higher plants. Evol Biol 30:93–120

Gaut BS, Muse SV, Clark W, Clegg MT (1992) Relative substitution at the *rbcL* locus in monocotyledonous plants. J Mol Evol 35:292–303

Gray MW (1983) The bacterial ancestry of plastids and mitochondria. Bioscience 33:693–699

Källersjö M, Farris JS, Chase MW, Bremer B, Fay MF, Humphries CJ, Petersen G, Seberg O, Bremer K (1998) Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. Plant Syst Evol 213:259–287

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

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and branching order in Hominoidea. J Mol Evol 229:170–179

Kolodner R, Tewari KK (1972) Physiochemical characterization of mitochondrial DNA from pea leaves. Proc Natl Acad Sci USA 69:1830–1834

Kubo T, Yanai Y, Kinoshita T, Mikami T (1995) The chloroplast *trnP-trnW-petG* gene cluster in the mitochondrial genomes of *Beta vulgaris*, *B. trigyna* and *B. webbiana*: evolutionary aspects. Curr Genet 27:285–289

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

Lonsdale DM (1989) The plant mitochondrial genome. In: Marcus A. (ed) The biochemistry of plants, vol 15. Academic Press, New York, pp 229–295

Lonsdale DM, Hodge TP, Howe CJ, Stern DB (1983) Maize mitochondrial DNA contains a sequence homologous to the ribulose-1,5-bisphosphate carboxylase large subunit gene of chloroplast DNA. Cell 34:1007–1014

Maeda N, Kitano K, Fukui T, Ezaki S, Atomi H, Miki K, Imanaka T (1999) Ribulose bisphosphate carboxylase/oxygenase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 is composed solely of large subunits and forms a pentagonal structure. J Mol Biol 293:57–66

Marienfeld J, Unseld M, Brennicke A (1999) The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. Trends Plant Sci 4:495–502

Moon E, Kao T-H, Wu R (1987) Rice chloroplast DNA molecules are heterogeneous as revealed by DNA sequences of a cluster of genes. Nucleic Acids Res 15:611–630

Nakazono M, Hirai A (1993) Identification of the entire set of transferred chloroplast DNA sequences in the mitochondrial genome of rice. Mol Gen Genet 236:341–346

Nugent JM, Palmer JD (1988) Location, identity, amount and serial entry of chloroplast DNA sequences in crucifer mitochondrial DNAs. Curr Genet 14:501–509

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced for the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. J Mol Biol 223:1–7

Olmstead RG, Michaels HJ, Scott KM, Palmer JD (1992) Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL*. Ann Mo Bot Gard 79:249–265

Palmer JD (1992) Comparison of chloroplast and mitochondrial genome evolution in plants. In: Herrmann RG (ed) Cell organelles. Springer, Berlin Heidelberg New York, pp 99–133

Palmer JD, Shields CR (1984) Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature 307:437–440

Qiu Y-L, Chase MW, Les DH, Parks CR (1993) Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene *rbcL*. Ann Mo Bot Gard 80:587–606

Rodman J, Price RA, Karol K, Conti E, Sytsma KJ, Palmer JD (1993) Nucleotide sequences of the *rbcL* gene indicate monophyly of mustard oil plants. Ann Mo Bot Gard 80:686–699

Schuster W, Brennicke A (1988) Interorganellar sequence transfer: plant mitochondrial DNA is nuclear, is plastid, is mitochondrial. Plant Sci 54:1–10

Sennblad B, Endress ME, Bremer B (1998) Morphology and molecular data in phylogenetic fraternity: the tribe Wrightieae (Apocynaceae) revisited. Am J Bot 85:1143–1158

Siculella L, Palmer JD (1988) Physical and gene organization of mitochondrial DNA in fertile and male sterile sunflower. CMS-associated alterations in structure and transcription of the *atpA* gene. Nucleic Acids Res 16:3787–3799

Siemeister G, Hachtel W (1990) Structure and expression of a gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase (*rbcL*) in the colourless euglenoid flagellae *Asasia longa*. Plant Mol Biol 14:825–833

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL*, and *atpB* sequences. Bot J Linn Soc 133:381–461

Stern DB (1987) DNA transposition between plant organellar genomes. J Cell Sci Suppl 7:145–154

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

Stern DB, Palmer JD (1984) Extensive and widespread homologies between mitochondrial DNA and chloroplast DNA in plants. Proc Natl Acad Sci USA 81:1946–1950

Stern DB, Palmer JD (1986) Tripartite mitochondrial genome of spinach: physical structure, mitochondrial gene mapping, and location of transposed chloroplast DNA sequences. Nucleic Acids Res 14:5651–5666

Swofford DL (2002) PAUP* 4.0 beta 10. *Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, Mass.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

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

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

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

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 15:555–556

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yang Z, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino-acid sequences. Genetics 141:1641–1650